

Masterarbeit – GEO 511

Geographisches Institut der Universität Zürich, Abteilung Geographische Informationssysteme
Winterthurerstrasse 190, 8057 Zürich

In Zusammenarbeit mit:

Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft WSL, Forschungseinheit für Waldressourcen
und Waldmanagement, Gruppe GIS
Zürcherstrasse 111, 8903 Birmensdorf

Geodaten als Linked Data

Eine Untersuchung zur Strukturierung und
Vernetzung von Umweltdaten für das Semantic Web

31.01.2012

Simon Fleischli

Matrikel-Nr.: 05-053-681

Betreuung Eidg. Forschungsanstalt WSL

Sandro Bischof, M.Sc. UZH

Betreuung Institut für Informatik, Universität Zürich

Dr. Thomas Scharrenbach

Betreuung und Fakultätsvertretung

Prof. Dr. Robert Weibel

Danksagung

Ich möchte mich an dieser Stelle bei allen Personen bedanken, die mich während meiner Studienzeit und besonders im Verlaufe meiner Masterarbeit unterstützt haben.

Ein grosser Dank geht an meinen Betreuer, Sandro Bischof, der mich stets mit seinem fachlichen und technischen Wissen bei meiner Arbeit unterstützt und mit vielen interessanten Diskussionen und Ideen neue Gedanken angestossen hat.

Ebenfalls möchte ich mich bei meinen Mitbetreuern, Dr. Thomas Scharrenbach und Prof. Dr. Robert Weibel, für die wertvollen Impulse und generelle fachliche Unterstützung danken.

Meinen Eltern, die mir das Studium ermöglicht und mich immer unterstützt haben, gilt ein besonderer Dank. Ebenso herzlich möchte ich Vivian Kubon für die tatkräftige und moralische Unterstützung danken.

Linked Data stellt ein neues Paradigma dar, um die heutigen vorhandenen Datenmassen im World Wide Web in eine Ordnung zu bringen und untereinander zu verknüpfen. Dadurch soll neben dem heutigen *Web der Dokumente* auch ein *Web der Daten* geschaffen werden. Das Web der Daten entspricht einem Datenraum, der durchsucht werden kann, analog zu dem heutigen “surfen” im Web.

Diese Arbeit beschäftigt sich mit der Integration von *räumlichen* Daten in dieses Web der Daten. Unter Integration wird in diesem Rahmen die Umformung von Informationen aus den Geodatenätzen in das RDF-Datenmodell verstanden. Das heisst, die Informationen werden strukturiert, indem sie basierend auf einem Vokabular beschrieben und in Beziehung zu anderen Informationen gesetzt werden. Dabei wird im Wesentlichen zwischen Vektor- und Rasterdaten unterschieden, da jeweils die Ansätze zu deren Integrierung unterschiedliche Entwicklungsstände vorweisen. Vektordaten werden bereits von diversen Organisationen und Projekten im RDF-Format bereitgestellt. Dazu zählen beispielsweise die Projekte LinkedGeoData.org¹ und GeoNames² sowie Arbeiten am Ordnance Survey of Great Britain (OS) und beim U.S. Geological Survey (USGS). Alle vier Organisationen verfolgen einen ähnlichen Grundsatz, nämlich die Umformung der Sachattribute, wie auch der Geometrien in RDF. Dazu gehört auch die Abbildung der räumlichen Beziehungen, die sich aus den Geodaten ergeben, im RDF-Datenmodell.

Über die Grundprinzipien der Umformung von *Vektordaten* herrscht eine gewisse Einigkeit. So werden die Geometrien entsprechend einer bestimmten Syntax abgelegt, in der Regel in Form eines Strings. Dieser String kann beispielsweise Informationen zur Art der Geometrie (z.B. Punkt, Linie, Polygon), sowie die Stützpunkte (als Koordinatenpaare) enthalten. Räumliche Beziehungen werden vor der Umformung in RDF ermittelt und explizit als Tripel gespeichert. Trotz des Gebrauches ähnlicher Prinzipien, existiert aber kein einheitlicher Standard zur Umformung, so dass sich die Syntax von Organisation zu Organisation unterscheiden kann. Ein aufkommender Standard ist GeoSPARQL, welcher gleichzeitig eine Abfragesprache sowie ein Datenformat darstellt.

Bei den *Rasterdaten* existieren wenige Ansätze zu deren Umformung und dementsprechend müssen erst Methoden entwickelt werden, wie solch eine Umformung realisiert werden könnte. Bei den Arbeiten des USGS wurde ein möglicher Lösungsweg aufgezeigt, der die GML-Spezifikation miteinbezieht.

In dieser explorativen Arbeit werden zwei eigene Ansätze vorgestellt, welche die Benutzung von räumlichen – insbesondere Rasterdaten – Daten im Zusammenhang mit Linked Data ermöglichen sollen. Zu Beginn der Arbeit wurden Anwendungsfälle (Use Cases) formuliert, die unter Zuhilfenahme von Linked Data beantwortet werden sollten. Dementsprechend geben sie auch die Entwicklungsrichtung der Ansätze vor. Der erste Ansatz der “Vektorisierten Kleinsten Einheiten” (VKE) versucht Vektor-, wie auch Rasterdaten umfassend im RDF-Datenmodell zu repräsentieren. Somit können die RDF-Daten zur Be-

¹ <http://linkedgeodata.org/About>

² <http://www.geonames.org/>

antwortung der Use Cases direkt abgefragt werden. Der zweite Ansatz der “Ausgegliederten Räumlichen Operationen” (ARO) versucht die Vorteile von Linked Data mit externen Technologien, wie beispielsweise einem GIS, zu kombinieren. Dadurch können zur Beantwortung der Anwendungsfälle komplexe räumliche Operationen durchgeführt werden.

Basierend auf dem Ansatz ARO konnten alle Anwendungsfälle gelöst werden. Die Resultate und Implementierungen werden im Rahmen der Arbeit genauer beleuchtet und unter dem Aspekt von Linked Data diskutiert.

Zusammenfassung	I
Abbildungsverzeichnis	VI
Tabellenverzeichnis	VIII
Quelltextverzeichnis	IX
1 Einleitung	1
1.1 Die Idee hinter Semantic Web und Linked Data	1
1.1.1 Das World Wide Web heute	1
1.1.2 Das Semantic Web	2
1.1.3 Linked Data	3
1.1.4 Unterschied von Semantic Web und Linked Data	3
1.2 Problemstellung und Motivation	4
1.3 Forschungsfragen und Zielsetzung	4
1.4 Struktur der Arbeit	5
2 Theoretische Grundlagen	6
2.1 Resource Description Framework (RDF)	6
2.1.1 Uniform Resource Identifier (URI)	8
2.1.2 Exkurs: Blank Nodes	8
2.1.3 Exkurs: RDF-Collection	9
2.2 Serialisierung von RDF	9
2.2.1 Format RDF/XML	10
2.2.2 Format Notation3	10
2.2.3 Exkurs: Format RDFa	11
2.3 Ontologien, Vokabulare und Datenmodelle	12
2.3.1 Der Begriff der Ontologie und Abgrenzung	12
2.3.2 RDF Schema (RDFS) und Web Ontology Language (OWL)	12
2.3.3 Bestehende Vokabulare	14
2.4 Abfragesprache SPARQL	15
2.5 Triple Store Sesame	16
2.6 Die Prinzipien von Linked Data	16
2.7 Bestehende Projekte	17

2.7.1	Linking Open Data (LOD)	17
2.7.2	DBpedia	17
2.8	Fazit	20
3	Geodaten als Linked Data	21
3.1	Bestehende Projekte	22
3.1.1	LinkedGeoData.org	22
3.1.2	GeoNames	22
3.1.3	Administrative Geography des Ordnance Survey (OS)	23
3.1.4	The National Map des USGS	23
3.2	Umformung von Vektordaten	25
3.2.1	Punktdaten mit dem Basic Geo Vocabulary	25
3.2.2	Vektordaten mit GeoRSS	27
3.2.3	Vektordaten beim Ordnance Survey	28
3.2.4	Vektordaten mit Geo OWL	29
3.2.5	Vektordaten mit NeoGeo	30
3.2.6	Vektordaten beim USGS	32
3.3	Umformung von Rasterdaten	33
3.4	GeoSPARQL - Abfragesprache und Format	34
3.5	Diskussion	35
3.6	Fazit (Forschungslücken)	36
4	Methodisches Vorgehen und Daten	38
4.1	Methodisches Vorgehen	38
4.2	Verwendete Datensätze	40
4.2.1	Jagdbanngebiete	40
4.2.2	Exkurs: Geschichte der Jagdbanngebiete	40
4.2.3	Exkurs: Jagdbanngebiete heute	41
4.2.4	Landschaftsmodelle VECTOR25 / VECTOR200	42
4.2.5	Digitale Höhenmodelle DHM25 / DHM RIMINI	44
4.2.6	Biogeographische Regionen	46
5	Ansatz “Vektorisierte Kleinste Einheiten” (VKE)	47
5.1	Konzeptionelle Grundlagen	48
5.2	Vergleich VKE und USGS	49
5.3	Modellgrundlagen VKE	50
5.3.1	Implementierung der räumlichen Beziehungen	50
5.3.2	Datenmodell	50
5.4	Datenprozessierung	52
5.5	Umformung in RDF	57
5.6	Resultate des vereinfachten Use Case	58
5.7	Diskussion und Fazit	61

6	Ansatz “Ausgegliederte Räumliche Operationen” (ARO)	64
6.1	Konzeptionelle Grundlagen	64
6.2	Vergleich VKE und ARO	66
6.3	Modellgrundlagen ARO	67
6.3.1	Modellierung der räumlichen Beziehungen	67
6.3.2	Datenmodell	68
6.4	Umformung der Grundlagendaten	70
6.5	Verknüpfung der Grundlagendaten mit externen Quellen	71
6.6	Diskussion	72
7	Einfache Use Cases	74
7.1	Vorgehen und Resultate	74
7.2	Fazit	77
8	Komplexer Use Case “Waldflächen und Höhenstufen”	78
8.1	Use Case spezifische Grundlagen	78
8.1.1	Höhenstufen	78
8.1.2	Massstab	80
8.2	Datenaufbereitung der Höhenstufen	81
8.3	Datenprozessierung	82
8.4	Resultate	85
8.5	Diskussion	85
9	Komplexer Use Case “Betroffene Gemeinden”	87
9.1	Einbindung von GeoNames	87
9.2	Datenprozessierung	88
9.3	Resultate	90
9.4	Diskussion	91
10	Komplexer Use Case “Ähnliche Gebiete”	93
10.1	Bestimmung der Ähnlichkeit	93
10.2	Datenprozessierung	94
10.3	Resultate	96
10.4	Diskussion	97
11	Schlussfolgerungen und Ausblick	98
11.1	Schlussdiskussion	98
11.2	Erreichtes und Erkenntnisse	100
11.3	Fazit	102
11.4	Ausblick	102
	Literaturverzeichnis	XI

2.1	Die Tripel-Typen als Graphen	8
2.2	Graph, der den Namen eines Mitarbeiters repräsentiert	10
2.3	Das LOD Cloud Diagramm	18
2.4	Suchresultate nach dem Begriff “Matterhorn”	19
2.5	RDF-Repräsentation des Berges “Matterhorn”	19
3.1	LGD Browser Ausschnitt vom Gebiet um den Zoo Zürich	23
3.2	Ausschnitt von Genf aus GeoNames	24
3.3	Ausschnitt des grafischen Resultates der SPARQL-Abfrage aus Quelltext 3.1	25
3.4	Darstellung des NeoGeo Vokabulars	31
4.1	Die 41 Eidgenössischen Jagdbanngebiete der Schweiz	41
4.2	Ausschnitt der Primärflächen aus VECTOR25 und VECTOR200	44
4.3	Digitales Höhenmodell DHM25	45
4.4	Digitales Höhenmodell DHM RIMINI	45
4.5	Die biogeographischen Regionen der Schweiz	46
5.1	Datenmodell des Ansatzes VKE	51
5.2	Primärflächen im Originalzustand	53
5.3	Primärflächen auf das Jagdbanngebiet “Urirotstock” zugeschnitten	53
5.4	Digitales Höhenmodell DHM25 auf die Jagdbanngebiete zugeschnitten und Problematik der Rasterextraktion	54
5.5	Digitales Höhenmodell DHM25 vektorisiert	54
5.6	Union aus Jagdbanngebieten, Primärflächen und vektorisiertem DHM25	55
5.7	Vergrösserte Darstellung der Kleinsten Einheiten	55
5.8	Resultat der SPARQL-Abfrage in m ² nach der Waldfläche im Jagdbanngebiet “Trescolmen”	59
6.1	Schematische Darstellung des Ansatzes ARO	66
6.2	Datenmodell des Ansatzes ARO	69
6.3	Resultate der Suche nach “Jagdbanngebiet” über den Service <sameAs>	72
7.1	Ausschnitt der Ergebnisse der Abfrage aus Quelltext 7.1	75
7.2	Die Teilgebiete des Jagdbanngebietes “Haut de Cry/Derborence”	75
7.3	Das Teilgebiet “JB476966” in detaillierter Ansicht	76
7.4	Ergebnis der Abfrage aus Quelltext 7.2	77

8.1	Zeitmessung von den kleinsten zu den grössten Teilgebieten	81
8.2	Zeitmessung von den grössten zu den kleinsten Teilgebieten	81
8.3	Ablaufdiagramm des komplexen Use Case “Waldfläche und Höhenstufen”	83
8.4	Ergebnisse der zwei Intersect-Operationen	84
8.5	Ursprüngliches digitales Höhenraster und dessen Reklassifizierung gemäss den Höhenstufen	84
9.1	Ablaufdiagramm des komplexen Use Case “Betroffene Gemeinden”	89
10.1	Ablaufdiagramm des komplexen Use Case “Ähnliche Gebiete”	95

3.1	Die räumlichen Beziehungen in GeoSPARQL	35
4.1	Thematische Ebenen von VECTOR25	42
4.2	Mögliche Objektarten der Flächenobjekte von VECTOR25	42
4.3	Thematische Ebenen von VECTOR200	43
4.4	Mögliche Objektarten von VECTOR200	43
5.1	Speicherbedarf der RDF-Daten	62
8.1	Höhenstufenunterteilung	79
8.2	Ergebnis - Waldfläche je Höhenstufe	85
9.1	Resultat der Abfrage nach den Jagdbanngebieten “Kärpf”, “Grand Muveran” und “Piz Ela”	91
10.1	Aufstellung der Resultate der Ähnlichkeitsanalyse mit dem Jagdbanngebiet “Kiental” . . .	96

2.1	Name einer Person mit einem Blank Node	9
2.2	Name einer Person ohne Blank Node	9
2.3	Graph aus Abbildung 2.2 im RDF/XML-Format	10
2.4	Graph aus Abbildung 2.2 im N3-Format	11
2.5	RDFa in einem XHTML Dokument	11
2.6	Beispiel für die Verwendung von RDFS	13
2.7	Beispiel-Tripel für SPARQL-Abfragen	15
2.8	SPARQL-Abfrage.	15
3.1	SPARQL-Abfrage zu The National Map	24
3.2	Punkt im Basic Geo Vocabulary Modell	26
3.3	Punkt bei GeoNames	26
3.4	Punkt bei LinkedGeoData.org	26
3.5	Polygon im GeoRSS Simple Modell	27
3.6	Polygon im GeoRSS GML Modell	28
3.7	Polygon bei LinkedGeoData.org	28
3.8	Polygon bei der Ordnance Survey	28
3.9	Punkt im Geo OWL Modell	30
3.10	Polygon im NeoGeo-Modell	31
3.11	Linie beim USGS	32
3.12	Polygon im GeoSPARQL Modell	35
5.1	Waldobjekt im N3-Format	50
5.2	Kleinste Einheit im Format-N3	56
5.3	Tripel für die Fläche eines Jagdbanngebietes	57
5.4	Mapping für die Namen der Jagdbanngebiete	57
5.5	Benennung eines Jagdbanngebietes	57
5.6	SPARQL-Abfrage für die Waldfläche	58
5.7	SPARQL-Abfrage der Durchschnittshöhe	60
5.8	SPARQL-Abfrage der durchschnittlichen, minimalen und maximalen Höhe des Waldes	60
6.1	Auszug der generierten Tripel der Primärflächen gemäss der Klasse ProJBKomponenten	70
6.2	Auszug der generierten Tripel der Primärflächen gemäss der Klasse Geodaten	70
6.3	Verknüpfung der Klasse Jagdbanngebiet mit DBpedia	71
7.1	SPARQL-Abfrage für den einfachen Use Case “Attribute abfragen”	75
7.2	SPARQL-Abfrage für den einfachen Use Case “Aggregation”	76
8.1	Tripel der subalpinen Stufe (nordexponiert) in den Südalpen	82

9.1	Tripel der Gemeinde “Collonges” der Klasse OrteGN	89
9.2	Abfrage der betroffenen Gemeinden der Jagdbanngebiete “Kärpf”, “Grand Muveran” und “Piz Ela”	90

Das World Wide Web (oder einfach Web) hat enorme Ausmasse angenommen, bietet eine Fülle an Informationen und eröffnet dementsprechend vielfältige Nutzungsmöglichkeiten. Für eine Menge Menschen ist es selbstverständlich Informationen im Internet zu suchen, sei es beispielsweise um herauszufinden, wann der nächste Zug fährt, um Preise von Onlinehändlern zu vergleichen oder um Literatur und Ideen für eine Arbeit zu suchen. Jedoch kann gerade dieses grosse Angebot auch zu Unübersichtlichkeit führen und die effiziente wie auch effektive Informationsverarbeitung erschweren. Daraus kann das Bedürfnis entstehen diese Fülle in eine Ordnung zu bringen. In dieser Arbeit werden mögliche Ansätze aufgezeigt, wie Informationen (insbesondere geographische Informationen) organisiert bereitgestellt werden können.

1.1 Die Idee hinter Semantic Web und Linked Data

1.1.1 Das World Wide Web heute

Der Erfolg des heutigen Web kann unter anderem auf die Vorteile der Informationsbeschaffung über das Web zurückgeführt werden: Aufgrund der kurzen Übertragungszeiten von Sender und Empfänger können die Informationen in ständiger Aktualität bereitgestellt und abgerufen werden. Öffentlich zugängliche Daten sind universell verfügbar (im Gegensatz zu einem Intranet), das heisst, Personen mit einem Internetanschluss haben grundsätzlich Zugang zu diesen Informationen (abgesehen von beispielsweise staatlichen Zensuren), wodurch auch breitere Bevölkerungsschichten erreicht werden. Zudem hat das Web eine Liberalisierung der Bereitstellung von Informationen mit sich gebracht (Hitzler u. a., 2007).

Diese einfache Verfügbarkeit in solch grossem Umfang bringt aber auch Herausforderungen, teilweise gar Nachteile, mit sich. Oft muss eine Selektion vorgenommen werden, um an die relevanten Informationen zu gelangen. Werkzeuge wie beispielsweise Suchmaschinen unterstützen uns dabei, doch kann der personelle Aufwand durch manuelle Priorisierung trotzdem erheblich sein. Zudem besteht die Gefahr, dass wichtige Informationen in der Fülle untergehen und nicht bemerkt werden. Frühere Versionen von Suchmaschinen führten letztendlich eine auf Stichworten basierte Suche durch, verglichen also die angegebenen Zeichenketten mit den Webinhalten (Antoniou u. van Harmelen, 2008; Hitzler u. a., 2007).

Die Repräsentation der Webinhalte richtet sich bisher vor allem nach dem Menschen und weniger nach den Maschinen. So “versteht” die Maschine nicht, was für Informationen in einem Dokument enthalten sind. Eine weitere Problematik ist die Heterogenität der angebotenen Informationen. So werden diese beispielsweise auf Webseiten unterschiedlich repräsentiert oder sind in verschiedenen Sprachen verfasst. Dadurch wird die globale Verwendung und Weiterverarbeitung der angebotenen Informationen erschwert. Zusätzlich wäre es wünschenswert im Web implizit vorhandene Informationen zu extrahieren. Das heisst, aufgrund der vorhandenen Daten lassen sich allenfalls durch Folgerungen neue Informationen generieren

(Hitzler u. a., 2007).

1.1.2 Das Semantic Web

Die grundlegende Idee des Semantic Web besteht in der Aufbereitung der zur Verfügung gestellten Informationen, so dass diese durch Maschinen verarbeitet werden können. Dies soll unter anderem eine effizientere Informationsverarbeitung ermöglichen (Hitzler u. a., 2007). Dazu sind gemäss Antoniou u. van Harmelen (2008) folgende Grundvoraussetzungen nötig:

- Die Daten sollen mit expliziten Metadaten (Informationen zu Struktur und Bedeutung) angereichert werden, wodurch semantische Informationen eingebunden werden. Wird beispielsweise auf einer Webseite das beobachtete Verbreitungsgebiet einer bestimmten Tierart angegeben (inklusive räumlicher Koordinaten), stellt diese Information für die Maschine einfach eine Aneinanderreihung von Zeichen dar. Wird jedoch dieser Zeichenkette eine Bedeutung zugemessen, indem explizit angegeben wird, dass es sich um ein Polygon in einem bestimmten Bezugssystem oder eine Beschreibung der Tierart handelt, “weiss” die Maschine allenfalls wie diese zu interpretieren sind. So könnte mittels des Polygons ein räumlicher Vergleich mit anderen Verbreitungsgebieten gemacht werden.
- Die Definition von Ontologien (Begriffserklärung im Kapitel 2) ermöglicht ein gemeinsames Verständnis eines Interessenbereiches, indem versucht wird Objekte eines Wirklichkeitsausschnittes in Klassen und Eigenschaften einzuteilen und diese miteinander in Beziehung zu setzen. Wird beispielsweise die gleiche Ontologie für Arzneimittel von verschiedenen Pharmaunternehmen verwendet, ist allen Benutzern bekannt, was unter “Schmerzmittel” zu verstehen ist und welche Medikamente zu dieser Gruppe gehören.
- Mit Hilfe der formalen Logik können Methoden zur Schlussfolgerung (*reasoning*) von neuen Informationen aus den bestehenden Informationen entwickelt werden. Damit soll das implizit vorhandene Wissen extrahiert werden (Hitzler u. a., 2007). Ein einfaches Beispiel dafür ist: In der Ontologie wird definiert, dass Auerhühner zur Familie der Fasanenartigen gehören und diese wiederum zur Ordnung der Hühnervögel. Daraus kann die Maschine nun schliessen, dass Auerhühner auch Hühnervögel sind.

Zu diesen Grundvoraussetzungen benennen Hitzler u. a. (2007) zusätzlich die Interoperabilität. Sie bezieht sich auf die Verwendung von einheitlichen und offenen Standards, damit Informationen von jeglichen Anwendungen und Plattformen ausgetauscht werden können (Hitzler u. a., 2007). In diesen Bereich fallen auch die Ontologien, da sie die Interoperabilität fördern können.

Die Organisationen World Wide Web Consortium (W3C)¹ und Open Geospatial Consortium (OGC)² spielen bei der Realisierung der Interoperabilität von Webdiensten eine entscheidende Rolle. Die W3C entwickelt und publiziert Standards für das Web,³ während die OGC Standards für räumliche Komponenten des Webs entwickelt und publiziert.⁴

Das Semantic Web kann auch als eine Erweiterung des heutigen auf Dokumenten (z. B. HTML) basierenden Web (“Web der Dokumente”), zu einem “Web der Daten” verstanden werden (Hitzler u. a., 2007).

¹ <http://www.w3.org/>

² <http://www.opengeospatial.org/>

³ <http://www.w3.org/Consortium/>, 29.12.2011

⁴ <http://www.opengeospatial.org/ogc>, 29.12.2011

1.1.3 Linked Data

Linked Data ist eine fundamentale Grundlage für das Web der Daten (Semantic Web). Heute existieren viele Organisationen und Einzelpersonen, die Daten öffentlich zugänglich machen und/oder weiterverarbeiten. Dazu müssen die Daten möglichst gut strukturiert werden (Metadaten), um eine einfache Wiederverwendung zu ermöglichen. Zudem wird durch die Strukturierung die Interoperabilität gefördert. Die Idee von Linked Data kann analog zum heute vorherrschenden Web der Dokumente betrachtet werden. Der Text in diesen Dokumenten wird hauptsächlich für die Darstellung strukturiert (oft unter Verwendung von HTML). Aber der eigentliche Inhalt (die Informationen im Text) bleibt unstrukturiert. An diesem Punkt setzt Linked Data an, das unter anderem die Strukturierung von Daten zum Ziel hat. Dies ist bereits ansatzweise durch Mikroformate und Programmierschnittstellen (*application programming interface*, kurz API) möglich (Heath u. Bizer, 2011).

Mikroformate ermöglichen die Strukturierung von Daten für spezifische Bereiche, wie Veranstaltungen, Organisationen usw. und deren Einbindung in HTML Dokumente.⁵ Probleme bei den Mikroformaten sind, dass sie kein gemeinsames Datenmodell besitzen, jeweils nur über eine sehr eingeschränkte und fachspezifische Auswahl an Entitäten verfügen und nicht mit einander kombiniert werden können. Zudem können oft Beziehungen zwischen den Entitäten nicht abgebildet werden (Heath u. Bizer, 2011; Hausenblas, 2009).

APIs ermöglichen den Zugang zu strukturierten Daten eines Anbieters über das Hypertext Transfer Protocol (HTTP). Jedoch besteht hier das Problem, dass jeder Anbieter eine etwas andere Programmierschnittstelle zur Verfügung stellt und dadurch kein einheitlicher Zugang möglich wird. Zudem sind die Daten nicht mit fremden Datenquellen verknüpft, sie werden also in einem eher isolierten Kontext angeboten (Heath u. Bizer, 2011).

Neben der Strukturierung von Dokumenten sind Links von einem Dokument zum anderen ein weiterer wichtiger Bestandteil des Webs der Dokumente. Dadurch entsteht ein Netz, durch das man sich durchklicken und zu Dokumenten gelangen kann. Das Gleiche soll bei Linked Data auf der Datenebene realisiert werden. Das heisst, Daten sollen entdeckt, zugänglich gemacht, integriert und verwendet werden können (Heath u. Bizer, 2011).

1.1.4 Unterschied von Semantic Web und Linked Data

Der Unterschied von Linked Data und Semantic Web ist nicht eindeutig und wird von der Community diskutiert. Manche fassen den Begriff Linked Data als eine Neubenennung (*rebranding*) auf, also als moderne Bezeichnung für Semantic Web. Tom Heath⁶ sieht die Begriffe nicht als synonym, sondern eher ergänzend. So umschreibt seiner Meinung nach Linked Data eine der zugrunde liegenden Technologien für das Semantic Web. Anstelle von Semantic Web verwendet Heath auch den Begriff "Web of Data" (Web der Daten), weil dies für viele verständlicher ist und man sich darunter mehr vorstellen kann. Auch im Rahmen dieser Arbeit wird Linked Data als Grundlage für das Web der Daten angesehen. Die Arbeit liegt thematisch im Linked Data Bereich, da die Bereitstellung, Vernetzung und Abfrage der Daten im Vordergrund stehen.

⁵ <http://microformats.org/wiki/introduction>, 29.12.2011

⁶ <http://tomheath.com/blog/2009/03/linked-data-web-of-data-semantic-web-wtf/>, 22.06.2011

1.2 Problemstellung und Motivation

Die zwei Grundideen von Linked Data, Daten zu strukturieren und mit fremden Datenquellen zu verknüpfen, müssen nicht nur spezifisch für das Web gelten, sondern können unter anderem auch auf die Datenhaltung- und -verarbeitung innerhalb von Institutionen übertragen werden. Dies mag besonders sinnvoll erscheinen, wenn grössere Datenbestände vorhanden sind. So könnten diese Daten in eine einheitliche Ontologie integriert werden und gegenseitige Verknüpfungen würden ein Web der Daten innerhalb der Institution schaffen. Dies bringt folgende Vorteile mit sich:

- Die Daten der Institution könnten vereinfacht mit fremden Datenquellen über das Web verknüpft werden. Dies würde einerseits die öffentliche Zugänglichkeit dieser Daten unterstützen (falls beabsichtigt) und es könnten neue potentiell interessante Daten entdeckt werden.
- Die Verwendung einer gemeinsamen Ontologie vereinfacht das Verständnis der verschiedenen Datensätze und unterstützt den Austausch innerhalb der Institution.
- Das Durchführen von Folgerungen (*reasoning*) könnte in den Datensätzen implizit vorhandene Informationen zu Tage fördern.

Eine Institution, die grössere Datenbestände unterhält, ist die Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft⁷ (WSL). Das Datenzentrum Natur und Landschaft (DNL) der Forschungsanstalt unterhält hauptsächlich ein relationales Datenbanksystem mit heterogenen Daten, bestehend aus Dokumenten, Sachdaten, räumlichen Daten (inklusive attributiven und geometrischen Daten) sowie vielen weiteren Datenarten. Die Informationsgewinnung auf diesen Datenbeständen erfolgt derzeit mittels Abfragen in relationalen Datenbanken und auf Ontologien basiertem Suchen. Zu diesen bestehenden Abfrage-Paradigmen soll nun Linked Data hinzugefügt werden, indem erste Erfahrungen in der Datenhaltung und -verarbeitung anhand eines kleineren Datensatzes gewonnen werden sollen. Dieser Geodatensatz umfasst die Jagdbanngebiete der Schweiz, weshalb aufgrund des Raumbezuges die Verwendung von räumlichen Daten mit Linked Data im Fokus der Arbeit steht.

1.3 Forschungsfragen und Zielsetzung

In dieser Arbeit wurden aufgrund des explorativen Vorgehens keine Forschungsfragen im klassischen Sinne definiert, sondern praktische Anwendungsfälle (*Use Cases*) geben die Leitplanken für die Arbeit vor. Darüber hinaus werden im Rahmen dieser Anwendungsfälle auch allgemeine, nicht fallspezifische Fragen verfolgt und beantwortet, die als **Forschungsfragen** gelten können:

- Welche bestehenden Ansätze können verwendet werden, um räumliche Daten in Linked Data zu integrieren?
- Was für Möglichkeiten bieten sich an, um Rasterdaten in Linked Data zu integrieren und wie lässt sich dies praktisch umsetzen?
- Wie lassen sich räumliche Daten sinnvoll mit anderen Linked Data Datensätzen verknüpfen, um einen Beitrag zum Web der Daten zu leisten?
- Was ist der mögliche Nutzen für potentielle Benutzer mit Linked Data zu arbeiten und im Speziellen, was ist der Nutzen räumliche Daten im Kontext von Linked Data zu verwenden?

⁷ <http://www.wsl.ch/>

Auf diesen allgemeinen Forschungsfragen und den Anwendungsfällen basieren die **Zielsetzungen** dieser Arbeit:

- Diskussion der bestehenden Ansätze, um räumliche Daten in den Linked Data Kontext zu integrieren.
- Entwicklung möglicher Ansätze zum Umgang mit Rasterdaten im Kontext von Linked Data.
- Prüfen der verschiedenen Ansätze durch Implementierung praktischer Anwendungsfälle (*proof of concept*).

1.4 Struktur der Arbeit

Der Aufbau der Arbeit richtet sich stark nach deren explorativen Charakter und der Entwicklung von Ansätzen. Sie zeigt dementsprechend die vollführten Arbeitsschritte auf. Zu Beginn wird eine Einführung in die theoretischen Grundlagen von Linked Data gegeben, wozu wichtige Konzepte, Standards und Technologien zählen (Kapitel 2). Zudem wird ein kurzer Einblick in zwei bekannte Projekte aus dem Linked Data Bereich gegeben. Danach werden spezifischere Grundlagen bezüglich räumlicher Daten im Linked Data Kontext besprochen (Kapitel 3). Dabei sind bestehende Ansätze im Umgang mit räumlichen Daten, sowie die mit räumlichen Daten verbundenen Schwierigkeiten im Zentrum, zu denen besonders Rasterdaten zählen. Hier werden einige Projekte vorgestellt, die räumliche Daten in den Linked Data Kontext integriert haben. Im Kapitel 4 wird ein kurzer Überblick zum methodischen Vorgehen und zu den verwendeten Datensätzen gegeben. In den Kapiteln 5 und 6 werden die im Rahmen dieser Arbeit entwickelten Ansätze vorgestellt. Die darauffolgenden Kapitel beschäftigen sich mit der Implementierung von praktischen Anwendungsfällen (Use Cases), die sich auf die vorgängig beschriebenen Ansätze stützen. Das Kapitel 11 diskutiert die Ergebnisse der Arbeit, stellt Schlussfolgerungen an und gibt einen Ausblick.

Theoretische Grundlagen

Zur Realisierung von Linked Data bzw. des Semantic Webs werden einige Konzepte und Technologien benötigt, die teilweise spezifisch für diesen Zweck entwickelt wurden. In diesem Kapitel wird ein kurzer Überblick zu den wichtigsten verwendeten Technologien gegeben. Dazu zählen die theoretischen Grundlagen, wie auch deren möglicher Einsatz in der Praxis. Zum Schluss werden zwei wichtige Projekte vorgestellt, die im Linked Data Bereich eine zentrale Rolle spielen.

2.1 Resource Description Framework (RDF)

Resource Description Framework (RDF) ist ein vom W3C¹ (World Wide Web Consortium) standardisiertes Datenmodell, um strukturierte Informationen zu beschreiben. Obwohl RDF streng genommen ein Datenmodell ist, wird RDF oft auch als formale Sprache bezeichnet, was auch für diese Arbeit zutrifft (Antoniou u. van Harmelen, 2008; Hitzler u. a., 2007). Ursprünglich war RDF dazu gedacht Metadaten von Webressourcen zu repräsentieren. Da jedoch heute unter Ressourcen nicht nur webbasierte Dinge verstanden werden, wird RDF nun auch für die Repräsentation von Informationen unterschiedlichen Ursprungs verwendet (Manola u. a., 2004). Zudem soll RDF imstande sein zwischen verschiedenen im Web verwendeten Datenmodellen zu vermitteln (Heath u. Bizer, 2011). Dadurch können Daten zwischen Anwendungen ohne Informationsverlust ausgetauscht werden (Hitzler u. a., 2007).

Die Informationen werden in RDF als gerichtete Graphen² (Menge von Knoten, die durch gerichtete Kanten verbunden werden) repräsentiert, wobei Knoten und Kanten eindeutig durch sogenannte Uniform Resource Identifiers (URIs) identifiziert und benannt werden (Blank Nodes und Literale sind davon ausgenommen) (Hitzler u. a., 2007; Heath u. Bizer, 2011). Drei fundamentale Konzepte liegen RDF zugrunde (Antoniou u. van Harmelen, 2008):

Ressourcen (*resources*) bilden einen zentralen Baustein. Sie repräsentieren die im Interesse liegenden Objekte. Das bedeutet, Ressourcen können sich auf jegliche denkbare oder real existierende Objekte (z. B. ein Auto, eine Webseite, eine mathematische Gleichung) beziehen. Jede Ressource wird durch einen URI (Uniform Resource Identifier) eindeutig identifiziert und benannt (Antoniou u. van Harmelen, 2008; Manola u. a., 2004).

Eigenschaften (*properties*) beschreiben die Beziehungen zwischen den Ressourcen und stellen somit eine spezielle Art von Ressourcen dar, die ebenfalls eindeutig durch einen URI identifiziert werden (Antoniou u. van Harmelen, 2008).

Aussagen (*statements*) beschreiben die Ressourcen in Form von Tripeln (*triples*). Ein Tripel setzt sich immer aus einem Subjekt, Prädikat und Objekt zusammen und weist somit die Struktur eines einfachen

¹ <http://www.w3.org/TR/tr-technology-stds>

² Wenn nicht anders vermerkt, handelt es sich um endliche Graphen. Endliche Graphen haben eine endliche Knotenmenge im Gegensatz zu unendlichen Graphen, die eine unendliche Knotenmenge haben (Diestel, 2006).

Satzes auf (Antoniou u. van Harmelen, 2008). Das Subjekt ist der URI der Ressource (auch Blank Node möglich), welche beschrieben werden soll. Das Objekt kann ein Literal (z. B. Zahl, String) oder der URI einer anderen Ressource (auch Blank Node möglich) sein. Das Prädikat definiert die Beziehung zwischen dem Subjekt und Objekt und wird ebenfalls durch einen URI identifiziert. Ein Tripel entspricht einem einfachen RDF-Graphen. Prinzipiell lassen sich zwei verschiedene Tripel-Typen unterscheiden (Heath u. Bizer, 2011):

- **Literal-Tripel:** Diese haben einen Literal als Objekt. Literale repräsentieren in RDF Datenwerte, die einem bestimmten Datentyp zugeordnet werden können. Der Wert eines Literals wird durch eine Zeichenkette wiedergegeben, wobei der angegebene Datentyp die Interpretation dieses Wertes festlegt (Hitzler u. a., 2007).

Literale ohne Datentyp werden wie Zeichenketten (String) interpretiert (Hitzler u. a., 2007). Sie können zusätzlich mit einem Bezeichner für eine natürliche Sprache, wie Deutsch oder Englisch, kombiniert werden (Heath u. Bizer, 2011). Bei Literalen mit einem Datentyp wird dieser durch einen URI identifiziert. Er entspricht üblicherweise einem XML-Schema Datentypen³ (z. B. Integer oder Date). Literale können nicht als Subjekte oder Prädikate verwendet werden (Heath u. Bizer, 2011; Hitzler u. a., 2007; Manola u. a., 2004). Literale haben also einen beschreibenden Charakter.

- **RDF-Verknüpfungen (*RDF Links*):** Dieser Typ beschreibt die Beziehung zwischen zwei Ressourcen (keine Literale) und setzt sich somit aus drei URIs zusammen (jeweils für Subjekt, Prädikat und Objekt). Der URI des Prädikates definiert den Beziehungstyp zwischen den Ressourcen.

Hilfreich ist die Unterscheidung zwischen internen und externen RDF-Verknüpfungen. Interne verbinden Ressourcen innerhalb derselben Datenquelle, während externe eine Verbindung zu fremden Datenquellen herstellen. Das heisst, die Ressourcen von Subjekt und Objekt befinden sich in unterschiedlichen Datenbeständen. Besonders diese externen Verknüpfungen sind ein Grundstein für Linked Data und somit auch Voraussetzung für die Schaffung eines Web der Daten. Durch diese Verknüpfungen werden nicht nur die betroffenen Datenquellen verbunden, sondern auch alle anderen, die mit diesen Ressourcen an anderer Stelle verknüpft sind (Heath u. Bizer, 2011).

Die Abbildung 2.1 zeigt drei Tripel als Graphen, wovon zwei Literal-Tripel sind und einer eine RDF-Verknüpfung darstellt. Ein Tripel identifiziert eine oder einen Mitarbeiter/in mit dem URI `http://beispiel.org/MitarbeiterID/1234` (Subjekt), welche/r in der Abteilung mit dem URI `http://beispiel.org/Abteilung/09` (Objekt) arbeitet. Das Prädikat wird durch den URI `http://beispiel.org/arbeitetIn` repräsentiert und beschreibt die Beziehung zwischen diesen beiden Ressourcen. Das Tripel ist vom Typ her eine RDF-Verknüpfung, da Ressourcen miteinander referenziert werden.

Diese/r Mitarbeiter/in (Subjekt) hat einen Namen (`http://beispiel.org/Name`), der **Erika Musterfrau** lautet (Literal). Hierbei handelt es sich um ein Literal-Tripel ohne Angabe eines Datentypes.

Das Subjekt `http://beispiel.org/Abteilung/09` des letzten Tripel (in einem anderen Tripel war es das Objekt) steht für eine Abteilung, die eine bestimmte Anzahl Angestellte `http://beispiel.org/anzahlAngestellte` hat. Sie beträgt zwölf (`'12'`³`http://www.w3.org/XMLSchema#integer`). Diesem Literal-Tripel ist ein XML-Schema Datentyp angefügt, in dem die eigentliche Anzahl durch `http://www.w3.org/XMLSchema#integer` ergänzt wird. Dadurch wird festgelegt, dass es sich bei 12 um einen Integer handelt und dieser entsprechend zu interpretieren ist.

³ <http://www.w3.org/TR/xmlschema-2/>



Abbildung 2.1: Die Tripel-Typen “Literal-Tripel” und “RDF-Verknüpfung” als Graphen dargestellt (verändert nach: Hitzler, 2007, S. 39).

2.1.1 Uniform Resource Identifier (URI)

Uniform Resource Identifiers (URIs) dienen als eindeutige Bezeichner für Ressourcen, z. B. um Webseiten oder E-Mail-Adressen zu identifizieren. Sie müssen sich aber nicht zwangsweise auf das Web beziehen, sondern könnten auch als generelle Bezeichner eingesetzt werden. Ein URI besteht aus mehreren Teilen, wobei zu Beginn ein URI-Schema angegeben wird, das den URI-Typ festlegt. Im Web häufig verwendete Schema sind beispielsweise “http” und “mailto” (Hitzler u. a., 2007). Nach dem Schema folgt die Angabe von Server (*authority*), Pfad (*path*), Anfrage (*query*) und Fragment (*fragment*), wobei davon nur der Pfad (neben dem Schema) zwingend vorhanden sein muss (Hitzler u. a., 2007).⁴ Ein HTTP URI für ein PDF-Dokument, das die Jagdbannggebiete der Schweiz beschreibt, könnte folgendermassen aussehen:

`http://www.beispiel.org/Jagdbannggebiete/Beschreibung_JB.pdf`

Somit ist `http` das Schema, `www.beispiel.org` der Server und `Jagdbannggebiete/Beschreibung_JB.pdf` der Pfad.

Da URIs generell als eindeutige Bezeichner verwendet werden können, müssen diese nicht zwingend über das Web abrufbar sein (Heath u. Bizer, 2011; Hitzler u. a., 2007). In Bezug auf Linked Data dient ein URI nicht nur der Identifizierung von Ressourcen, sondern sie sollte auch dereferenzierbar sein, d. h. über HTTP aufgelöst werden können und gleichzeitig eine Beschreibung der Ressource liefern. Die Verwendung von weltweit einheitlich gestalteten URIs bringt einige Vorteile: Sie können dezentralisiert von den jeweiligen Datenanbietern erstellt werden, sofern diese Anbieter über den benutzten Domain-Namen verfügen (Heath u. Bizer, 2011). Über einen URI identifizierte Ressourcen können einfach referenziert, verknüpft und abgerufen werden (Shadbolt u. a., 2006). URIs dienen der Benennung von Subjekt, Prädikat und Objekt (Manola u. a., 2004).

2.1.2 Exkurs: Blank Nodes

Ein Blank Node ist eine Ressource, die über keinen URI verfügt und deshalb als anonyme Ressource bezeichnet wird. Sie werden oft im Zusammenhang mit Aggregationen (zusammengehörigen Ressourcen) verwendet, wobei bedacht werden muss, dass anonyme Ressourcen von aussen nicht direkt angesprochen werden können (da sie über keinen URI verfügen). Trotzdem müssen Blank Nodes identifiziert und differenziert werden, um sie innerhalb von Tripel repräsentieren zu können. Sie werden durch `_:irgendEinName` identifiziert. Blank Nodes können nicht als Prädikate verwendet werden (Manola u. a., 2004). Beispielsweise kann der Name einer Person durch zwei Tripel beschrieben werden, wie im Quelltext 2.1 zu sehen ist. Hier wird eine anonyme Ressource (`_:NameVon2143`) hinzugezogen, der die eigentlichen Tripel mit den Informationen angehängt wird (Zeilen 5 und 6). Damit wird die Zusammengehörigkeit dieser Werte aus-

⁴ <https://tools.ietf.org/html/rfc3986>, 24.01.2012

gedrückt. Natürlich könnten die Tripel ohne den Blank Node direkt geschrieben werden, wobei dann die Kennzeichnung der Zugehörigkeit verloren geht.

```

1 @prefix vok:<http://beispiel.org/> .
2 @prefix mitID:<http://beispiel.org/MitarbeiterID/> .
3
4 mitID:2143 vok:Name _:NameVon2143 .
5 _:NameVon2143 vok:Vorname "Maria" .
6 _:NameVon2143 vok:Nachname "Feiertag" .

```

Quelltext 2.1: Name einer Person mit einem Blank Node.

Der Quelltext 2.2 zeigt die gleichen Tripel, aber ohne Blank Node. Dadurch wird ersichtlich, dass eine zusätzliche Ressource mit einem URI (mitName:ID2143) eingeführt werden muss, was durch die Verwendung von Blank Nodes umgangen werden kann. Dies macht besonders Sinn, wenn der “Platzhalter”, welcher die Zugehörigkeit angibt, gar nicht direkt angesprochen werden soll.

```

1 @prefix vok:<http://beispiel.org/> .
2 @prefix mitID:<http://beispiel.org/MitarbeiterID/> .
3 @prefix mitName:<http://beispiel.org/MitarbeiterName/> .
4
5 mitID:2143 vok:Name mitName:ID2143 .
6 mitName:ID2143 vok:Vorname "Maria" .
7 mitName:ID2143 vok:Nachname "Feiertag" .

```

Quelltext 2.2: Name einer Person ohne Blank Node.

2.1.3 Exkurs: RDF-Collection

Die RDF-Collection beschreibt eine Gruppe von Dingen, die in einer Listenstruktur in RDF repräsentiert werden, wie beispielsweise eine Fussballmannschaft, die aus mehreren Spielern besteht (Manola u. a., 2004). Die RDF-Collection ist eine abgeschlossene Liste, d. h. sämtliche Fussballspieler wären in dieser Liste vorhanden (Manola u. a., 2004). Wohingegen der sogenannte RDF-Container eine offene Liste darstellt und daher potentielle Mitglieder dieser Gruppe nicht unbedingt enthalten sind. Jedoch muss beachtet werden, dass diese Listen lediglich Kurzschreibweisen für RDF-Graphen sind, d. h. anstatt jedes Tripel einzeln niederzuschreiben, können diese in einer Liste zusammengefasst werden (Hitzler u. a., 2007).

2.2 Serialisierung von RDF

Serialisierung bezeichnet die Repräsentation der Tripel gemäss einer bestimmten Syntax und deren Niederschreiben in einer sequentiellen Folge. Vom W3C wurden zwei RDF (Serialisierungs-)Formate standardisiert – RDF/XML⁵ und RDFa⁶. Daneben existieren einige nicht-standardisierte Formate (z. B. N3, Turtle), die aber ebenfalls breite Verwendung finden (Heath u. Bizer, 2011).

⁵ <http://www.w3.org/TR/rdf-syntax-grammar/>

⁶ <http://www.w3.org/TR/rdfa-syntax/>

2.2.1 Format RDF/XML

Das Format RDF/XML basiert auf der XML Syntax und dient der Repräsentation sowie dem Austausch von RDF-Daten (Hitzler u. a., 2007). Der Quelltext 2.3 stellt den Graphen aus Abbildung 2.2 im RDF/XML-Format dar.

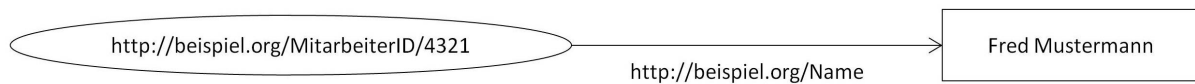


Abbildung 2.2: Graph, der den Namen eines Mitarbeiters repräsentiert.

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3   xmlns:vok="http://beispiel.org/">
4
5   <rdf:Description rdf:about="http://beispiel.org/MitarbeiterID/4321">
6     <vok:Name>Fred Mustermann</vok:Name>
7   </rdf:Description>
8
9 </rdf:RDF>
  
```

Quelltext 2.3: Graph aus Abbildung 2.2 im RDF/XML-Format (verändert nach: Manola u. a., 2004, Kapitel 3.1).

Die erste Zeile deklariert den Folgenden Inhalt als XML, inklusive Versionshinweis. Die zweite Zeile definiert mit `rdf:RDF`, dass es sich im folgenden um RDF handelt. Zudem werden in der zweiten und dritten Zeile Namensräume (*namespaces*) festgelegt. Namensräume ermöglichen eine abgekürzte Schreibweise, indem für einen URI ein Präfix definiert wird, der anstelle des URI verwendet werden kann. So steht der Präfix `vok` (Zeile 3) für den URI `http://beispiel.org/`. Wird nun `vok:Name` geschrieben ist dies äquivalent zu `http://beispiel.org/Name`. Dadurch lässt sich die Schreibweise nicht nur abkürzen, sondern es kann auf externe Namensräume zugegriffen werden (hinter welchen ein RDF-Dokument steht), wie dies beispielsweise bei `rdf:RDF` der Fall ist (Antoniou u. van Harmelen, 2008).

Die Zeilen 5 bis 7 repräsentieren den Graphen aus Abbildung 2.2. Durch `rdf:Description` wird gekennzeichnet, dass eine Beschreibung einer RDF-Ressource folgt, die durch `rdf:about` benannt wird (entspricht dem Subjekt). Nachfolgend werden in Zeile 6 das Prädikat `vok:Name` und Objekt `Fred Mustermann` (Literal ohne Datentyp) angegeben (Heath u. Bizer, 2011).

Das RDF/XML-Format ist wenig benutzerfreundlich für Menschen, da es nicht einfach zu lesen und zu schreiben ist (Heath u. Bizer, 2011).

2.2.2 Format Notation3

Das Format Notation3 (auch als N3 bezeichnet) ist ein kompaktes, für Menschen einfach lesbares bzw. schreibbares Format, das zusätzliche Ausdrucksmöglichkeiten gegenüber RDF/XML bietet.⁷ Die Notation folgt dem Muster eines Tripel (Subjekt, Prädikat und Objekt).⁸ Zusätzlich existieren Unterformate, die in ihrer Funktionalität gegenüber Notation3 limitiert sind: Dazu gehören Turtle und N-Triples.⁹ Im Rahmen

⁷ <http://www.w3.org/TeamSubmission/n3/>, 23.12.2011

⁸ <http://www.w3.org/2000/10/swap/Primer>, 27.12.2011

⁹ <http://www.w3.org/TeamSubmission/n3/>, 23.12.2011

dieser Arbeit wird mehrheitlich auf das N3-Format zurückgegriffen. Normalerweise wird für ein Tripel je eine Zeile verwendet. Es kann jedoch aus Platzgründen vorkommen, dass sie zu lang sind und deshalb untereinander mit Einrückung geschrieben werden.

Der Quelltext 2.4 zeigt den Graphen aus der Abbildung 2.2 im N3-Format.

```
1 @prefix vok:<http://beispiel.org/> .
2 @prefix mitID:<http://beispiel.org/MitarbeiterID/> .
3
4 mitID:4321 vok:Name "Fred Mustermann" .
```

Quelltext 2.4: Graph aus Abbildung 2.2 im N3-Format.

In den Zeilen 1 und 2 werden die Namensräume definiert. Danach wird das Tripel gemäss dem Muster Subjekt `mitID:4321`, Prädikat `vok:Name` und Objekt `"Fred Mustermann"` formuliert. Der Punkt am Schluss signalisiert, dass dies ein Tripel ist.

2.2.3 Exkurs: Format RDFa

Das Serialisierungsformat RDFa erlaubt die Einbindung von RDF in XHTML Dokumente. Bereits HTML (Hypertext Markup Language) ermöglicht eine minimale Strukturierung von Webseiten und Daten. Jedoch beschränkt sich diese vor allem auf Textinhalte und weniger auf die Daten selbst (Heath u. Bizer, 2011). Von der Grundidee her ist RDFa ähnlich wie die Mikroformate. Es erlaubt jedoch die Einbettung von beliebigen RDF-Graphen in XHTML, weshalb es als eine vollwertige Semantic Web Technologie angesehen werden kann. RDFa basiert auf dem einheitlichen Datenmodell RDF und ermöglicht die Kombination von verschiedenen Vokabularen (Hausenblas, 2009). Das Beispiel im Quelltext 2.5 zeigt die Einbettung von RDFa in ein XHTML Dokument.

```
1 div typeof="foaf:Person" xmlns:foaf="http://xmlns.com/foaf/0.1/">
2   <p property="foaf:name">
3     Alice Birpemschwick
4   </p>
5 </div>
```

Quelltext 2.5: RDFa in einem XHTML Dokument (verändert nach: <http://www.w3.org/TR/xhtml-rdfa-primer/>, Kapitel 3.1, 16.08.2011).

Die direkte Einbindung in die XHTML-Tags wird ersichtlich. So ist der ursprüngliche Tag `<p>` durch `<p property="foaf:name">` ergänzt worden (Zeile 3). Das externe Vokabular Friend-of-a-Friend (FOAF) wird miteinbezogen. Folgendes Tripel lässt sich aus dem Beispiel ableiten:¹⁰

```
:blanknode1          foaf:name          'Alice Birpemschwick'.
```

Der wesentliche Unterschied zu RDF besteht darin, dass RDFa dem Ausdruck von RDF innerhalb von XHTML dient und somit die Daten auch in für Menschen lesbarer Form dargestellt werden können. Wohingegen RDF zur effektiven Datenrepräsentation in maschinenlesbarer Form verwendet wird (Hausenblas, 2009; Heath u. Bizer, 2011).

¹⁰<http://www.w3.org/TR/xhtml-rdfa-primer/>, 16.08.2011

2.3 Ontologien, Vokabulare und Datenmodelle

2.3.1 Der Begriff der Ontologie und Abgrenzung

Der vielseitig verwendete Begriff Ontologie bezieht sich auf einen bestimmten Wirklichkeitsausschnitt (*domain of interest*), welcher präzise durch kurze Statements beschrieben wird (Hitzler u. a., 2009). Eine Ontologie stellt also ein Abbild eines Wirklichkeitsausschnittes dar, indem unter anderem versucht wird, die existierenden Objekte in Klassen und Eigenschaften einzuteilen, sowie diese miteinander in Beziehung zu setzen. Die Ontologie legt klar fest, wie ein Wirklichkeitsausschnitt abgebildet wird und aus welchen Begriffen er sich zusammensetzt (Antoniou u. van Harmelen, 2008). Typischerweise gehören auch eine hierarchische Gliederung und die optionale Festlegung von Wertebeschränkungen und disjunkten Elementen dazu (z. B. eine Herde Gämsen muss mindestens 5 Individuen zählen) (Antoniou u. van Harmelen, 2008). In der Regel besteht eine Ontologie aus einer definierten Liste von Klassen und Beziehungen zwischen den Klassen (Antoniou u. van Harmelen, 2008).

Ontologien dienen vor allem dem gemeinsamen Verständnis einer Domäne (Wirklichkeitsausschnitt). Dazu müssen jeweils die gleichen Ontologien verwendet oder Übersetzungen zwischen unterschiedlichen Ontologien bereitgestellt werden (Antoniou u. van Harmelen, 2008). Eine solche Übersetzung kann beispielsweise klarstellen, dass die Klasse “Angestellte” das Gleiche repräsentiert wie die Klasse “Beschäftigte” einer anderen Ontologie. Somit können Applikationen unterschiedliche Datenquellen benutzen. Die Interoperabilität wird gefördert. Zusätzlich führt dies zu einer Verbesserung von Abfragen über das Web, da gezielt nach Ressourcen gesucht werden kann, die entsprechend ontologisch eingeordnet sind (Antoniou u. van Harmelen, 2008).

Die Unterscheidung der Begriffe “Vokabular” und “Ontologie” ist nicht eindeutig. Allgemein wird für eher komplexe Abbilder der Begriff Ontologie verwendet, während für einfachere Sachverhalte der Begriff Vokabular gebräuchlich ist. Vokabulare erfüllen vor allem zwei Zwecke: Zum einen sollen sie bei der Integration von Daten aus verschiedenen Quellen helfen, indem die verwendeten Begriffe einander gegenübergestellt und in Beziehung gesetzt werden. Zum anderen kann ein Vokabular zur Organisation von Daten (Wissen) eingesetzt werden, um die Vorteile von Linked Data nutzen zu können.¹¹ Im Rahmen dieser Arbeit werden beide Begriffe synonym verwendet, zusätzlich kommt der Begriff des Datenmodells hinzu. Ein Datenmodell ist ebenfalls ein abstrahiertes Abbild eines Wirklichkeitsausschnittes und beschreibt die grundsätzlichen Strukturen, Beziehungen und möglichen Eigenschaften eines solchen Abbildes (Bill u. Zehner, 2001). Der Begriff Datenmodell wird in dieser Arbeit verwendet, wenn das Abbild nicht ausschliesslich durch eine Ontologiesprache (wie RDFS) definiert wird (z. B. UML ähnliche Darstellungen).

2.3.2 RDF Schema (RDFS) und Web Ontology Language (OWL)

RDF ermöglicht die Beschreibung von Ressourcen mittels Tripel, jedoch sind damit abstraktere Beschreibungen auf einer übergeordneten Ebene, wie die Semantik einer Domäne, nicht möglich. Dazu müssen Ontologien (bzw. Vokabulare) hinzugezogen werden (Heath u. Bizer, 2011).

RDF Schema (RDFS) ermöglicht die Festlegung von einfachen Ontologien (*lightweight ontologies*), die oft auch als Vokabulare bezeichnet werden (Heath u. Bizer, 2011). Die Sprache RDFS wurde im Jahre 2004 vom W3C zu einem Standard erklärt (Shadbolt u. a., 2006). Die grundlegenden Elemente eines RDFS-Vokabulars sind die Definition von Klassen (*class*) und Eigenschaften (*property*). Das heisst, die mit RDF

¹¹<http://www.w3.org/standards/semanticweb/ontology>, 21.06.2011

beschriebenen Ressourcen und deren Prädikate werden den Elementen (Primitiven) eines RDFS Vokabulars zugeordnet (Heath u. Bizer, 2011). Eine Klasse umfasst bestimmte Typen von Instanzen, wobei die Instanzen den RDF-Ressourcen entsprechen. So können beispielsweise Ressourcen, welche die einzelnen Mitarbeiter einer Firma repräsentieren zur Klasse “Angestellte” zusammengefasst werden (Antoniou u. van Harmelen, 2008).

Das Beispiel im Quelltext 2.6 (N3-Format) veranschaulicht die Verwendung von RDFS in Bezug auf den Graphen in Abbildung 2.2. In den ersten beiden Zeilen werden die Namensräume für das `rdf` und `rdfs` Vokabular definiert. Aus historischen Gründen teilen sich die heutigen Primitiven von RDFS auf diese beiden Vokabulare auf (Heath u. Bizer, 2011). In den Zeilen 3 und 4 folgen die bereits aus Quelltext 2.4 bekannten Namensräume. Das Tripel in Zeile 6 bestimmt über `rdf:type`, dass `vok:Angestellter` eine `rdfs:Class` ist. Das heisst der URI `vok:Angestellter` identifiziert eine Klasse, der später einzelne Instanzen zugeordnet werden können. In Zeile 9 ordnet das Tripel nun die konkrete Instanz `mitID:4321`, welche eine Mitarbeiterin repräsentiert, der Klasse `vok:Angestellter` zu. In der Zeile 7 wird das Attribut `vok:Name` explizit in RDFS als Eigenschaft (`rdf:Property`) definiert. Ausserdem fällt auf, dass die Primitive `rdf:Property` aus dem RDF-Vokabular stammt, aber eigentlich zu RDFS gehört.

```

1 @prefix rdf:<http://www.w3.org/1999/01/rdf-schema#> .
2 @prefix rdfs:<http://www.w3.org/2000/02/22-rdf-syntax-ns#> .
3 @prefix vok:<http://beispiel.org/> .
4 @prefix mitID:<http://beispiel.org/MitarbeiterID/> .
5
6 vok:Angestellter rdf:type rdfs:Class .
7 vok:Name rdf:type rdf:Property .
8
9 mitID:4321 rdf:type vok:Angestellter .
10 mitID:4321 vok:Name "Gertrude Langhans" .

```

Quelltext 2.6: Beispiel für die Verwendung von RDFS.

Ressourcen können zusätzlich für Menschen lesbar beschrieben werden, indem die RDFS-Eigenschaften für eine Bezeichnung (`rdfs:label`) und eine Beschreibung (`rdfs:comment`) verwendet werden. Dadurch können potentielle Benutzer das definierte Vokabular besser verstehen (Heath u. Bizer, 2011). Um die Beziehungen unter den Klassen und Eigenschaften genauer zu beschreiben, kann auf weitere RDFS-Primitiven zurückgegriffen werden. Sie erlauben die Definition von einfachen Regeln (Heath u. Bizer, 2011):

- Durch `rdfs:subClassOf` kann die gleichzeitige Zugehörigkeit von Instanzen einer Klasse zu einer oder mehreren anderen Klassen ausgedrückt werden (Heath u. Bizer, 2011). Es entsteht eine hierarchische Gliederung. So kann beispielsweise ein Mountainbike der Klasse “Fahrräder” zugeordnet werden und die Klasse “Fahrräder” kann eine Subklasse von “Fahrzeugen” sein, womit das Mountainbike ebenfalls zur Klasse der “Fahrzeuge” gehört.
- Mit `rdfs:subPropertyOf` können Eigenschaften hierarchisch geordnet werden. Dadurch wird definiert, dass Ressourcen mit einer Eigenschaft auch mit einer weiteren Eigenschaft verknüpft sind (Heath u. Bizer, 2011; Manola u. a., 2004). Beispielsweise eine Ressource, die eine Person repräsentiert, kann die Eigenschaft “Projektleiter” haben, wobei diese eine Sub-Eigenschaft von “Mitarbeiter” ist. Dadurch würde diese Ressource durch die Eigenschaften Projektleiter und Mitarbeiter beschrieben werden.

- Mit `rdfs:domain` wird der Definitionsbereich eingeschränkt (Hitzler u. a., 2007; Manola u. a., 2004), indem eine Eigenschaft einer oder mehreren Klassen zugeordnet wird (Heath u. Bizer, 2011). So hat beispielsweise die Eigenschaft “hat Mitarbeiternummer” die Klasse “Angestellter” als Domäne. Dadurch wird jede Ressource, die über diese Eigenschaft verfügt, automatisch auch der Klasse “Angestellter” zugeordnet.
- Mit `rdfs:range` wird der Wertebereich eingeschränkt (Hitzler u. a., 2007), indem die Werte einer Eigenschaft Instanzen einer oder mehrerer Klassen sind (Heath u. Bizer, 2011). So könnte beispielsweise die Eigenschaft “ist Projektleiterin” durch den Wertebereich der Klasse “Angestellter” begrenzt sein. Dadurch können nur Instanzen aus der Klasse “Angestellter” als Werte für diese Eigenschaft verwendet werden.

Komplexere Sachverhalte können mit RDFS nicht abgebildet werden, weshalb auf die Web Ontology Language (OWL) zurückgegriffen werden kann (Hitzler u. a., 2007). Die Erstellung einer Ontologie oder eines Vokabulars bildet ebenfalls eine zentrale Funktion, wozu die Möglichkeit spezifische Informationen zu bestimmten Objekten abzulegen kommt.¹² Dies reicht von Kardinalitäten über disjunkte Klassen, bis hin zu speziellen Bedingungen einzelner Klassen oder Eigenschaften (Antoniou u. van Harmelen, 2008). Im Rahmen dieser Arbeit wird OWL nicht verwendet, mit Ausnahme der Primitive `owl:sameAs`. Sie ermöglicht die Verknüpfung von identischen Instanzen, z. B. aus verschiedenen Datenquellen (Smith u. a., 2004).

Ein Vokabular ermöglicht also die Definition von Konzepten und Beziehungen, indem Klassen gebildet, mögliche Eigenschaften und Beziehungen beschrieben, sowie Beschränkungen festgelegt werden.¹³ Einzelne Ressourcen (Instanzen) können den Klassen zugeordnet werden und der Aufbau einer Hierarchie ist möglich (Manola u. a., 2004).

2.3.3 Bestehende Vokabulare

Die gemeinsame Verwendung von einheitlichen Vokabularen ist wünschenswert, da dadurch keine Übersetzungen zwischen diesen bereitgestellt werden müssen und somit die dadurch repräsentierten Daten allgemein verständlich sind. Es existieren einige Vokabulare, die Beschreibungen auf eine allgemeine Weise ermöglichen und dadurch breite Verwendung finden können (Heath u. Bizer, 2011). Bewegt man sich in einem spezifischen Bereich, wie z. B. in dieser Arbeit mit den Jagdbanngeländen, ist es oft nötig für gewisse Begriffe ein eigenes Vokabular zu definieren.

Zwei wichtige Vertreter bestehender Vokabulare sind:

- Das “Dublin Core Metadata Element Set” der Organisation Dublin Core Metadata Initiative (DCMI) hat fünfzehn Eigenschaften definiert, die zur Beschreibung von Ressourcen dienen. Dazu zählen beispielsweise der Ersteller (*creator*), die Sprache (*language*) oder das Thema einer Ressource (*subject*).¹⁴ Im folgenden wird mit “dc” dieses Vokabular angesprochen und im RDF-Format durch den gleichnamigen Namensraum vertreten (`@prefix dc: <http://purl.org/dc/elements/1.1/>`)¹⁵.
- Das “Friend-of-a-Friend” (FOAF) Vokabular stellt Primitiven bereit zur Beschreibung von Personen, deren Beziehungen zu anderen Personen und Objekten sowie deren Aktivitäten (Heath u. Bizer, 2011).

¹²<http://www.w3.org/TR/2009/REC-owl2-primer-20091027/>, 21.06.2011

¹³<http://www.w3.org/standards/semanticweb/ontology>, 21.06.2011

¹⁴<http://dublincore.org/documents/dces/>, 28.12.2011

¹⁵<http://dublincore.org/documents/dc-rdf/>, 28.10.2011

2.4 Abfragesprache SPARQL

Um Anfragen auf Linked Data Graphen durchzuführen, wird die RDF spezifische Abfragesprache SPARQL (SPARQL Protocol and RDF Query Language) verwendet.¹⁶ Das grundlegende Prinzip von SPARQL basiert auf dem Vergleich von Graphen aus der Abfrage und den gespeicherten Graphen (z. B. in einem Triple Store) (Prud’hommeaux u. Seaborne, 2008).

Die Mehrheit der SPARQL-Abfragen bestehen aus einem oder mehreren sogenannten Tripelmustern (*triple patterns*), die zusammengefasst als “basic graph pattern” bezeichnet werden. Ein Tripelmuster ist wie ein normales Tripel aufgebaut, ausser dass Subjekt, Prädikat oder Objekt durch Variablen ersetzt werden können. Diese Muster werden mit den bestehenden Tripeln verglichen und die übereinstimmenden Ressourcen zurückgegeben (Prud’hommeaux u. Seaborne, 2008).

Der Quelltext 2.7 enthält zwei Tripel, die einem Mitarbeiter mit der ID 657 einen Namen und Telefonnummer zuweisen. Die SPARQL-Abfrage im Quelltext 2.8 fragt im ersten Tripelmuster (Zeile 6) nach einer Ressource, die das Attribut `vok:Name` mit dem Wert `"Ramon"` besitzt. Im zweiten Tripelmuster (Zeile 7) wird nach dem Wert des Attributes `vok:Tel` dieser Ressource gesucht.

Eine SPARQL-Abfrage besteht in der Regel aus zwei Teilen: Im SELECT-Abschnitt werden die Variablen aufgezählt, die in der Ausgabe angezeigt werden sollen. Im WHERE-Abschnitt werden die Tripelmuster eingefügt, die mit dem RDF-Datensatz verglichen werden sollen. Die aufgeführte Abfrage verlangt zwei Übereinstimmungen. Erst werden Graphen gesucht, die mit dem ersten Tripelmuster (Zeile 6) übereinstimmen. Danach wird basierend auf den Resultaten nach Entsprechungen für das zweite Tripelmuster (Zeile 7) gesucht (Prud’hommeaux u. Seaborne, 2008). Das Resultat der Abfrage wäre für die Variablen `?ID` der Wert `mitID:657` und für `?Telefon` der Wert `098-765-43-21`. Gäbe es mehrere Personen im Datensatz mit den gleichen Attributen, würden alle diese ausgegeben werden.

```

1 @prefix vok:<http://beispiel.org/> .
2 @prefix mitID:<http://beispiel.org/MitarbeiterID/> .
3
4 mitID:657 vok:Name "Ramon" .
5 mitID:657 vok:Tel "098-765-43-21" .

```

Quelltext 2.7: Beispiel-Tripel für SPARQL-Abfragen.

```

1 @prefix vok:<http://beispiel.org/> .
2 @prefix mitID:<http://beispiel.org/MitarbeiterID/> .
3
4 SELECT ?ID ?Telefon
5 WHERE{
6 ?ID vok:Name "Ramon" .
7 ?ID vok:Tel ?Telefon .
8 }

```

Quelltext 2.8: SPARQL-Abfrage.

SPARQL verfügt zudem über weitere Funktionen um genauere und komplexere Abfragen zu gestalten. So können beispielsweise Resultate gefiltert oder arithmetisch verrechnet werden (Prud’hommeaux u. Seaborne, 2008). In dieser Arbeit wird die SPARQL-Version 1.1 verwendet¹⁷.

¹⁶<http://www.w3.org/standards/semanticweb/query>, 28.12.2011

¹⁷<http://www.w3.org/TR/sparql11-query/>

Eine Weiterentwicklung von SPARQL, für den Umgang mit räumlichen Daten, ist GeoSPARQL. GeoSPARQL ist eine vom OGC vorgeschlagene Anpassung von SPARQL, um räumliche Abfragen durchführen zu können, aber auch um räumliche Daten zu repräsentieren.¹⁸ Es befindet sich im Prozess der Standardisierung, d. h., wird voraussichtlich demnächst offiziell als Standard veröffentlicht. GeoSPARQL wird im Kapitel 3 genauer behandelt.

2.5 Triple Store Sesame

Triple Stores sind für RDF-Daten optimierte Datenbanksysteme zur Datenspeicherung, -verwaltung und -abfrage (Spanos u. a., 2010). Triple Stores dienen also der spezifischen Speicherung von RDF-Daten. Je nachdem liegen die Stärken der Stores mehr im Umgang mit sehr grossen Datenmengen oder in der Folgerung (*reasoning*) aus den impliziten Informationen (Shadbolt u. a., 2006).

Sesame¹⁹ ist ein frei erhältliches Produkt der niederländischen Softwarefirma Aduna und wird in dieser Arbeit verwendet (Versionen 2.5.1 und 2.6.1). Dieser Triple Store ermöglicht die Speicherung, Inferenzierung und Abfrage von RDF-Daten. Die Abfragesprache SPARQL sowie die gängigsten RDF-Datenformate werden vollumfänglich unterstützt.²⁰

Es bestehen unterschiedliche Möglichkeiten auf Sesame zuzugreifen und mit den Daten zu arbeiten. Dazu gehören neben einer eigenen Eingabekonsole²¹ und einer Programmierschnittstelle²² (API) auch eine grafische Benutzeroberfläche ("workbench"), die über den Browser aufgerufen werden kann, sowie der Zugriff über HTTP Methoden²³ (wie GET, PUT usw.). Es wurden die Workbench, die Eingabekonsole und die HTTP Methoden in Kombination mit Python (Version 2.6) verwendet.

2.6 Die Prinzipien von Linked Data

Wie der Name Linked Data suggeriert, sollten Daten nicht einfach bereitgestellt, sondern miteinander verknüpft werden, damit das Web der Daten durchsucht werden kann. Tim Berners-Lee hat dazu vier Regeln aufgestellt.²⁴

- Dinge sollen mittels einem URI benannt werden.
- Grundsätzlich sollten HTTP URIs verwendet werden, wodurch sie auch eindeutig im Web identifizierbar und abrufbar sind. HTTP ist der universelle Zugangsmechanismus des Web, dadurch sollen die durch URIs identifizierten Ressourcen über das Web zugänglich sein (Heath u. Bizer, 2011).
- Der URI sollte auf nützliche Informationen verweisen, welche in einem standardisierten Format (z. B. RDF, SPARQL) dargeboten werden. Auch beim jetzigen Web der Dokumente ist die allgemeine Akzeptanz eines gemeinsamen Standards (HTML) Voraussetzung (Heath u. Bizer, 2011).
- Die letzte Regel verlangt die Einbindung von Verknüpfungen zu anderen URIs. Diese RDF-Verknüpfungen sind Beziehungen und haben somit im Gegensatz zu den normalen Hyperlinks im Web der Dokumente eine explizite Bedeutung, wodurch die Beziehung beschrieben wird. Sie ermöglichen die Schaffung

¹⁸<http://www.opengeospatial.org/projects/groups/geosparqlswg>, 28.12.2011

¹⁹<http://www.openrdf.org/>

²⁰<http://www.openrdf.org/about.jsp>, 23.11.2011

²¹<http://www.openrdf.org/doc/sesame2/users/ch07.html>, 23.11.2011

²²<http://www.openrdf.org/doc/sesame2/users/ch08.html>, 23.11.2011

²³<http://www.openrdf.org/doc/sesame2/system/ch08.html#d0e304>, 23.11.2011

<http://www.jenitennison.com/blog/node/153>, 23.11.2011

²⁴<http://www.w3.org/DesignIssues/LinkedData.html>, 16.08.2011

des Web der Daten, da unterschiedliche Datensätze, aber auch Vokabulare miteinander verbunden werden (Heath u. Bizer, 2011).

Diese Prinzipien sind allgemein anerkannt und werden dementsprechend auch in dieser Arbeit angewandt.

2.7 Bestehende Projekte

2.7.1 Linking Open Data (LOD)

Das Ziel des Linking Open Data Projektes ist die Bereitstellung öffentlich zugänglicher Datensätze im RDF-Format und deren Verknüpfung mit anderen Datensätzen zur sogenannten LOD Cloud. Das Projekt hat 295 verschiedene Datensätze, bestehend aus 31 Milliarden RDF-Tripel und 504 Millionen RDF-Verknüpfungen, veröffentlicht (Stand September 2011).²⁵ Die Abbildung 2.3 (*LOD cloud diagram*) visualisiert diese Datensätze und Verknüpfungen. Die Knoten repräsentieren die bereitgestellten Datensätze, während die Kanten auf Verknüpfungen dazwischen deuten. Die Pfeile geben die Richtung der Verknüpfungen an, wobei ihre Strichdicke mit der Anzahl Verknüpfungen zwischen zwei Datensätzen korrespondiert.²⁶ Dabei fallen die zahlreichen gerichteten Verknüpfungen zu DBpedia, aber auch zu GeoNames ins Auge.

2.7.2 DBpedia

Das Projekt DBpedia extrahiert strukturierte Informationen aus Wikipedia und stellt sie über das Web als RDF-Beschreibung bereit. Diese Daten entsprechen jeweils dem aktuellen Stand von Wikipedia. In den letzten Jahren hat sich DBpedia zu einem zentralen Punkt für das Web der Daten kristallisiert, d. h. eine Menge Datenanbieter haben ihre Datensätze mit DBpedia verknüpft. Die breite Abdeckung von unterschiedlichsten Themenfeldern hat diese Stellung weiter gefestigt (Bizer u. a., 2009b). Durch die Strukturierung der in den Wikipedia Artikeln enthaltenen Informationen werden komplexere und spezifische Abfragen möglich. Beispielsweise können Personen gesucht werden, die vor 1900 in Zürich geboren wurden oder Fußballspieler, die aus einem Land mit mehr als 10 Millionen Einwohner stammen und als Torhüter in einem Club tätig waren, dessen Heimstadion mehr als 40'000 Sitzplätze anbietet.²⁷ Das Beispiel in Abbildung 2.4 zeigt einen Ausschnitt der Suchresultate nach dem Begriff "Matterhorn" in der Benutzeroberfläche "Public Faceted Web Service Interface"²⁸. Diese Abfrage ist textbasiert und präsentiert Ressourcen, in welchen dieser Begriff vorkommt, wozu beispielsweise neben dem Berg auch die gleichnamige Attraktion im Disneyland zählt. In der Abbildung 2.5 ist ein Ausschnitt der RDF-Repräsentation des Berges enthalten. Hier wird die Strukturierung der Informationen gut ersichtlich. So sind beispielsweise die Koordinaten durch die Attribute `geo:lat` und `geo:long` separat ausgewiesen oder es wird durch das Attribut "sameAs" auf Ressourcen in fremden Datensätzen verwiesen, die ebenfalls den Berg Matterhorn repräsentieren.

²⁵<http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>, 25.12.2011

²⁶<http://www4.wiwi.fu-berlin.de/lodcloud/state/>, 25.12.2011

²⁷<http://wiki.dbpedia.org/OnlineAccess>, 29.12.2011

²⁸<http://dbpedia.org/fct>

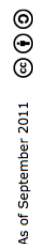


Abbildung 2.3: Das LOD Cloud Diagramm stellt die veröffentlichten Datensätze und deren Verknüpfungen dar (Quelle: <http://richard.cyganiak.de/2007/10/loд/imagemap.html>, 25.12.2011).

Entity	Title	Text excerpt
dbpedia:Matterhorn_%28ride%29	Matterhorn (ride)	For the Disneyland attraction, see Matterhorn Bobsleds. The Matterhorn or Flying Bobs, sometimes known by alternate names such as Musik Express or Terminator in... For the Disneyland attraction, see Matterhorn Bobsleds. The Matterhorn or Flying Bobs,...
dbpedia:Switzerland	Switzerland	Contrasted landscapes between the Matterhorn area in the high Alps, the Sanetsch region and the plateau...
category:Matterhorn_Gothard_Bahn_stations	Matterhorn Gotthard Bahn stations	Matterhorn Gotthard Bahn stationsMatterhorn Gotthard Bahn stations.
yago:MatterhornGotthardBahnStations	Matterhorn Gotthard Bahn stations	Matterhorn Gotthard Bahn stations.
category:Matterhorn_Gothard_Bahn_locomotives	Matterhorn Gotthard Bahn locomotives	Matterhorn Gotthard Bahn locomotivesMatterhorn Gotthard Bahn locomotives.
yago:MatterhornGotthardBahnLocomotives	Matterhorn Gotthard Bahn locomotives	Matterhorn Gotthard Bahn locomotives.
dbpedia:Matterhorn_Gothard_Bahn	Matterhorn Gotthard Bahn	Matterhorn Gotthard BahnDie Matterhorn Gotthard Bahn MGB ist eine Schmalspurbahn in der Schweiz mit... Die Matterhorn Gotthard Bahn MGB ist eine Schmalspurbahn in der

Abbildung 2.4: Suchresultate nach dem Begriff “Matterhorn” (Quelle: <http://dbpedia.org/fct/facet.vsp?cmd=text&sid=7459>, 29.12.2011).

sameAs	http://linkedgeodata.org/triplify/node/26863664#idfbase:m/01p8j1 http://sws.geonames.org/2659729/
rdfs:comment	»more«
rdfs:label	Matterhorn »more«
foaf:name	Matterhorn Monte Cervino
geo:lat	45.976387(xsd:float)
geo:long	7.658333(xsd:float)
has abstract	The Matterhorn (German), Monte Cervino (Italian) or Mont Cervin (French), is a mountain in the Pennine Alps on the border between Switzerland and Italy. Its summit is 4,478 metres (14,690 ft) high, making it one of the highest peaks in the Alps. The four steep faces, rising above the surrounding glaciers, face the

Abbildung 2.5: RDF-Repräsentation des Berges “Matterhorn” (Quelle: <http://dbpedia.org/fct/facet.vsp?cmd=text&sid=7459>, 29.12.2011).

2.8 Fazit

In diesem Kapitel wurden die wichtigsten Grundlagen, Technologien und Standards für die Umsetzung von Linked Data und somit zur Realisierung eines Web der Daten vorgestellt. Zentrale Bestandteile sind die Bereitstellung von strukturierten Informationen und deren Verknüpfung mit anderen Datenquellen. Es existieren unterschiedliche Möglichkeiten strukturierte Informationen anzubieten, wie beispielsweise Mikroformate, spezielle Datenformate oder der direkte Zugang über APIs. Zudem gibt es gemäss Heath u. Bizer (2011) einige Gründe, die für die Verwendung der Linked Data Technologien sprechen:

RDF stellt ein weltweit einheitliches Datenmodell bereit, womit Daten repräsentiert, identifiziert und zueinander in Beziehung gesetzt werden können. Die Benutzung der bereitgestellten Daten wird vereinfacht. Dazu gehört auch die Verwendung gemeinsamer Vokabulare (Ontologien) bzw. Übersetzungen, womit Klarheit über den Kontext und die Bedeutung der gemeinsam verwendeten Daten geschaffen wird. Die Daten sollten also selbsterklärend sein (Heath u. Bizer, 2011).

Der Einbezug von HTTP erlaubt den standardisierten Zugriff auf die angebotenen Daten und ist bereits durch das heutige Web etabliert, so dass beispielsweise ein Browser für den Datenzugriff verwendet werden kann. HTTP URIs identifizieren die Ressourcen global eindeutig. Durch die Verknüpfung mit Ressourcen aus fremden Datenquellen wird das Web der Daten geschaffen. Also ein gemeinsamer globaler Datenraum, der auf der Suche nach neuen Informationen durchstöbert werden kann (Heath u. Bizer, 2011).

Geodaten als Linked Data

Im riesigen Angebot des Web finden sich unter anderem Daten mit direktem räumlichen Bezug oder räumlichem Hintergrund, denen eine relativ hohe Bedeutung zugemessen werden darf. Dieses Potential wurde auch von kommerziellen Anbietern erkannt. So existieren beispielsweise eine Menge Onlinekarten-Anbieter, unter denen sich auch grosse Unternehmen wie Google oder Microsoft befinden (Hart u. Dolbear, 2007). Oft werden nicht nur statische Basiskarten angeboten, sondern sie sind mit zusätzlichen Daten angereichert, die interaktiv durch den Benutzer abgerufen werden können. So lassen sich beispielsweise nebst dem Kartenausschnitt auch die nächsten Bushaltestellen inklusive Abfahrtszeiten oder der nächste Geldautomat anzeigen.

Räumliche Informationen bilden vielfach auch eine Verknüpfung zwischen verschiedenen Datensätzen. Verfügt beispielsweise eine Umweltbehörde über Angaben zu Tiervorkommen und den Jagdrevieren einer Region, so kann durch Flächenverschneidung (*polygon/map overlay*) überprüft werden, ob diese Vorkommen innerhalb eines solchen Gebietes liegen (Hart u. Dolbear, 2007).

Daher war es vorhersehbar, dass sich diese Bedeutsamkeit räumlicher Daten früher oder später auch in Bezug auf das Semantic Web widerspiegeln wird. Somit entstand das Bedürfnis, Geodaten in das RDF-Format konvertieren zu können. Dies ist jedoch aufgrund ihrer spezifischen Eigenschaften nicht trivial: Geodaten weisen untereinander räumliche Beziehungen auf. Je nach Datenmodell (z. B. Spaghetti-Modell) sind diese topologischen Informationen nicht explizit abgelegt (im Gegensatz z. B. zum Topologischen Vektormodell). Angesichts der Möglichkeit geometrische Berechnungen durchzuführen, beispielsweise mit einem GIS, besteht auch keine Notwendigkeit dazu. So können jeweils die räumlichen Beziehungen flexibel und spezifisch für die gewählten Objekte ad hoc berechnet werden. Auch als Linked Data sollte die Darstellung räumlicher Beziehungen möglich sein, um das Potential von Geodaten auszuschöpfen. Die Lösungsansätze können verschiedene Ebenen betreffen. So könnten auf der konkreten Datenebene geometrische Berechnungen eingeführt werden. Auf der Ebene der Ontologie könnten durch deren Anpassung räumliche Beziehungen definiert werden (Hart u. Dolbear, 2007).

Für die Repräsentation und Abfrage von räumlichen Daten im RDF-Format existiert noch kein einheitlicher Standard, weshalb viele Organisationen (auch kommerzielle) ihre eigenen Strategien und Ontologien dazu entwickelt haben. Jedoch wäre ein gemeinsamer konsistenter Standard von Vorteil. Dementsprechend wurden in der Vergangenheit bereits einige Anläufe unternommen einen Standard für räumliche Daten im RDF-Format zu definieren, wovon sich keiner durchsetzen konnte (Battle u. Kolas, 2011). Grundsätzlich kann frei darüber entschieden werden, welche Informationen als Linked Data abgelegt werden sollen. So könnte das Spektrum bei den Geodaten von einer einfachen Benennung des Datensatzes (im Sinne "es existiert ein Datensatz") bis hin zur detaillierten Umformung sämtlicher Informationen (inklusive den Geometriedaten) in das RDF-Format reichen.

In diesem Kapitel werden zuerst einige Projekte vorgestellt, die die Einbindung von räumlichen Daten in

Linked Data zum Ziel haben. Nachfolgend werden diverse bestehende Ansätze erläutert, wie Geodaten als RDF repräsentiert werden könnten. Zu den Ansätzen werden jeweils Beispiele (gekürzte und teilweise abgeänderte Versionen der Originale) gezeigt.

3.1 Bestehende Projekte

3.1.1 LinkedGeoData.org

Das Projekt LinkedGeoData.org¹ (LGD) hat die Integration von OpenStreetMap² (OSM) Daten in das Semantic Web zum Ziel (Stadler u. a., 2011). OpenStreetMap ist ein Projekt, das die Erstellung von frei verfügbaren Karten anstrebt. Registrierte Mitglieder erheben eigenhändig die Kartendaten. Diese dürfen kostenlos verwendet und auch weiterverarbeitet werden.³

Die Verantwortlichen von LinkedGeoData.org sehen in der Bereitstellung von in RDF konvertierten Daten einen grossen Nutzen für die alltägliche Informationsbeschaffung, da die Kartendaten mit anderen externen Daten verbunden werden können. Beispielsweise könnte die Suche nach Lebensmittelgeschäften innerhalb eines bestimmten Gebietes nicht nur deren Standort, sondern auch deren Angebote ergeben (Stadler u. a., 2011). Zentrale Aspekte der Arbeit von LGD sind: Konvertierung der OSM-Daten in das RDF-Format, Bereitstellung einer einfachen Ontologie für diese Daten, Verknüpfung mit anderen Linked Data Quellen wie DBpedia oder GeoNames, sowie die Verwendung von verschiedenen Kartensymbolen und Beschriftungen aus unterschiedlichen Quellen (Stadler u. a., 2011).

Zusätzlich wurde ein LinkedGeoData Browser und Editor⁴ kreiert, der auf Kartenbasis Knoten und Wege analysiert, sowie Filtermöglichkeiten anbietet. Registrierte OSM-Benutzer können damit die angezeigten Elemente direkt editieren.⁵

Die Abbildung 3.1 zeigt einen Ausschnitt um den Zürcher Zoo. Auf der linken Seite des Browsers werden die im Bildausschnitt befindlichen Knoten angezeigt, auf der rechten Seite können diese gefiltert werden. Im Bild wurde die Tramhaltestelle beim Zoo ausgewählt (gelb hervorgehoben). Die zugehörigen Informationen (RDF-Tripel) können dem unteren Informationsfenster entnommen werden.

3.1.2 GeoNames

Bei GeoNames⁶ handelt es sich um eine kostenlose geographische Datenbank, mit über zehn Millionen geographischen Namen (entsprechend einem *Gazetteer*). Diese sind oft in verschiedene Sprachen übersetzt und können mit weiteren Daten, wie der Bevölkerungszahl oder Höhe des Ortes angereichert sein. Die Daten werden über Webservices abgerufen oder können in Form eines Datenpaketes (aufgeteilt nach Staaten) heruntergeladen werden.⁷ Zusätzlich werden die Daten im RDF-Format mit entsprechender Ontologie angeboten, sind also mittels Linked Data referenzierbar. Die Daten untereinander sind ebenfalls verknüpft, beispielsweise ein Land mit den Nachbarländern, was auch topologisch ausgewertet werden kann.⁸

Abbildung 3.2 zeigt einen Ausschnitt von Genf. In GeoNames vorhandene Städte- und Ortsnamen sind

¹ <http://linkedgeodata.org/>

² <http://www.openstreetmap.org/>

³ http://www.openstreetmap.ch/faq.php#was_ist_osm, 08.11.2011

⁴ <http://browser.linkedgeodata.org>

⁵ <http://linkedgeodata.org/LGD%20Browser>, 08.11.2011

⁶ <http://www.geonames.org/>

⁷ <http://www.geonames.org/about.html>, 08.11.2011

⁸ <http://www.geonames.org/ontology/documentation.html>, 08.11.2011

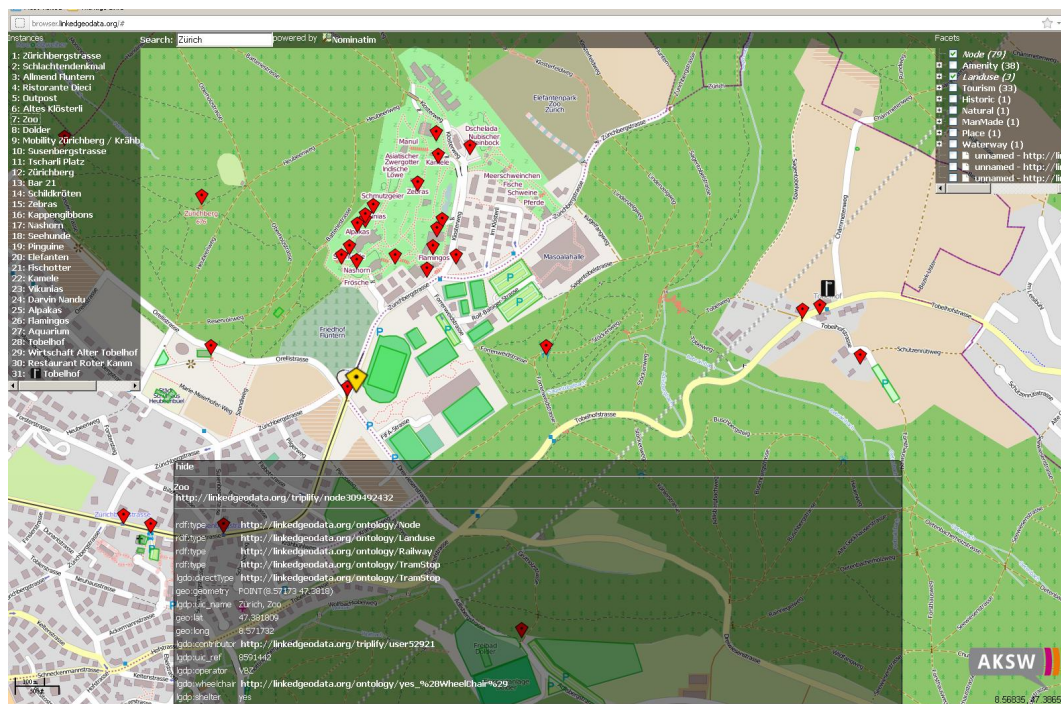


Abbildung 3.1: LGD Browser Ausschnitt vom Gebiet um den Zoo Zürich mit ausgewählter Tramhaltestelle (Quelle: <http://browser.linkedgedata.org/#>, 14.11.2011).

mit einem “P” markiert. Das Pop-up zeigt die vorhandenen Daten für den Ort Genf, inklusive einem Link zum passenden Wikipedia Artikel.

3.1.3 Administrative Geography des Ordnance Survey (OS)

Die Landestopographie (*national mapping agency*) von Grossbritannien, der Ordnance Survey (OS), untersucht unter anderem die Verwendung von semantischen Webtechnologien als Informationsquelle geographischer Daten. Im Zentrum stehen die Entwicklung von “Daten-Prototypen” und die Bereitstellung von Daten im RDF-Format (Goodwin u. a., 2008). Diese Daten werden zudem im Rahmen einer Initiative zur Veröffentlichung von Behördendaten angeboten.⁹

Einer dieser Prototypen ist die Umformung der administrativen Grenzen von Grossbritannien in das RDF-Format (Goodwin u. a., 2008). Später wurde der Datenbestand durch die Verknüpfung sämtlicher Postleitzahlen mit den administrativen Grenzen erweitert.¹⁰ Die Behörde stellte die Anforderung, dass nebst den Sachabfragen auch topologische Abfragen möglich sein sollten (Goodwin u. a., 2008). Die Daten selbst wurden basierend auf einem bestehenden Datensatz generiert, der die offiziellen Grenzen der administrativen Einheiten gespeichert hat (Goodwin u. a., 2008).

3.1.4 The National Map des USGS

Die US-Amerikanische Behörde U.S. Geological Survey (USGS) entwickelt im Rahmen einer Fallstudie eine Ontologie für “The National Map” und konvertiert ausgewählte Testdaten in das RDF-Format. Dazu haben sie Methoden entwickelt, die nicht nur die Daten im RDF-Format zugänglich, sondern auch die

⁹ <http://data.ordnancesurvey.co.uk/.html>, 28.11.2011

¹⁰ <http://blog.ordnancesurvey.co.uk/2010/11/linked-data-at-ordnance-survey/>, 28.11.2011

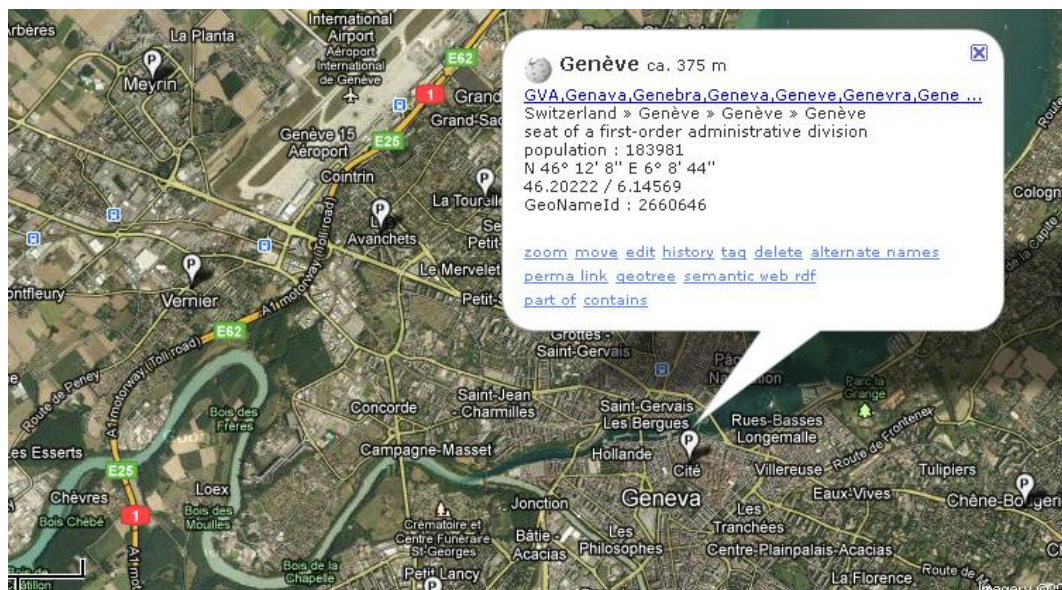


Abbildung 3.2: Ausschnitt von Genf aus GeoNames (Quelle: http://www.geonames.org/maps/google_46.202_6.146.html, 22.11.2011).

Resultate von semantischen Abfragen grafisch darstellbar machen (Usery u. Varanka, 2011). The National Map ist ein Programm von Behörden und lokalen Partnern zur Bereitstellung von unterschiedlichen geographischen Informationen über die Vereinigten Staaten von Amerika.¹¹ Die Abbildung 3.3 zeigt einen Ausschnitt des grafischen Resultates der folgenden Abfrage:

```

1 PREFIX ogc: <http://www.opengis.net/rdf/#>
2 PREFIX fid: <http://cegis.usgs.gov/rdf/nhd/featureID/#>
3 SELECT ?feature ?type
4 WHERE {
5   fid:_102217454 ogc:hasGeometry ?geo1.
6   ?geo1 ogc:touches ?geo2.
7   ?feature ogc:hasGeometry ?geo2.
8   ?feature a ?type. }

```

Quelltext 3.1: SPARQL-Abfrage zu The National Map (verändert nach: Usery u. Varanka, 2011, S. 11).

Der Abfrage im Quelltext 3.1 ist zu entnehmen, dass sämtliche Instanzen gesucht werden, deren Geometrien die Geometrie der Instanz mit dem Identifikator `fid:_102217454` berühren (`ogc:touches`), welche den Fluss “West Hunter Creek” repräsentiert. Dieser ist in der Abbildung 3.3 rot eingefärbt (mit dem URI <http://cegis.usgs.gov/rdf/nhd/featureID#102217454>). In blau sind die Resultate der Abfrage, welche Zuflüsse darstellen und somit topologisch gesehen den West Hunter Creek berühren. Der Hintergrund ist mit einer Rasterkarte vom USGS hinterlegt (Usery u. Varanka, 2011).

¹¹<http://nationalmap.gov/about.html>, 28.11.2011

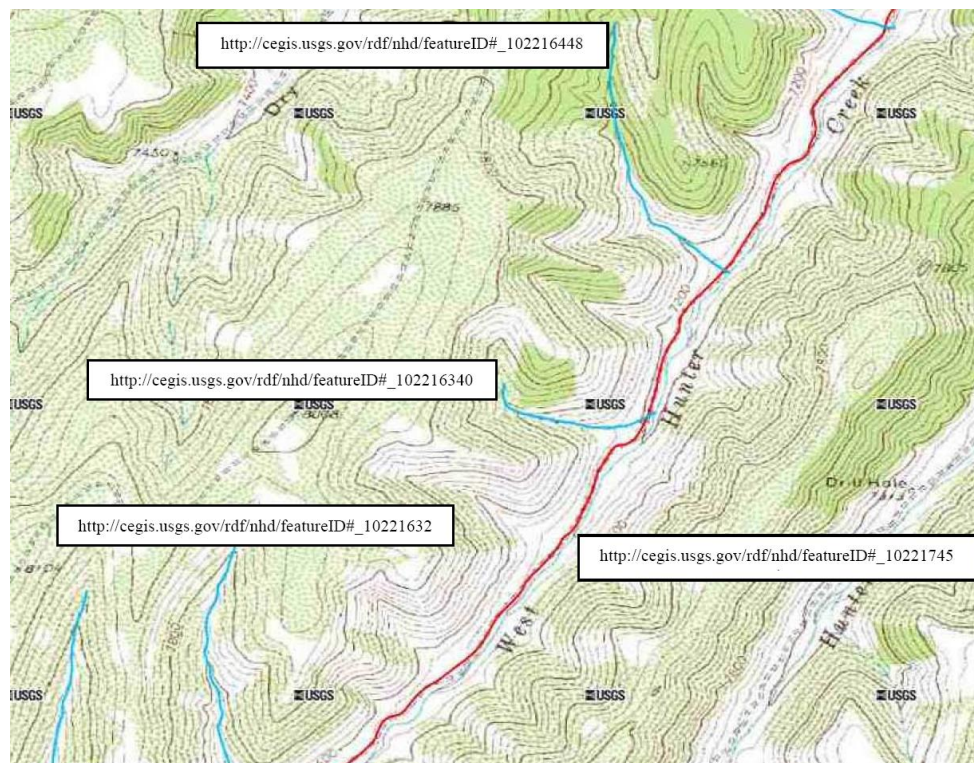


Abbildung 3.3: Ausschnitt des grafischen Resultates der SPARQL-Abfrage aus Quelltext 3.1 (Quelle: Usery u. Varanka, 2011, S. 12).

3.2 Umformung von Vektordaten

Das Vektordatenmodell, dessen grundlegende Bausteine Punkte, Linien und Polygone sind, eignet sich besonders zur Darstellung von Geoobjekten. Abgrenzbare geographische Objekte, wie beispielsweise Häuser, Strassen oder auch Jagdbanngebiete werden vorzugsweise durch Vektoren modelliert. Geographische Objekte (auch als Entitäten bezeichnet) besitzen Attribute und können räumlich klar abgegrenzt, sowie räumlich verortet werden. Attribute können meist deutlich dem ganzen Objekt zugeordnet werden (Burrough u. McDonnell, 2006).

3.2.1 Punktdaten mit dem Basic Geo Vocabulary

Eine etablierte Weise Punkte im RDF-Format abzulegen, ist das Basic Geo Vocabulary des W3C, welches jedoch weder standardisiert wurde noch auf dem Weg dahin ist. Es stellt ein Vokabular zur Repräsentation der geographischen Breite und Länge von Punktobjekten im WGS84 Referenzsystem bereit. Die dementsprechend vorhandene Klasse `geo:Point` wird durch die Attribute `geo:lat` (Breite), `geo:long` (Länge) und `geo:alt` (Höhe) beschrieben (Brickley, 2006). Das folgende Beispiel stellt die alte Sternwarte von Bern im RDF/XML-Format dar, welche als Fundamentalpunkt für das ältere schweizerische Bezugssystem CH1903 dient.¹²

¹²<http://www.swisstopo.admin.ch/internet/swisstopo/de/home/topics/survey/sys/refsys/switzerland.html>, 30.08.2011

```

1 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
2   xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#">
3   <geo:Point>
4     <geo:lat> 46.951081</geo:lat>
5     <geo:long> 7.438637</geo:long>
6   </geo:Point>
7 </rdf:RDF>

```

Quelltext 3.2: Punkt im Basic Geo Vocabulary Modell (verändert nach: Brickley, 2006).

In den ersten zwei Zeilen des Quelltextes 3.2 werden die Namensräume angegeben. Gleich danach folgt die Definition des Punktes durch Angabe der geographischen Breite (`geo:lat`) und der Länge (`geo:long`) im WGS84 Referenzsystem (Zeilen 4 und 5), wobei mit `geo:Point` in den Zeilen 3 und 6 die Geometrie explizit als Punkt definiert wird.

Wie den folgenden Beispielen zu entnehmen ist, bedienen sich die Projekte LinkedGeoData.org wie auch GeoNames des Basic Geo Vocabulary, um Punkte in RDF zu repräsentieren (Auer u. a., 2009).

Bei der Punktdefinition von GeoNames (im RDF/XML-Format) im Quelltext 3.3 werden neben den Punktkoordinaten (`wgs84_pos:lat`, `wgs84_pos:long`) noch weitere Attribute wie `gn:population` (Zeile 6) für die Bevölkerung und `gn:wikipediaArticle` (Zeile 9) als Verweis auf einen passenden Wikipedia Artikel definiert. Weiter wird durch `rdfs:seeAlso` in Zeile 11 die Verknüpfung zu einem externen Datensatz, in diesem Fall DBpedia.org, festgelegt.

```

1 <rdf:RDF
2   xmlns:gn="http://www.geonames.org/ontology#"
3   xmlns:wgs84_pos="http://www.w3.org/2003/01/geo/wgs84_pos#">
4   <gn:Feature rdf:about="http://sws.geonames.org/3020251/">
5     <gn:name>Embrun</gn:name>
6     <gn:population>7069</gn:population>
7     <wgs84_pos:lat>44.56387</wgs84_pos:lat>
8     <wgs84_pos:long>6.49526</wgs84_pos:long>
9     <gn:wikipediaArticle rdf:resource="http://de.wikipedia.org/wiki/
10      Embrun"/>
11     <rdfs:seeAlso rdf:resource="http://dbpedia.org/resource/Embrun\%2C_
12      Hautes-Alpes"/>
13   </gn:Feature>
14 </rdf:RDF>

```

Quelltext 3.3: Punkt bei GeoNames (verändert nach: <http://sws.geonames.org/3020251/about.rdf> und <http://www.geonames.org/ontology/documentation.html>, 11.11.2011).

Im Beispiel von LinkedGeoData.org (im N3-Format) im Quelltext 3.4 wird in den letzten zwei Tripeln (Zeilen 8 - 13) der Punkt im Namensraum des Basic Geo Vocabulary ausgedrückt. Zusätzlich erfolgt eine Definition (Zeilen 4 - 7) durch ein eigenes Format eines privaten Anbieters (OpenLink¹³). Dies ermöglicht bei Verwendung der OpenLink Virtuoso Datenbank die Indexierung in einem R-Baum (Stadler u. a., 2011). Aus den Tripeln kann herausgelesen werden, dass es sich um eine Bowling Anlage handelt (erstes Tripel).

```

1 <http://linkedgeo.org/triplify/node1000003273>
2   <http://www.w3.org/2000/01/rdf-schema#label>
3     "Classic Bowling Lanes" .
4 <http://linkedgeo.org/triplify/node1000003273>

```

¹³<http://virtuoso.openlinksw.com/>

```

5 <http://www.w3.org/2003/01/geo/wgs84_pos#geometry>
6 "POINT(-86.54 39.18)"^^<http://www.openlinksw.com/schemas/virttrdf#
7   Geometry> .
8 <http://linkedgeodata.org/triplify/node1000003273>
9 <http://www.w3.org/2003/01/geo/wgs84_pos#lat>
10 "39.18"^^<http://www.w3.org/2001/XMLSchema#double> .
11 <http://linkedgeodata.org/triplify/node1000003273>
12 <http://www.w3.org/2003/01/geo/wgs84_pos#long>
13 "-86.54"^^<http://www.w3.org/2001/XMLSchema#double> .

```

Quelltext 3.4: Punkt bei LinkedGeoData.org (verändert nach: http://downloads.linkedgeodata.org/preview.php?file=releases_sl_110406_sl_LGD-Dump-110406-RelevantNodes.sorted.nt.bz2, 11.11.2011).

3.2.2 Vektordaten mit GeoRSS

In den Jahren 2005 bis 2007 fand eine Weiterentwicklung des Basic Geo Vocabulary zum GeoRSS Modell statt, um auch Linien und Polygone beschreiben zu können. Jedoch wurde auch dieses nie standardisiert, fand aber trotzdem Verwendung in der Praxis (Battle u. Kolas, 2011). Wie der Name bereits suggeriert, wurde GeoRSS ursprünglich entwickelt, um räumliche Informationen in RSS Feeds (vgl. Exkurs: RSS) darstellen zu können. Es existieren die Versionen GeoRSS Simple und GeoRSS GML (Reed, 2006). GeoRSS Simple unterstützt Punkte, Linien, Rechtecke und Polygone. Zudem können Beziehungen, Höhen und Radien definiert werden. Jedoch ist das Referenzsystem auf WGS84 beschränkt. GeoRSS GML basiert auf dem Standard GML¹⁴ (Geography Markup Language¹⁵) und kann als eine vereinfachte Version davon verstanden werden, sowie auch als eine Erweiterung von GeoRSS Simple. So kann beispielsweise ein anderes Referenzsystem als WGS84 bestimmt werden (Reed, 2006).

Der Quelltext 3.5 definiert ein Polygon im GeoRSS Simple Format (<georss:polygon>), das die Schadenzone eines Hochwassers abbildet (<title>). Eine gewisse "Hanna Muster" ist die Autorin dieses Eintrages (<author>).

```

1 <?xml version="1.0" encoding="utf-8"?>
2 <feed xmlns="http://www.w3.org/2005/Atom"
3     xmlns:georss="http://www.georss.org/georss"
4     xmlns:gml="http://www.opengis.net/gml">
5   <title>Schadenzonen Hochwasser</title>
6   <author>
7     <name>Hanna Muster</name>
8   </author>
9   <entry>
10     <georss:polygon>
11       45.256 -110.45 43.84 -109.86 45.256 -110.45
12     </georss:polygon>
13   </entry>
14 </feed>

```

Quelltext 3.5: Polygon im GeoRSS Simple Modell (verändert nach: <http://www.georss.org/simple>, Abschnitt Point, 30.08.2011).

Der gleiche Eintrag im GeoRSS GML Format ist im Quelltext 3.6 aufgeführt (nur abweichender Teil). Vom Aufbau her sind beide Modelle praktisch identisch, jedoch bietet GeoRSS GML mehr Möglichkeiten

¹⁴<http://www.opengeospatial.org/standards/gml>

¹⁵GML dient der Speicherung und dem Transfer von Geodaten

zur Spezifikation der Geometrie, da es wie schon erwähnt auf dem GML-Standard beruht.

```

1 [...]
2   <entry>
3     <georss:where>
4       <gml:Polygon>
5         <gml:exterior>
6           <gml:LinearRing>
7             <gml:posList>
8               45.256 -110.45 43.84 -109.86 45.256 -110.45
9             </gml:posList>
10            </gml:LinearRing>
11          </gml:exterior>
12        </gml:Polygon>
13      </georss:where>
14    </entry>
15 </feed>

```

Quelltext 3.6: Polygon im GeoRSS GML Modell (verändert nach: <http://www.georss.org/gml>, Abschnitt Point, 30.08.2011).

LinkedGeoData.org verwendet neben dem Basic Geo Vocabulary für Punkte das GeoRSS Simple Modell um Linien und Flächen darzustellen. Der Quelltext 3.7 im N3-Format zeigt ein Polygon, welches als Spielplatz deklariert ist.

```

1 <http://linkedgedata.org/triplify/way100000065>
2   <http://linkedgedata.org/ontology/directType>
3   <http://linkedgedata.org/ontology/Playground> .
4 <http://linkedgedata.org/triplify/way100000065>
5   <http://www.georss.org/georss/polygon>
6   "60.010 30.350 60.011 30.350 60.011 30.350 60.010 30.350" .

```

Quelltext 3.7: Polygon bei LinkedGeoData.org (verändert nach: http://downloads.linkedgedata.org/preview.php?file=releases_sl_110406_sl_LGD-Dump-110406-RelevantWays.sorted.nt.bz2, 11.11.2011).

3.2.3 Vektordaten beim Ordnance Survey

Der Ordnance Survey benutzte für das Projekt der Umformung der administrativen Grenzen von Grossbritannien zwar eine eigene Ontologie, doch wird die Geometrie ähnlich wie bei GeoRSS GML repräsentiert, d. h. auch der OS greift auf den GML-Standard zurück.¹⁶ Das Beispiel im Quelltext 3.8 zeigt im RDF/XML-Format die Geometrie der Stadt Southampton mit Angabe der Fläche (<geometry:hectares>).

```

1 <rdf:RDF>
2   <rdf:Description rdf:about="http://data.ordnancesurvey.co.uk/id/geometry/
3     96957">
4     <rdf:type rdf:resource="http://data.ordnancesurvey.co.uk/ontology/
5       geometry/
6       AbstractGeometry"/>
7     <geometry:asGML rdf:parseType="Literal">
8       <gml:Polygon srsName="os:BNG">
9         <gml:exterior>
10          <gml:LinearRing>

```

¹⁶<http://data.ordnancesurvey.co.uk/doc/geometry/96957.rdf>, 08.11.2011

```

11         <gml:posList srsDimension="2">
12             443348.1 108678.3 [...] 443326.7 108696.6 443348.1 108678.3
13         </gml:posList>
14     </gml:LinearRing>
15 </gml:exterior>
16 </gml:Polygon>
17 </geometry:asGML>
18 <geometry:hectares>5638.515</geometry:hectares>
19 </rdf:Description>
20 </rdf:RDF>

```

Quelltext 3.8: Polygon bei der Ordnance Survey (verändert nach: <http://data.ordnancesurvey.co.uk/doc/geometry/96957.rdf>, 11.11.2011).

Exkurs: RSS

Really Simple Syndication (RSS) ist ein Format, um sogenannte RSS-Feeds zu definieren. Diese beinhalten oft sich schnell ändernde Inhalte, wie Nachrichtenüberschriften oder Blogeinträge und machen diese öffentlich zugänglich. Interessiert sich beispielsweise jemand für die aktuellsten Sportnachrichten von verschiedensten Webseiten, so kann diese Person mit einem sogenannten Feed-Reader die verschiedenen RSS-Feeds, welche Sportnachrichten anbieten, abonnieren (sofern von den Nachrichtenanbietern bereitgestellt) und sich direkt anzeigen lassen, ohne dass die Informationen über mehrere Seiten hinweg manuell zusammengesucht werden müssen. RSS strukturiert die Informationen, ist maschinenlesbar und weist somit Ähnlichkeiten zu RDF auf (Reed, 2006).

3.2.4 Vekordaten mit Geo OWL

Das W3C strebte eine Erweiterung des Basic Geo Vokabulars an, um den räumlichen Inhalten der zu beschreibenden Objekten (Ressourcen) besser gerecht zu werden. Diese orientiert sich stark am Ansatz von GeoRSS, sollte aber einfacher zu implementieren sein. Dementsprechend wurde auf das existierende GeoRSS Vokabular zurückgegriffen, welches die Beschreibung von Punkten, Linien, Rechtecken und Polygonen erlaubt. Damit wurde eine deutliche Einschränkung in Kauf genommen, verglichen mit den Möglichkeiten des etablierten GML Simple Features Modell¹⁷ des OGC (Liebermann u. a., 2007). Zusätzlich gab es Überlegungen zur Abbildung räumlicher Beziehungen, doch fanden sie keinen Eingang in das Modell (Liebermann u. a., 2007). Das Resultat dieser Anstrengungen war zwar ein ausgereiftes Modell zur Repräsentation räumlicher Daten, trotzdem wurde Geo OWL kein offizieller Standard. Zudem beschränkte sich Geo OWL wie GeoRSS Simple auf das WGS84 Referenzsystem (Battle u. Kolas, 2011).

¹⁷<http://www.opengeospatial.org/standards/gml>, 26.10.2011

Das Beispiel eines Punktes im Quelltext 3.9, in dessen Nähe sich die Person “Josh Lieberman” befindet, zeigt die Ähnlichkeit zu GeoRSS GML:

```

1 <rdf:RDF xmlns="http://xmlns.com/foaf/0.1/"
2   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3   xmlns:gml="http://www.opengis.net/gml/">
4 <Person>
5   <name>Josh Lieberman</name>
6   <based_near>
7     <gml:Point>
8       <gml:pos>42.34 -71.21</gml:pos>
9     </gml:Point>
10  </based_near>
11 </Person>
12 </rdf:RDF>

```

Quelltext 3.9: Punkt im Geo OWL Modell (verändert nach: Lieberman, 2007, Kapitel 1.1, 11.11.2011).

3.2.5 Vektordaten mit NeoGeo

Da sich trotz verschiedener Anläufe kein allgemeiner Standard etablieren konnte, welcher den Anforderungen zur Repräsentation räumlicher Daten genügte, entwickelten viele Datenanbieter ihr eigenes Vokabular. Die NeoGeo Community¹⁸ wollte dagegen steuern und unternahm weitere Anstrengungen einen Standard zu definieren. Deren Vokabular basiert auf dem GML Simple Features Modell, wodurch sie gleichzeitig mit einem GIS kompatibel sein sollten (Salas u. a., 2011). Die Grundidee ist, möglichst alles als Ressourcen zu repräsentieren. Beispielsweise wird jeder Knotenpunkt einer Linie als eigenständige Ressource betrachtet, welche in einer RDF-Collection zusammengefasst werden, während in den früheren Ansätzen sämtliche Knoten innerhalb einer Ressource liegen. Dadurch wird jeder einzelne Punkt durch Vergabe eines URI ansprechbar. Das NeoGeo Vokabular erlaubt nicht nur die Definition von Linien, Punkten und Polygonen, sondern auch von mehrteiligen Geometrien (Multipoint, Multiline und Multipolygon) (Salas u. a., 2011). Zusätzlich können räumliche Beziehungen basierend auf RCC8 (region connection calculus)¹⁹ beschrieben werden (Battle u. Kolas, 2011).

Das Beispiel im Quelltext 3.10 zeigt ein Polygon. In den letzten Zeilen werden die jeweiligen Stützpunkte als eigene Ressourcen in einer RDF-Collection definiert (Zeilen 8 - 11). Dabei wird auf das bereits bekannte Basic Geo Vocabulary zurückgegriffen. Die Definition des Polygons selbst mit `nggeo:Polygon`, `nggeo:exterior`, `nggeo:LinearRing` und `nggeo:posList` lässt die Struktur der GML Basis erkennen.

¹⁸<http://geovocab.org/>, 06.09.2011

¹⁹vgl. Cohn u.a., 1997

```

1 @prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
2 @prefix ngeo: <http://geovocab.org/geometry#> .
3
4 _:polygon rdf:type ngeo:Polygon ;
5           ngeo:exterior [
6
7               rdf:type ngeo:LinearRing ;
8               ngeo:posList (
9
10                  [ geo:lat "-29"; geo:long "16" ]
11                  [ geo:lat "-28"; geo:long "33" ]
12                  [ geo:lat "-34"; geo:long "27" ]
13                  [ geo:lat "-29"; geo:long "16" ]
14              )
15          ] ;

```

Quelltext 3.10: Polygon im NeoGeo Modell (verändert nach: <http://geovocab.org/doc/neogeo.html>, Kapitel 3.4 Polygon, 10.11.2011).

Die Abbildung 3.4 veranschaulicht das Vokabular von NeoGeo mit den wichtigsten Klassen und Beziehungen. Hier zeigt sich, dass schlussendlich jede Geometrie aus Punkten des Basic Geo Vocabulary zusammengesetzt ist, wobei aber die einzelnen Geometrietypen klar miteinander in Beziehung gesetzt und hierarchisch geordnet werden. Wie auch Battle und Kolas (2011) erkannt haben, liegt aber ein Nachteil

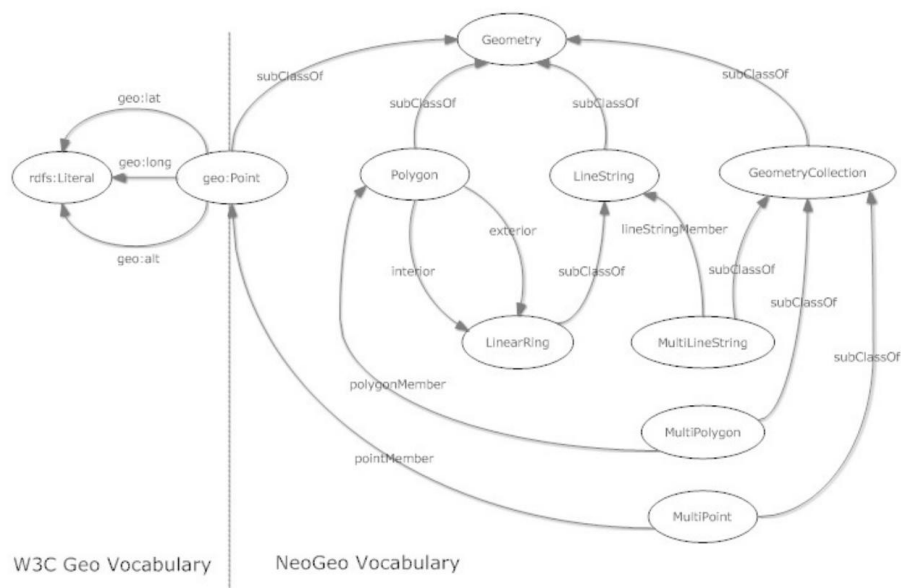


Abbildung 3.4: Darstellung des NeoGeo Vokabulars (Quelle: <http://geovocab.org/doc/neogeo.html>, Kapitel 3. Vocabulary).

genau in der Speicherung der Punkte als RDF-Collection. Dies führt nicht nur zu einem zusätzlichen Aufblähen der Datenmenge, sondern es kann auch der zusätzliche Nutzen in Frage gestellt werden. Zwar kann jeder einzelne Punkt eines Polygons referenziert werden, doch nützt ein einziger Punkt isoliert betrachtet nur wenig. Zudem bereitet die Abfrage von RDF-Listen in SPARQL gewisse Schwierigkeiten (Battle u. Kolas, 2011).

3.2.6 Vektordaten beim USGS

Jegliche Resultate aus den SPARQL-Abfragen sollen grafisch darstellbar sein, weshalb die RDF-Daten mit den Koordinaten angereichert werden sollen. Damit hat der USGS eine klare Anforderung an die Datenkonvertierung gestellt (Usery u. Varanka, 2011). Die Daten wurden erst in das GML-Format und dann mit GML2RDF, einem vom USGS entwickelten Programm, in das RDF-Format N3 konvertiert. Die Attributdaten wurden ausgelesen und als Tripel abgelegt, wohingegen für die Geometriedaten der gesamte GML-Code als String abgelegt wurde. Damit kann durch einfaches Auslesen des GML-Strings die Geometrie dargestellt werden. Zusätzlich werden noch die Koordinaten der Geometrien separat als Tripel abgelegt, wodurch diese direkt mittels einer SPARQL Abfrage aufgerufen werden können. Die topologischen Beziehungen wurden im Voraus, nach dem 9-intersection Model²⁰, berechnet und ebenfalls als Tripel abgelegt (Bulen u. a., 2011).

```

1 @prefix trans:    <http://cegis.usgs.gov/rdf/trans#> .
2 @prefix RoadSegment: <http://cegis.usgs.gov/rdf/trans/RoadSegment#> .
3 @prefix roadClass: <http://cegis.usgs.gov/rdf/trans/roadClass#> .
4 @prefix geom:     <http://cegis.usgs.gov/rdf/geometry#> .
5
6 RoadSegment:_125988612
7   geom:geometryType "LineString"^^<http://www.w3.org/2001/XMLSchema#string>;
8   geom:gml "<gml:featureMember>
9             <ogr:Trans_RoadSegment fid=\"F1235\">
10              <ogr:geometryProperty>
11               <gml:LineString srsName=\"EPSG:4326\">
12                <gml:coordinates>
13                 -108.38,39.89 [...] -108.39,39.90 -108.39,39.90
14                </gml:coordinates>
15              </gml:LineString>
16            </ogr:geometryProperty>
17            <ogr:County_Route>24</ogr:County_Route>
18          </ogr:Trans_RoadSegment>
19        </gml:featureMember>" ;
20   geom:intersects RoadSegment:_125999838 ;
21   geom:touches RoadSegment:_125999838 ;
22   trans:countyRoute "24"^^<http://www.w3.org/2001/XMLSchema#string>;
23   trans:geometryProperty "LINESTRING (-108.38 39.89, [...] -108.39 39.90)";

```

Quelltext 3.11: Linie beim USGS (verändert nach: http://usgs-ybother.srv.mst.edu/ontology/CO_14050006.zip, 11.11.2011).

Quelltext 3.11 stellt einen Strassenabschnitt dar, woraus die vom USGS verwendete RDF-Struktur gut ersichtlich wird. Das Prädikat `geom:gml` hat als Objekt den String mit den gesamten Informationen (Geometrie und sämtliche Attribute) im GML-Format (Zeilen 8 - 19). Die reine Geometrie ist unter `trans:geometryProperty` abgelegt (Zeile 23). In diesem Beispiel sind zwei Beziehungen `geom:intersects` (Zeile 20) und `geom:touches` (Zeile 21) auszumachen. Aufgrund der Übersichtlichkeit wurde der GML-String auf mehrere Zeilen verteilt, was aber an der Grundstruktur nichts ändert.

²⁰Für mehr Informationen: Egenhofer, M.J., and Herring, J.R., 1991, Categorizing Binary Topological Relations between Regions, Lines, and Points in Geographic Databases. <http://www.spatial.maine.edu/~max/9intReport.pdf>.

3.3 Umformung von Rasterdaten

Ansätze oder Stossrichtungen zur Verwendung von Rasterdaten als Linked Data sind, im Gegensatz zu den Vektordaten, nur spärlich vorhanden. Obwohl den Rasterdaten nicht weniger Bedeutung zugemessen werden sollte.

Anders als das Vektormodell, eignen sich Raster besonders zur Repräsentation kontinuierlicher Daten, wie beispielsweise Temperaturen oder Höhen (Burrough u. McDonnell, 2006). In einem Raster wird der abgebildete Raum in regelmässige Zellen (oft Quadrate) aufgeteilt. Diesen Pixeln werden Attribute zugewiesen, wodurch die geographische Variation ausgedrückt wird. Die Variation innerhalb einer Zelle geht aber verloren. Jede Zelle erhält nur einen Wert zugewiesen (Longley u. a., 2011). Das heisst, Objekte sind nur approximiert abgegrenzt und der Zugriff auf sie ist erschwert, da sie nicht explizit durch die Datenstruktur definiert werden (Usery u. Varanka, 2011).

Raster sind in diesem Sinne, ohne weitere Verarbeitung, für den Menschen eher arm an Semantik, da zum einen keine konkreten Objekte benannt werden und zum anderen topologische Beziehungen fehlen. Beispielsweise ein kategoriales Raster, das die Bodenbedeckung repräsentiert, bietet einen breiten Interpretationsspielraum und stellt verschiedenste Objekte dar (approximiert zur Realität). So kann eine einzelne Zelle, die ein Moor repräsentiert, unterschiedlichen Objekten zugeordnet werden, wie etwa Schutzgebieten oder landwirtschaftlich ungenutzten Flächen usw. - je nachdem was von Interesse ist und dementsprechend an Objekten ausgeschieden werden soll. Eine einzelne Zelle ist für Menschen weniger bedeutend als eine Anhäufung von Zellen, die beispielsweise ein bestimmtes Mooregebiet repräsentieren. Aufgrund der approximierten Ausscheidung von Objekten und dem breiten Interpretationsspielraum fehlt eine gewisse Eindeutigkeit von Geometrie und Attributen, wie sie beim Vektormodell zu finden ist. Deshalb macht es Sinn zu entscheiden, was genau von den Rasterdaten (z. B. konkrete Objekte, Teile eines Rasters usw.) abgelegt werden sollte.

Eine einfache Option wäre die Beschränkung auf die Umformung von Metadaten in RDF, wie Auflösung, Zellgrösse, Datenbeschreibung usw. Damit wäre die Speicherung des eigentlichen Rasters nicht mehr von Belang und dennoch könnte auf grundsätzliche Informationen zum Raster zugegriffen werden.

Liegt das Interesse auf einzelnen Objekten, die im Rasterdatensatz enthalten sind, bestünde die Möglichkeit diese konkret zu beschreiben. Sie könnten aus dem Raster extrahiert und als eigene Objekte gespeichert werden. Jedoch stellt sich bei diesem Vorgehen grundsätzlich die Frage nach einer korrekten und eindeutigen Abgrenzung. Beispielsweise wird ein Berg in einem Höhenmodell nicht explizit ausgewiesen und muss anhand bestimmter Kriterien festgelegt werden²¹ (Usery u. Varanka, 2011).

Eine Alternative zur Speicherung von Objekten könnte die Angabe der Position eines Objektes innerhalb des Rasters sein, beispielsweise durch die Aufzählung der zusammengehörenden Zellen oder durch Angabe von räumlichen Koordinaten. Usery und Varanka (2011) vom USGS stellen in ihrer Veröffentlichung "Design and Development of Linked Data from The National Map" einen Ansatz vor, der in die genannte Richtung tendiert. Sie bestimmen und extrahieren die relevanten Objekte aus den Rasterdaten, behalten jedoch die Rastermatrix als geometrische Grundform bei. Dazu wird ein einzelnes Pixel oder eine Gruppe von Pixeln als geometrische Grundlage gewählt. Je nach Objekt setzt sich dieses aus einem oder mehreren solchen Einheiten zusammen. Ein Punktobjekt entspricht einem solchen Pixel, während ein Flächenobjekt aus mehreren besteht. Das heisst, die Objekte werden durch die entsprechend zugehörigen Rasterzellen repräsentiert. Welche Objekte, zugehörige Attribute und Beziehungen identifiziert werden sollen, gibt ein sogenanntes "Ontology Design Pattern" (ODP) vor. Diese Objekte werden in RDF im GML Coverage

²¹ vgl. Diplomarbeit: Derungs, Curdin: Was ist ein Berg? Wo ist ein Berg? Zürich: Geographisches Institut der Universität Zürich, 2007.

Format abgelegt, als sogenanntes gml:Grid Element (Usery u. Varanka, 2011). Ein solches Grid kann beispielsweise durch Punkte (Zentrum eines Pixels) oder auch durch die Pixel selbst repräsentiert werden (Portele, 2007). Der USGS bedient sich des Letzteren, jedoch besteht die Einschränkung, dass Pixel nur in einer rechteckigen Ausdehnung (*minimum bounding rectangle*) gespeichert werden können (Usery u. Varanka, 2011). Die genannten Möglichkeiten zur Rasterspeicherung könnten auf den gesamten Rasterdatensatz erweitert werden. Also weg von den einzelnen Objekten, hin zur ganzheitlichen Speicherung des Rasters.

3.4 GeoSPARQL - Abfragesprache und Format

Wie erläutert bestehen bereits einige Ansätze zur Repräsentation von Geodaten (besonders von Vektordaten) als RDF, welche aber bis anhin keine allgemeine Verbindlichkeit erreichten. Das OGC unternimmt mit GeoSPARQL einen weiteren Anlauf einen Standard einzuführen. Dieser befindet sich auf dem Weg vom W3C als Standard erlassen zu werden.²² Aus dem Namen GeoSPARQL könnte man schliessen, es handle sich ausschliesslich um eine räumlich erweiterte Version der SPARQL-Abfragesprache, was nur teilweise zutrifft. Neben der Funktion als Abfragesprache ist GeoSPARQL auch ein Datenformat zur Repräsentation räumlicher Daten (Battle u. Kolas, 2011). GeoSPARQL setzt sich aus drei zentralen Komponenten zusammen (Perry u. Herring, 2011):

- Das **Vokabular** definiert Klassen und Eigenschaften zur Beschreibung von räumlichen Daten, sowie um einfache SPARQL-Abfragen zu formulieren (Perry u. Herring, 2011). Die Ontologie basiert, wie bei einigen vorangegangenen Vorschlägen, auf dem GML Simple Features Modell, mit einigen Anpassungen an RDF. Die wichtigsten Klassen bilden `geo:SpatialObject` mit den Unterklassen `geo:Feature` und `geo:Geometry`, wobei die beiden Letzteren durch die Beziehung `geo:hasGeometry` verbunden werden (Battle u. Kolas, 2011). Beispielsweise das Jagdbanngebiet "Dixence" existiert in der Realität und ist ein `geo:Feature`. Dieses konzeptuelle Objekt hat auch eine festgelegte Geometrie, die durch `geo:Geometry` als Literal in GML oder WKT repräsentiert wird (Battle u. Kolas, 2011).
- Die **Filterfunktion** für GeoSPARQL-Abfragen soll räumliche Abfragen ermöglichen, womit gewisse Geoprozessierungsmöglichkeiten, wie Verschneidungen (*intersection*) oder Pufferanalysen (*buffer*), eingebunden werden (Perry u. Herring, 2011).
- Regeln zur **Transformation** von Abfragen ermöglichen GeoSPARQL attributive als auch geometrische Abfragen durchzuführen. Räumliche Analysen können aufgrund der Topologie, beispielsweise basierend auf RCC8, aber auch durch geometrische Berechnungen durchgeführt werden. Für das Erstere wurden basierend auf den Geometrien die räumlichen Beziehungen im Voraus ermittelt und als topologische Informationen in RDF abgelegt, wohingegen für das Letztere lediglich die Geometrien vorliegen müssen. In diesem Fall ermöglichen die Transformationsregeln eine Erweiterung der attributiven zur geometrischen Abfrage, bei gleich bleibender Spezifikation. So beinhaltet beispielsweise `ogc:within` beide Abfragetypen (Perry u. Herring, 2011). Zur Beschreibung von topologischen Beziehungen können die Vokabulare von Simple Features vom OGC, Egenhofers 9-intersection Modell und RCC8 verwendet werden. Jedoch wird die Möglichkeit für deren Benutzung durch die verwendete Technologie (z. B. Triple Store) bestimmt (Battle u. Kolas, 2011).

Die Tabelle 3.1 zeigt die möglichen räumlichen Beziehungen in GeoSPARQL und deren Äquivalente im RCC8 bzw. 9-intersection (Egenhofer) Modell.

²²<http://www.opengeospatial.org/projects/groups/geosparqlswg>, 06.10.2011

GeoSPARQL	RCC8	Egenhofer
ogc:equals	EQ	equal
ogc:disjoint	DC	disjoint
ogc:intersects	\neg DC	\neg disjoint
ogc:touches	EC	meet
ogc:within	NTPP + TPP	inside + coveredBy
ogc:contains	TPPi + NTTPi	contains + covers
ogc:overlaps	PO	overlap

Tabelle 3.1: Die räumlichen Beziehungen in GeoSPARQL und die Äquivalente in RCC8 bzw. 9-intersection (Quelle: Perry und Herring, 2011, S. 18).

Der Quelltext 3.12 zeigt die Definition eines Polygons im GeoSPARQL Modell im N3-Format.

```

1 <http://www.example.org/Points_of_Interest/Geometrie_Bern>
2   <http://www.opengis.net/rdf#asWKT>
3   "<http://www.opengis.net/def/crs/EPSG/0/4326> Polygon((5807 888,
4     [...] 5807 888))"^^<http://www.opengis.net/rdf#WKTLiteral> .

```

Quelltext 3.12: Polygon im GeoSPARQL Modell (verändert nach: Battle u. Kolas, 2011, S. 8).

Die dritte und vierte Zeile (Objekt) beschreiben die Geometrie des Subjektes `http://www.example.org/Points_of_Interest/Geometrie_Bern` der ersten Zeile. Dem String wird der Datentyp `WKTLiteral` zugeordnet. Innerhalb des Strings wird zuerst das Referenzsystem definiert, gefolgt von der Angabe des Geometrietypen und der Koordinatenliste (Battle u. Kolas, 2011).

3.5 Diskussion

Das Kapitel stellte Möglichkeiten zur Speicherung von räumlichen Daten im RDF-Format vor. Im Bereich der Vektordaten existieren einige mehr oder weniger ausgereifte Ansätze, die in unterschiedlichen Projekten verwendet werden, jedoch hat sich noch kein Format durchgesetzt. GeoSPARQL ist ein weiterer Versuch, einen Standard für Geodaten zu definieren.

Den meisten Ansätzen liegt das Prinzip zu Grunde, Vektordaten in einer Koordinatenliste (String) abzulegen, die einer bestimmten Syntax folgt und die Stützpunkte der Geometrie beinhaltet. Dies kann durchaus als vorteilhaft angesehen werden. Die Etablierung eines allgemeinen Standards würde nicht nur das Datenformat vereinheitlichen, sondern auch den Austausch vereinfachen. Zusätzlich greift GeoSPARQL zur Repräsentation der Geometrie auf die bereits etablierten OGC-Standards WKT und GML zurück (Battle u. Kolas, 2011). Unterstützend wirkt dabei die freie Zugänglichkeit der vom OGC festgelegten Standards, sowie die breite Verfügbarkeit von kostenlosen Applikationen, die auf diese zurückgreifen.²³ Dies fördert den Linked Data Gedanken, Daten einem möglichst breiten Publikum zugänglich zu machen.

Ein weiterer Vorteil ist die direkte Einbindung der Geometriedaten, womit sie auch direkt zu anderen Ressourcen räumlich verknüpft werden können (*spatial join*). Falls ausserdem einfache Geoprozessierungen implementiert werden, können ohne GIS oder ähnliche Systeme räumliche Fragen beantwortet werden.

Hingegen existieren bei den Rasterdaten noch keine verbreiteten Ansätze zu deren Umformung und Speicherung. Der USGS hat sich damit befasst und einen eigenen Weg entwickelt. Sie legen die Geometriedaten inklusive den extrahierten Rasterobjekten als String im GML-Format ab. Dazu muss im Voraus klar sein,

²³<http://www.opengeospatial.org/ogc>, 08.11.2011

welche Informationen aus dem Raster extrahiert und als eigene Objekte angeboten werden sollen. Die starke Abstützung auf das GML-Format, indem die gesamte GML-Definition als Tripel abgelegt wird, scheint vielversprechend und interessant. Durch Auslesen des GML-Strings können die Daten grafisch dargestellt und beispielsweise in Kombination mit einem GIS verwendet werden.

3.6 Fazit (Forschungslücken)

Die Idee von Linked Data einen globalen Datenraum (*Web of Data*) zu schaffen, hat in den letzten Jahren Einzug in das Web gehalten. So sind nicht mehr nur Dokumente miteinander verknüpft, sondern auch Daten (Bizer u. a., 2009a; Heath u. Bizer, 2011). Heute existieren einige Projekte, die erfolgreich Daten im RDF-Format anbieten und diese mit anderen Datenquellen verknüpfen (vgl. Abbildung LOD Cloud im Abschnitt 2.7.1). Dabei machen Daten mit einem geographischen Hintergrund einen beachtlichen Teil aus. Sie stellten im September 2011 gut einen Fünftel der Tripel (ca. 6.1 Milliarden Tripel) in der LOD Cloud.²⁴ Dies zeigt, dass räumlich referenzierte Daten eine wesentliche Rolle spielen und daher untersucht werden sollte, wie sie in den Linked Data Kontext integriert werden können.

Ein wichtiges Designprinzip von Linked Data ist die Beschreibung der zu repräsentierenden Ressource im RDF-Format (Heath u. Bizer, 2011). Damit im Zusammenhang steht die Forderung, dass “nützliche” Informationen in RDF angeboten werden sollen und dadurch direkt über Linked Data zugreifbar sind (Berners-Lee, 2009). Somit sollten auch Geodaten nicht nur mit fremden Datenquellen verknüpft, sondern mit den zugehörigen Informationen angereichert werden. Je nach Auffassung gehören dabei sowohl die Sachattribute als auch die Geometrien dazu. In diesem Kapitel wurden diverse Ansätze und Probleme erläutert, wie Daten mit räumlichem Bezug im RDF-Format repräsentiert werden können. Bis zum heutigen Zeitpunkt sind unterschiedliche Ansätze im Rahmen konkreter Projekte implementiert worden. Jedoch bestehen teilweise noch grössere, ungeklärte Lücken:

- Eine zentrale Thematik ist die **Konvertierung von Geometrien in RDF**. Im Bereich der Vektordaten scheinen die grundsätzlichen Fragen bezüglich der Umformung in RDF mehr oder weniger geklärt zu sein, bis auf die Tatsache eines fehlenden Standards zur Repräsentation. Hingegen bei den **Rasterdaten** sind erst wenige Ansätze vorhanden und es besteht noch Forschungsbedarf. Bisher hat die Rasterproblematik in der Literatur und der Community wenig Erwähnung gefunden.
- Ein weiteres wichtiges Aufgabengebiet neben der Datenkonvertierung, ist die **Implementierung von räumlichen Beziehungen**, welche mit RDF spezifischen Abfragesprachen (wie SPARQL) abgefragt werden können. Momentan werden solche Beziehungen vor allem durch die explizite Angabe der Topologie in Form von Tripeln und/oder in der Ontologie selbst definiert. Dies birgt die Einschränkung, dass räumliche Abfragen nur in dem Rahmen möglich sind, wie der Datenanbieter dies vorgesehen hat. Beispielsweise die Prüfung, ob eine Strasse durch ein bestimmtes Gebiet verläuft, ist nur möglich, wenn diese Information explizit extrahiert und bereitgestellt wurde (Battle u. Kolas, 2011).
- Die **Durchführung einfacher räumlicher Operationen** (Puffer, Punkt in Polygon usw.) ist zur Zeit praktisch nicht implementiert und gestaltet sich als rechnerisch ineffizient, da deren Entwicklung noch in den Kinderschuhen steckt (Heath u. Goodwin, 2011; Goodwin u. a., 2008). Damit

²⁴<http://www4.wiwi.fu-berlin.de/locloud/state/>, 07.12.2011

im Zusammenhang stellt sich auch die Frage nach einer **räumlichen Indexierung**, um räumliche Abfragen effizient durchführen zu können. Auf diesem Gebiet werden bereits erste Lösungen angeboten, die aber nur beschränkt verwendet werden können (z. B. nur für Punktgeometrien) (Battle u. Kolas, 2011). Insgesamt sind die Technologien aus dem Linked Data Umfeld von den Möglichkeiten, die räumliche Operationen oder ein GIS bieten, noch weit entfernt und es stellt sich die berechnete Frage, ob dies erstrebenswert sein soll.

Generell müssten die **Anforderungen an Geodaten im Linked Data Kontext** spezifiziert werden. Das heisst, sollen aufwändige Abfragen mit räumlichen Prozessierungen möglich sein, oder soll eher auf die Existenz der Datensätze und deren Metadaten verwiesen werden, ohne Zugriff auf die enthaltenen Objekte zu haben? Anhand der vorgestellten Projekte im Abschnitt 3.1 wurde ersichtlich, dass momentan der Trend in der Bereitstellung der Geodaten inklusive Sachattributen und Geometrien im RDF-Format liegt.

Im Rahmen dieser Arbeit soll ein Beitrag dazu geleistet werden, wie Geodaten in Linked Data eingebunden werden können.

4.1 Methodisches Vorgehen

Den Rahmen zur Entwicklung von Ansätzen zum Umgang mit Geodaten im Linked Data Kontext bilden zu Beginn definierte **Use Cases (Anwendungsfälle)**. Sie sind die Richtschnur dieser Arbeit. Das Ziel liegt in der erfolgreichen Beantwortung dieser Anwendungsfälle, unter Verwendung der Technologien aus dem Linked Data Bereich. Dementsprechend orientieren sich auch die konzeptionellen Grundlagen, die Modellierungen, die Datenprozessierungen und die Resultatgewinnung an diesen Anwendungsfällen. Einhergehend mit der Implementierung der Use Cases werden grundsätzliche Fragestellungen aus dem Linked Data Bereich untersucht, wie beispielsweise die Umformung von Geodaten in das RDF-Format. Dabei werden die bereits bestehenden Konzepte angewendet, sowie ergänzt und weiterentwickelt.

Die Use Cases stehen inhaltlich in direktem Zusammenhang zur Thematik der Eidgenössischen Jagdbanngebiete in der Schweiz und wurden von einem Experten des Bundesamts für Umwelt (BAFU) verifiziert. Das heisst, sie liegen im Interesse potentieller Benutzer und können somit in der Praxis Verwendung finden. Insgesamt wurden fünf verschiedene Anwendungsfälle definiert, die in zwei Gruppen, “Einfache Use Cases” und “Komplexe Use Cases”, aufgeteilt werden.

Die **einfachen Use Cases** lauten:

- **Einfacher Use Case 1 “Attribute abfragen”:**

Welche Jagdbanngebiete gehören einer bestimmten Schutzkategorie an bzw. beinhalten Teilgebiete, die dieser Schutzkategorie zugeordnet werden können?

- **Einfacher Use Case 2 “Aggregation”:**

Wie gross ist die Fläche aller Jagdbann- bzw. Teilgebiete, die einer bestimmten Schutzkategorie angehören?

Die beiden Anwendungsfälle stellen unkomplizierte Anforderungen dar und sollten jeweils mit einer einzigen Abfrage, ohne zusätzliche Schritte, beantwortet werden können. Die zu ermittelnden Resultate basieren auf bereits vorhandenen Werten in den Ursprungsdaten. Dadurch ist nur wenig Interpretationsspielraum gegeben und es werden eindeutige Resultate erwartet. Beispielsweise ist die Fläche eines Jagdbanngebietes festgelegt und wird nicht neu berechnet. So kann der Use Case “Aggregation” durch Summieren der betroffenen Flächen gelöst werden.

Die **komplexen Use Cases** lauten:

- **Komplexer Use Case 1 “Waldflächen und Höhestufen”:**

Wie verteilen sich die Waldflächen einzelner oder sämtlicher Jagdbanngebiete auf die verschiedenen ökologischen Höhestufen, unter Berücksichtigung der regionalen Variabilität dieser Höhestufen innerhalb der Schweiz?

Der Anwendungsfall benötigt neben den Geodaten noch zusätzliche Informationen. So sind die zu verwendenden Höhestufen (z. B. collin, montan usw.) und die Regionen mit der genannten Variabilität (z. B. Ostalpen, Mittelland usw.) zu definieren. Je nach Definition dieser Parameter fallen die Ergebnisse unterschiedlich aus.

Indem eines oder mehrere Jagdbanngebiete gleichzeitig abgefragt werden können, macht die Verwendung der Datensätze in unterschiedlichen Massstäben Sinn. Ist beispielsweise nur ein Gebiet gefragt, sollen grossmassstäbige und bei mehreren Gebieten eher kleinmassstäbige Daten verwendet werden.

- **Komplexer Use Case 2 “Betroffene Gemeinden”:**

Auf welchen Gemeinden liegt ein Jagdbanngebiet? Wie heissen die Gemeinden gemäss GeoNames und das zugehörige Jagdbanngebiet?

Bei diesem Anwendungsfall steht die Verknüpfung mit externen Datenquellen, in diesem Fall GeoNames, im Vordergrund. Das heisst, die entsprechenden GeoNames Einträge, welche eine solche Gemeinde repräsentieren, sollen gefunden, als Tripel abgelegt und zu den anderen im Triple Store vorhandenen Ressourcen in Beziehung gesetzt werden. Da GeoNames die Daten im RDF-Format anbietet und ebenfalls zu anderen Quellen verknüpft hat, wird durch deren Einbindung der Zugang zum Datenraum von Linked Data ermöglicht.

- **Komplexer Use Case 3 “Ähnliche Gebiete”:**

Können Teilgebiete bzw. ganze Jagdbanngebiete ausgeschieden werden, welche Ähnlichkeiten zu einem bestimmten Teilgebiet bzw. Jagdbanngebiet aufweisen?

Der dritte komplexe Anwendungsfall weist eine weite Umsetzungsspanne auf, da die Menge der Parameter, anhand der die Ähnlichkeit bestimmt wird, frei wählbar ist. In erster Linie sollen die bereits verwendeten Datensätze benützt und allenfalls durch weitere Datensätze ergänzt werden.

Zur Beantwortung der komplexen Use Cases sind voraussichtlich jeweils mehrere Schritte notwendig, bei denen teilweise zusätzliche und neue Daten generiert werden. Der Spielraum bezüglich Detaillierungsgrad und verwendeten Mitteln ist innerhalb der komplexen Anwendungsfälle wesentlich grösser als bei den einfachen. Beispielsweise beim Use Case “Ähnliche Gebiete” ergeben sich, abhängig von den Parametern zur Bestimmung der Ähnlichkeit, unterschiedliche Ergebnisse.

Im folgenden Abschnitt 4.2 wird eine kurze Einführung zu den wichtigsten Datensätzen gegeben, die als Ausgangsdaten bezeichnet werden. Die darauffolgenden Kapitel beschäftigen sich mit der Beantwortung der Anwendungsfälle und erläutern die dabei zugrundeliegenden Ansätze. Dazu werden erst die zwei Ansätze VKE (Vektorisierte Kleinste Einheiten) und ARO (Ausgegliederte Räumliche Operationen) beschrieben. Anschliessend folgt die Erläuterung der konkreten Umsetzung der Anwendungsfälle basierend auf dem Ansatz ARO. Zum Schluss werden die wichtigsten Erkenntnisse diskutiert und ein Ausblick für zukünftige Forschungen gegeben.

4.2 Verwendete Datensätze

Im Rahmen dieser Arbeit und im Zusammenhang mit den Anwendungsfällen werden verschiedene Geodatenätze verwendet. Es handelt sich dabei um Vektor- und Rasterdaten, die von amtlichen Stellen bereitgestellt werden. Im Zentrum steht der Datensatz der Jagdbanngebiete. Dieser wird um weitere Datensätze ergänzt, die im Kontext von Jagdbanngebieten von Interesse sind. Da es sich bei den Jagdbanngebieten um Schutzgebiete für Wildtiere handelt, liegt ein besonderes Augenmerk auf deren Lebensraum. Ein Lebensraum kann durch die landschaftliche Beschaffenheit beschrieben werden, welche beispielsweise durch die Datensätze VECTOR25 bzw. VECTOR200 (Landschaftsmodelle) sowie DHM25 bzw. DHM RIMINI (Höhenmodelle) widerspiegelt wird.

4.2.1 Jagdbanngebiete

Der vom Bundesamt für Umwelt (BAFU) herausgegebene Polygondatenatz der Jagdbanngebiete (vgl. Abbildung 4.1), wurde zuletzt im Jahre 2010 bearbeitet. Er wurde für die Arbeit vom Datenzentrum Natur und Landschaft (DNL) der WSL bezogen, kann aber auch frei von der Webseite des BAFU heruntergeladen werden.¹ Die im Datensatz enthaltenen Objekte (Jagdbanngebiete) wurden ab der Landeskarte manuell digitalisiert (BAFU, 2010). Die Objekte repräsentieren oft nur ein Teilgebiet und in wenigen Fällen ein gesamtes Jagdbanngebiet. Jagdbanngebiete werden anhand des Attributes “Schutzkategorie” in Teilgebiete mit unterschiedlichen Schutzbestimmungen aufgeteilt: Der Wert “10” steht für Gebiete mit integralen Schutzbestimmungen, “20” für Gebiete mit partiellen Schutzbestimmungen und “30” für den Wildschadenperimeter (BAFU, 2010).

Der Datensatz ist vollständig und mit 113 Objekten eher klein und übersichtlich. Dadurch qualifiziert er sich als idealer Testdatensatz für die Erprobung der entwickelten Ansätze.

4.2.2 Exkurs: Geschichte der Jagdbanngebiete

Die Huftierbestände in der Schweiz wurden im Laufe des 19. Jahrhunderts stark reduziert, teilweise sogar ausgerottet. Dafür verantwortlich waren ein hoher Jagddruck und der schlechte Zustand der Wälder.² Um dieser Entwicklung entgegenzutreten wurde 1874 die Möglichkeit des Bundes national verbindliche Bestimmungen zur Jagd festzulegen in der Bundesverfassung verankert. Gleich in den zwei darauffolgenden Jahren (1875/1876) wurden nationale Gesetze zu Jagd, Fischerei und Vogelschutz verabschiedet. Bereits mit diesem ersten eidgenössischen Jagdgesetz wurden nicht nur die Schonzeiten für das Wild vereinheitlicht und verlängert, sondern zusätzlich die Gebirgskantone zur Ausscheidung von einem oder mehreren sogenannten Bannbezirken verpflichtet. Zudem verbesserten sich die Lebensbedingungen für das Wild durch zunehmenden Schutz der Gebirgswälder (Bundi, 2007). Trotz dieser Regelungen konnte der Rückgang des Wildes nicht gestoppt werden, wofür nicht zuletzt auch die Wilderei verantwortlich war (Bundi, 2007). Deshalb wurde 1891 eine Revision des Jagdgesetzes angestrebt, welche aber erst 1902 (mit angepasstem Inhalt) den politischen Durchbruch schaffte. Inhaltlich wurden schärfere Strafbestimmungen für Gesetzesübertretungen, mehr Kompetenzen für Bund und Kantone und Verbote für gewisse Jagdmethoden (z.B. Selbstschussanlagen, Giftlegen) eingeführt. Diese Änderungen führten schlussendlich zu dem gewünschten Wildbestand und bahnten den Weg für die Wiederansiedlung von ausgerotteten oder abgewanderten Arten. So konnte beispielsweise der Steinbock im Jahre 1911 nach mehreren vorgängig

¹ <http://www.bafu.admin.ch/schutzgebiete-inventare/07843/index.html?lang=de>, 18.08.2011

² <http://www.bafu.admin.ch/schutzgebiete-inventare/07843/index.html?lang=de>, 19.12.2011

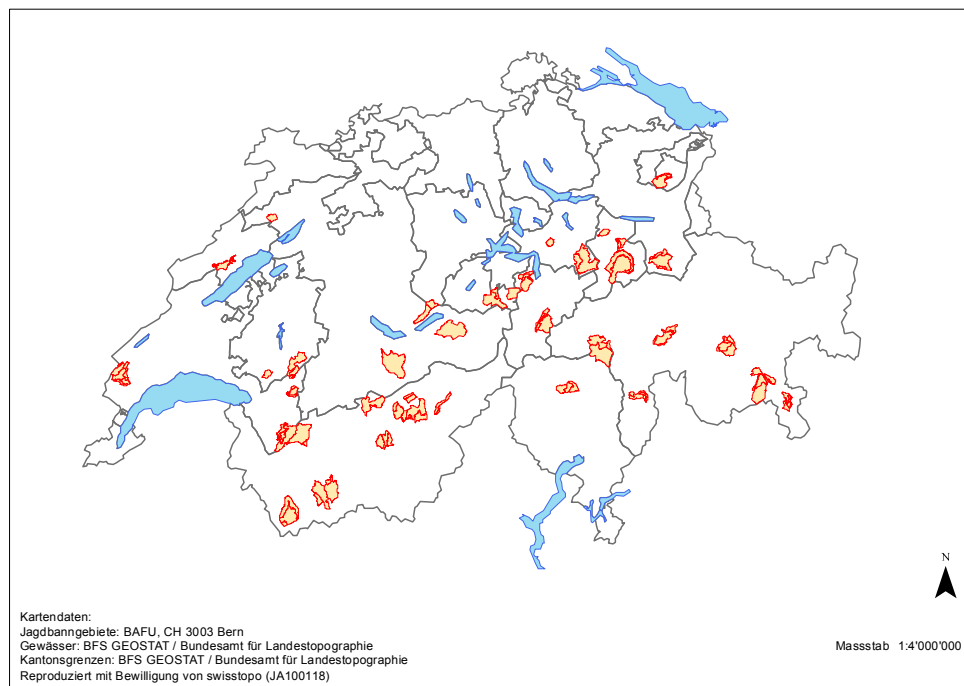


Abbildung 4.1: Die 41 Eidgenössischen Jagdbanngebiete der Schweiz.

fehlgeschlagenen Versuchen erfolgreich wiederangesiedelt werden (Bundi, 2007).

4.2.3 Exkurs: Jagdbanngebiete heute

Heute (Stand 2010) existieren insgesamt 41 Jagdbanngebiete mit einer Gesamtfläche von ca. 150'900 Hektaren. Sie dienen gegenwärtig weniger dem ursprünglichen Ziel der Anhebung der Huftierbestände, sondern übernehmen vor allem die Funktion des Arten- und Lebensraumschutzes. Zusätzlich liefern die Jagdbanngebiete Informationen zur natürlichen Entwicklung der Huftierpopulation. Die Jagd ist aber in diesen Gebieten nicht gänzlich verboten, denn in manchen gibt es einen Populationsüberschuss, der bejagt werden kann.³ Die Huftierbestände sind stets den örtlichen Verhältnissen anzupassen, d. h. eine natürliche Alters- und Geschlechtsstruktur, sowie Berücksichtigung der Interessen von Natur-, Landschaftsschutz und Landwirtschaft sind einzuhalten. Daher dürfen die Kantone jagdliche Regulierungsmassnahmen treffen. Ein Jagdbanngebiet kann in unterschiedliche Schutzzonen unterteilt sein. In integral geschützten Gebieten werden nur in Ausnahmefällen Massnahmen angeordnet, während in partiell geschützten Gebieten regelmässig die Bestände von Rehen, Gämsen, Rothirschen und Wildschweinen reguliert werden.⁴ Zusätzlich können Gebiete ausserhalb der Jagdbanngebiete bestimmt werden, in welchen Wildschäden vergütet werden (Wildschadenperimeter).⁵

Das Inventar jedes Jagdbanngebietes umfasst eine kartographische Darstellung des Perimeters, eine Beschreibung des Gebietes, sowie das Schutzziel und zugehörige Massnahmen bezüglich Arten-, Biotopschutz und jagdlichen Regulierungen.

³ <http://www.bafu.admin.ch/schutzgebiete-inventare/07843/index.html?lang=de>, 11.08.2011

⁴ http://www.admin.ch/ch/d/sr/922_31/a9.html, 11.08.2011

⁵ http://www.admin.ch/ch/d/sr/922_31/a2.html, 11.08.2011

4.2.4 Landschaftsmodelle VECTOR25 / VECTOR200

Das digitale Landschaftsmodell VECTOR25 (VEC25) basiert auf der Landeskarte 1:25'000 der Schweiz und repräsentiert natürliche sowie künstliche Landschaftsobjekte im Vektorformat. Insgesamt sind neun thematische Ebenen enthalten, wie der Tabelle 4.1 entnommen werden kann. Die Objekte werden durch Lage, Nachbarschaftsbeziehungen (Topologie), Form und weitere Sachattribute beschrieben (swisstopo, 2007).

Thematische Ebene	Beschreibung
Strassennetz	Strassen- und Wegnetz
Eisenbahnnetz	Eisenbahnnetz
Übriger Verkehr	Fähren, Seilbahnen usw.
Gewässernetz	Gewässerachsen und Uferlinien
Primärflächen	Primäre Bodenbedeckung (Wald, See usw.)
Gebäude	Diverse Gebäudearten
Hecken und Bäume	Diverse Objektarten der Vegetation
Anlagen	Künstliche Areale und Anlagen
Einzelobjekte	Diverse künstliche Objekte

Tabelle 4.1: Thematische Ebenen von VECTOR25 (Quelle: swisstopo, 2007, S. 3).

Im Rahmen dieser Arbeit wurde die Ebene der Primärflächen, und davon die Flächenobjekte (Polygone) verwendet. Sie beschreibt die primäre topographische Bodenbedeckung, wozu auch die Siedlungsgebiete, aber keine Einzelgebäude gehören. Die Grenzen zwischen den Flächen werden zusätzlich durch Linienobjekte mit eigenen Objektarten (z. B. Flussufer links, Flussufer rechts) dargestellt. Jedoch wurden diese weggelassen, weshalb im Folgenden mit dem Begriff Primärflächen nur die Flächenobjekte bezeichnet werden. Die Flächenobjekte bilden ein lückenloses Flächennetz, sind aber klar voneinander abgegrenzt, weshalb es keine Überschneidungen gibt (swisstopo, 2007). In der Tabelle 4.2 sind die vorhandenen Objektarten aufgeführt.

ObjectVal	Beschreibung(Objektart)	ObjectVal	Beschreibung(Objektart)
Z_BaumS	Baumschule	Z_Fels	Fels
Z_Fluss	Fluss	Z_Gebue	Gebüsch
Z_GerGeb	Geröll mit Busch	Z_GerGle	Geröll auf Gletscher
Z_Geroel	Geröll	Z_GerWa	Geröll in Wald
Z_GerWaO	Geröll in offenem Wald	Z_Glet	Gletscher
Z_GsPist	Graspiste	Z_HaPist	Piste mit Hartbelag
Z_KiGrub	Kiesgrube	Z_LeGrub	Lehmgrunbe
Z_ObstAn	Obstanlage	Z_Reben	Reben
Z_See	See	Z_Siedl	Siedlung
Z_StauDa	Staudamm	Z_StauMa	Staumauer
Z_SteBru	Steinbruch	Z_SumGeb	Sumpf und Gebüsch
Z_Sumpf	Sumpf	Z_SumWa	Sumpf in Wald
Z_SumWaO	Sumpf in offenem Wald	Z_Uebrig	Übriges Gebiet
Z_Wald	Wald	Z_WaldOf	Wald offen

Tabelle 4.2: Mögliche Objektarten der Flächenobjekte der Ebene Primärflächen von VECTOR25 (Quelle: swisstopo, 2007, S. 19).

Das digitale Landschaftsmodell VECTOR200 (VEC200) gibt ebenfalls die natürlichen und künstlichen Landschaftsobjekte im Vektorformat wieder, allerdings mit einem hohen Generalisierungsgrad. Die Genauigkeit entspricht der Landeskarte 1:200'000. VECTOR200 enthält sieben thematische Ebenen (siehe

Tabelle 4.3) und beschreibt die Objekte ebenfalls durch Lage, Nachbarschaftsbeziehungen (Topologie), Form und weitere Sachattribute (swisstopo, 2011).

Topics	Beschreibung
Transportation	Strassen- und Eisenbahnnetz, weitere Elemente im Zusammenhang mit dem Verkehr
Hydrography	Gewässerachsen und Seen
Landcover	Primäre Bodenbedeckung (Wald, Siedlung, usw.)
Buildings	Gebäude
Miscellaneous	Verschiedenes
Names	Namen
Administrative Boundaries	Administrative Grenzen (Landesgrenzen bis Gemeindegrenzen), Nationalparks

Tabelle 4.3: Thematische Ebenen von VECTOR200 (Quelle: swisstopo, 2011, S. 3).

Das Gegenstück aus VECTOR200 zu den Primärflächen aus VECTOR25, ist die thematische Ebene Landcover. Sie besitzt ausschliesslich Flächenobjekte, die die topographische Bodenbedeckung beschreiben. Die Flächen bilden wie bei VECTOR25 ein Flächennetz, können sich aber teilweise überlappen (z. B. Wald und Sumpf) (swisstopo, 2011, S. 35).

Wie aus Tabelle 4.4 ersichtlich, hat die Ebene Landcover aus VECTOR200 wesentlich weniger Objektarten als die Primärflächen aus VECTOR25. In Abbildung 4.2 ist ein visualisierter Ausschnitt der Primärflächen aus VECTOR25 (links) und VECTOR200 (rechts) zu sehen.

Mögliche Werte	Beschreibung	Mögliche Werte	Beschreibung
Wald	Wald	Fels	Fels
Geroell	Geröll	Gletscher	Gletscher
See	See	Stausee	Stausee
Siedl	Siedlung	Stadtzentr	Zentren grösserer Städte
Sumpf	Sumpf	Obstanlage	Obstanlage
Reben	Reben		

Tabelle 4.4: Mögliche Objektarten der Ebene Landcover von VECTOR200 (Quelle: swisstopo, 2011, S. 36).

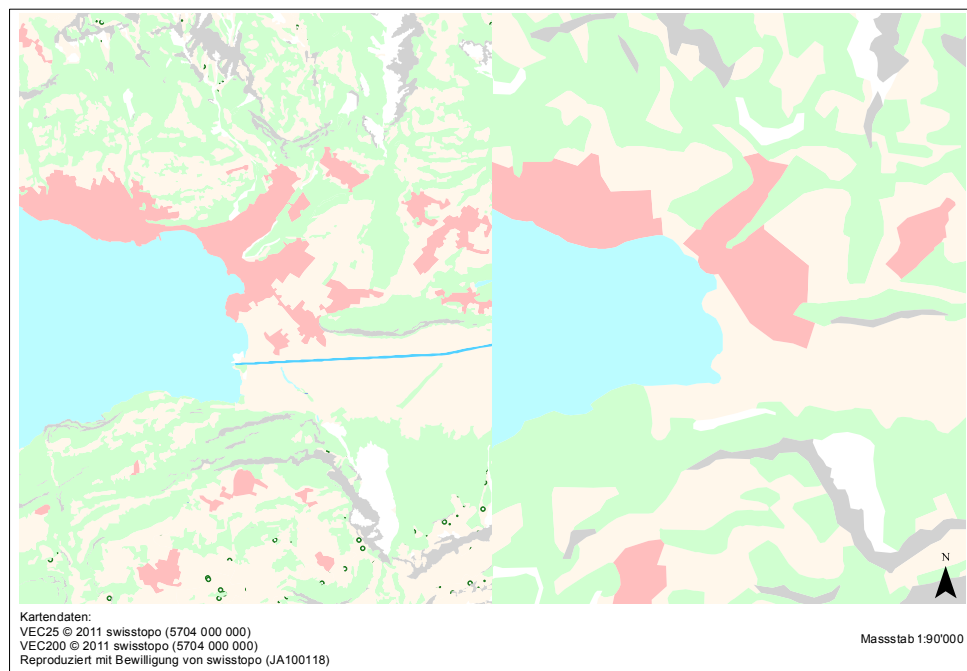


Abbildung 4.2: Ausschnitt der Primärflächen aus VECTOR25 (linker Kartenausschnitt) und VECTOR200 (rechter Kartenausschnitt).

4.2.5 Digitale Höhenmodelle DHM25 / DHM RIMINI

Wie VECTOR25 basiert auch das digitale Höhenmodell DHM25 (vgl. Abbildung 4.3) auf der Landeskarte 1:25'000 und beschreibt die dreidimensionale Form der Erdoberfläche (ohne Bebauung und Bewuchs). Dabei muss zwischen dem sogenannten Basismodell (Linien- und Punktgeometrien) und dem durch Interpolation abgeleiteten "Matrixmodell" (Raster) unterschieden werden. In dieser Arbeit wurde das Raster verwendet. Es weist eine Maschenweite von 25 Metern auf und jeder Zelle ist ein Höhenwert zugeordnet, welcher sich auf die Eckpunkte der Matrixzelle bezieht und nicht auf deren mittlere Höhe (swisstopo, 2005, S. 3 - 6).

Das digitale Höhenmodell DHM RIMINI (vgl. Abbildung 4.4) weist eine Maschenweite von 250 Metern auf und eignet sich daher für Arbeiten mit geringeren Genauigkeitsansprüchen.⁶

⁶ http://www.bfs.admin.ch/bfs/portal/de/index/infothek/erhebungen__quellen/blank/blank/arealstatistik/03/03_05.html, 09.01.2012
<http://www.geocat.ch/geonetwork/srv/deu/metadata.show?uuid=e55d9f69-508c-442c-89ab-6d1a8263ea3d&currTab=default>, 09.01.2012

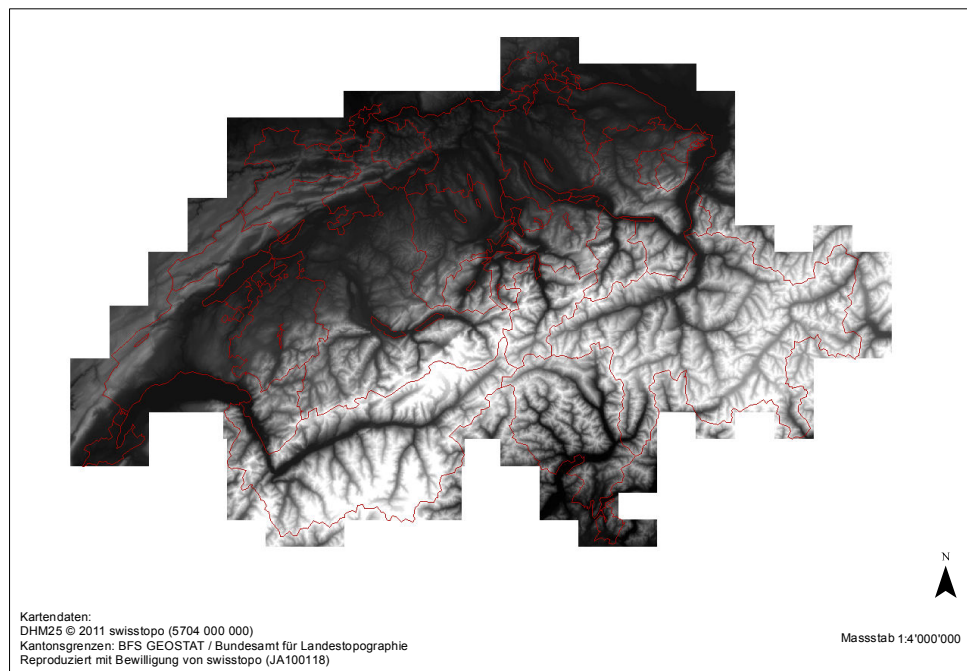


Abbildung 4.3: Digitales Höhenmodell DHM25.

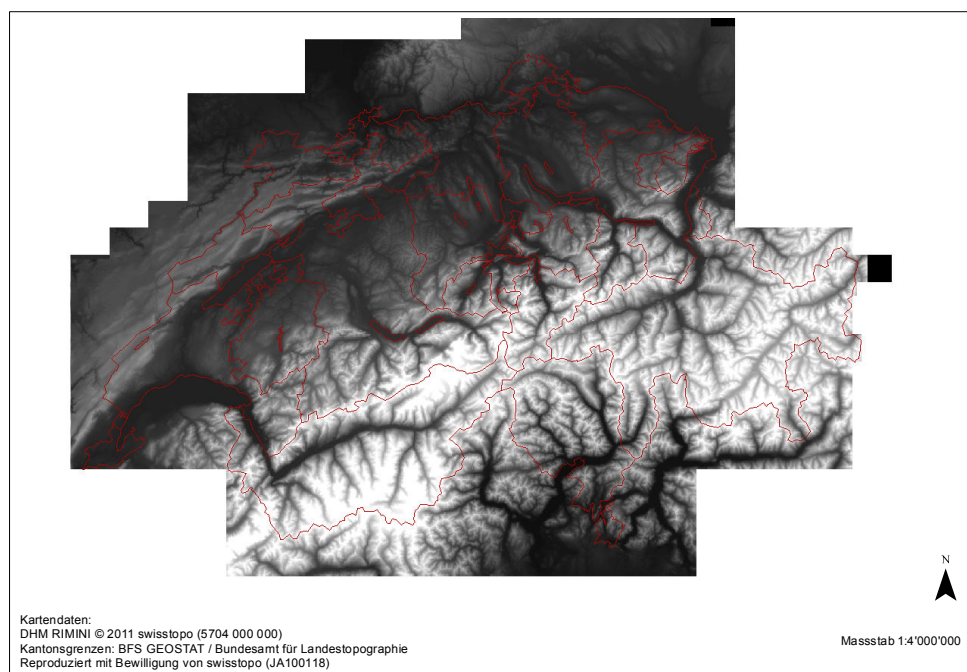


Abbildung 4.4: Digitales Höhenmodell DHM RIMINI.

4.2.6 Biogeographische Regionen

Der Datensatz Biogeographische Regionen wurde vom damaligen Bundesamt für Umwelt, Wald und Landschaft (BUWAL), heute Bundesamt für Umwelt (BAFU), bereitgestellt. Er basiert auf einem statistischen Gliederungsansatz, der die Flora und Fauna der Schweiz berücksichtigt. Insgesamt wurden sechs Grundregionen ausgeschieden (vgl. Abbildung 4.5), daneben gibt es eine feinere Unterteilung in zehn Regionen. Die Grundregionen sind (Gonseth u. a., 2001):

- Jura
- Mittelland
- Nordalpen (Alpennordflanke)
- Die Region Zentralalpen vereint die ursprünglich ausgeschiedenen Regionen Westliche und Östliche Zentralalpen. Durch deren Zusammenlegung entsprechen die biogeographischen Regionen den verwendeten Regionen aus den Höhenstufen (vgl. Kapitel 8).
- Südalpen (Alpensüdflanke)

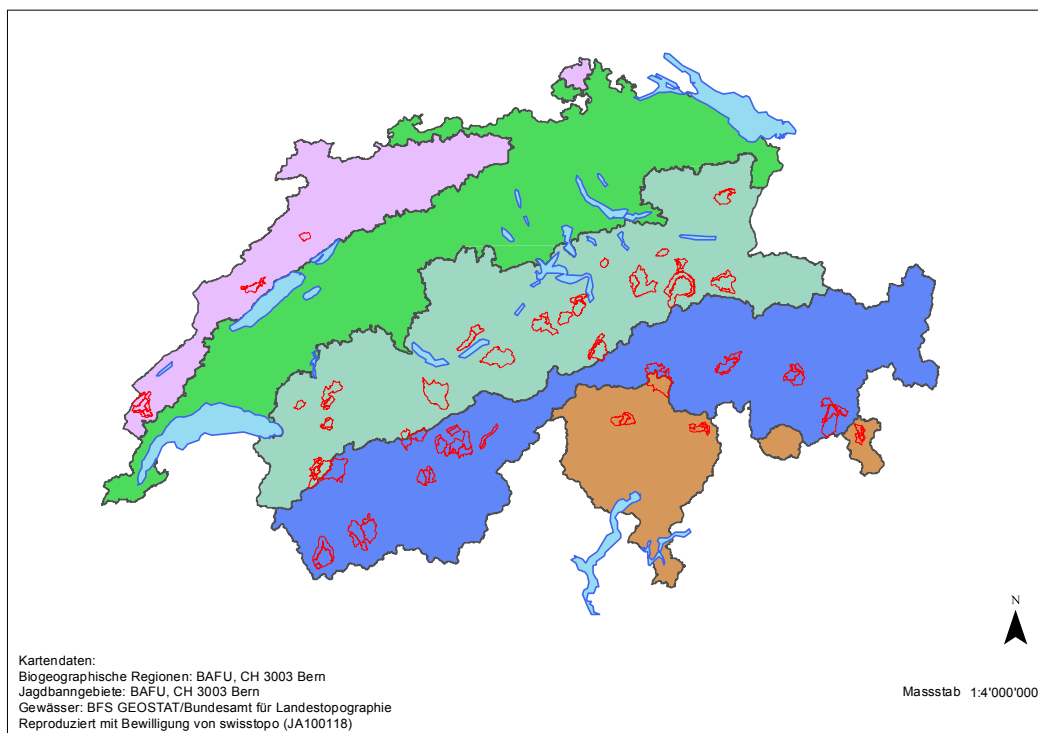


Abbildung 4.5: Die biogeographischen Regionen der Schweiz mit Jura, Mittelland, Nordalpen, Zentralalpen und Südalpen.

Ansatz “Vektorisierte Kleinste Einheiten” (VKE)

In den vorhandenen Ansätzen zur Konvertierung von Geodaten in das RDF-Format haben Rasterdaten bisher nur wenig Beachtung gefunden (vgl. Kapitel 3). Deshalb wurde ein Ansatz ausgearbeitet, der sowohl den Umgang mit Vektor-, als auch mit Rasterdaten ermöglichen soll. Der Ansatz Vektorisierte Kleinste Einheiten (VKE) wurde durch ein exploratives Vorgehen entwickelt. Dabei standen die Anwendung bestehender Konzepte bezüglich der Umformung von Vektordaten in RDF, sowie die Entwicklung und Erprobung möglicher Varianten zur Repräsentation von Rastern im Vordergrund. Anhand eines im Voraus ausgewählten Use Case wurde der Ansatz VKE im Detail ausgearbeitet und implementiert. Dadurch kann er als ein erster Prototyp für weitere Entwicklungen angesehen werden.

Der ausgewählte Anwendungsfall sollte die Überschaubarkeit der Prozesse und deren Nachvollziehbarkeit gewährleisten. Ausgesucht wurde eine vereinfachte Version des komplexen Use Case “Waldflächen und Höhenstufen” mit der Bezeichnung “Waldfläche und Höhe simpel”:

“Wie gross ist die Waldfläche innerhalb eines bestimmten Jagdbanngebietes und wie viel beitragen die durchschnittliche, maximale und minimale Höhe?”

Der Teil aus dem komplexen Anwendungsfall “Waldflächen und Höhenstufen”, welcher die Höhenstufen berücksichtigt, wurde durch eine einfache statistische Auswertung ersetzt. Zusätzlich beschränkt sich der Ansatz vorerst auf ein einzelnes Jagdbanngebiet, weshalb die grossmassstäbigen Datensätze (VECTOR25 und DHM25) gewählt wurden und somit eine Unterscheidung des Massstabes entfiel. Der vereinfachte Use Case schliesst Vektor- und Rasterdaten ein, deckt somit die zentralen Anforderungen des Ansatzes VKE ab (vgl. Abschnitt 5.1).

Im Verlauf der Umsetzung kristallisierte sich heraus, dass der Ansatz VKE die Erwartungen aufgrund der Anwendungsfälle nicht optimal erfüllen würde, weshalb keine Weiterentwicklung stattfand. Die nachfolgenden Erläuterungen vermitteln die Ideen und Kritikpunkte des Ansatzes VKE. Dieser Prototyp bildet die Vergleichsbasis für die Entwicklung weiterer Ansätze.

Dieses Kapitel beschreibt als erstes die konzeptionellen Grundlagen des entwickelten Ansatzes VKE und stellt ihn dem Ansatz des USGS gegenüber. Danach folgt die Vorstellung des Datenmodells von VKE. Anschliessend werden die Datenprozessierung, die aus dem Ansatz hervorgegangenen Resultate und deren Konsequenzen erläutert.

5.1 Konzeptionelle Grundlagen

VKE basiert auf den Ideen der Projekte von LinkedGeoData.org¹, GeoNames², bzw. den Arbeiten des Ordnance Survey (Goodwin u. a., 2008) und des USGS (Usery u. Varanka, 2011). Alle vier Organisationen formen neben den Sachdaten auch die Geometrien selbst in das RDF-Format um.

Dieser Grundsatz wurde übernommen, d. h. VKE sollte die Geodaten umfassend in RDF umwandeln und somit möglichst ohne Einschränkungen den potentiellen Benutzern zur Verfügung stellen. Damit verknüpft war die Bedingung, dass der exemplarische Anwendungsfall “Waldfläche und Höhe simpel” alleine durch die Möglichkeiten von Linked Data beantwortet werden musste. Es bestand nicht die Absicht ein Geographisches Informationssystem (GIS) zu imitieren. Vielmehr wird der auf Linked Data basierende Ansatz als Ergänzung (mit anderen Möglichkeiten) zu GIS betrachtet.

Da die zu RDF konvertierten Rasterdaten möglichst den Originaldaten entsprechen sollten, kam die Stossrichtung des Ansatzes von Usery und Varanka (2011) nicht in Frage. VKE soll nicht die im Voraus extrahierten Objekte des Rasters umwandeln, sondern das gesamte Raster.

Die detailliertesten Informationen eines Rasters stellen die einzelnen Zellen dar. Eine feinere Unterteilung als diese bringt keine neue Information hervor. Hier setzt die Idee von VKE an. Damit nicht nur die Vektordaten, sondern auch die Rasterdaten möglichst ohne Informationsverlust konvertiert werden können, werden die einzelnen Rasterzellen als eigene Einheiten angenommen und somit als Objekte des Rasters betrachtet. Konkret bedeutet dies, dass jede einzelne Rasterzelle als Linked Data abgelegt werden soll. Jedoch sei angemerkt, dass die Semantik eines solchen Objektes, das eine einzelne Rasterzelle repräsentiert, gering ist.

Folgende Möglichkeiten zur Konvertierung der Rasterzellen wurden in Betracht gezogen:

- Jede einzelne **Zelle wird referenziert** und kann im Raster verortet werden. In diesem Fall würde aber keine Geometrie als Tripel mit abgelegt, sondern die Tripel beschreiben lediglich Lage, Grösse und zugehörige Attribute einer Zelle. Die Lage könnte durch Koordinaten und/oder Angabe der benachbarten Objekte angegeben werden. Auch wäre die Nennung der Position in Bezug auf einen Eckpunkt des Rasters denkbar.
- Jede einzelne **Zelle wird vektorisiert** und zu einem Polygon (Fischnetz). Dadurch kann auch die Geometrie als RDF in den bereits bekannten Formaten für Vektoren abgelegt werden. Nachbarzellen, die das Gleiche repräsentieren, könnten aggregiert werden. Zusätzlich wäre auch eine den Objekten übergeordnete Beschreibung der Metadaten des gesamten Rasters denkbar.
- Der **Schwerpunkt der Zellen** (Centroid) wird als Punktgeometrie abgelegt. Somit repräsentiert ein Punkt eine Zelle. Diesem können die zugehörigen Attribute angehängt werden. Da ein Raster meist regelmässig ist und die Zellen eine gleichbleibend grosse Fläche repräsentieren, kann auch die Gesamtfläche gespeichert werden.

VKE bedient sich der Vektorisierung sämtlicher Rasterzellen zu Polygonen. Dadurch kann auf bereits erprobte Formate zurückgegriffen und können im Hinblick auf GeoSPARQL möglichst viele Optionen offen gehalten werden - gerade bezüglich geometrische Abfragen.

Mit der Repräsentation jeder einzelnen Rasterzelle als mindestens ein Tripel wurde ein Aufblähen der Knowledgebase bewusst in Kauf genommen. Dies geschah jedoch in der Annahme, dass Triple Stores für solch grosse Datenmengen ausgelegt sind und damit umgehen können. Im Allgemeinen wird davon

¹ <http://linkedgeodata.org/About>

² <http://www.geonames.org/>

ausgegangen, dass Tripel-Mengen im zweistelligen Millionenbereich ohne Probleme verarbeitet werden können, sofern die Hardwarevoraussetzungen erfüllt sind.³ Die Firma Franz Inc.⁴ hatte im August 2011 bekanntgegeben, erfolgreich eine Billion (10^{12}) Tripel in ihren Triple Store geladen zu haben.⁵

5.2 Vergleich VKE und USGS

Der Ansatz VKE bezieht ebenso wie der des USGS (Usery und Varanka, 2011) Rasterdaten mit ein. Jedoch verfolgen beide unterschiedliche Wege:

- VKE setzt **keine Extraktion** von relevanten Objekten aus dem Raster voraus, da dieses als Ganzes umgeformt wird. Dadurch sollte eine möglichst grosse Flexibilität, im Sinne einer vielseitigen Verwendbarkeit, gewährleistet und der Informationsverlust reduziert werden.
Je nach neuer Fragestellung müssten beim Ansatz des USGS jeweils die Objekte neu ausgeschieden werden. Die zu RDF konvertierten Daten können also nicht universell für unterschiedlichste Fragestellungen eingesetzt werden. Zudem gehen die um die Objekte herumliegenden Informationen bei der Konvertierung verloren. Bei VKE hingegen ist es möglich im Nachhinein (d. h. nach der Umformung) Objekte zu bestimmen. Ergänzend könnte man ebenfalls im Voraus extrahierte Objekte als Linked Data ablegen, indem die Geometrieinformationen indirekt durch Angabe der betroffenen Rasterzellen angefügt werden. Dadurch wären die Objekte und ein Abbild der Originaldaten vorhanden.
- Bei Usery und Varanka (2011) sind nur die extrahierten und in RDF umgeformten Objekte als Ganzes ansprechbar. Zwar bleibt die Information der Rasterzellen innerhalb dieser Objekte erhalten (durch Speicherung des Rasterausschnittes im GML-Format), doch sind sie nur indirekt durch Abfrage des gesamten Strings abrufbar. Bei VKE hingegen verkörpert jede **Rasterzelle ein eigenes Objekt**, womit jede dieser Informationseinheiten direkt ansprechbar ist und somit die dazugehörigen Informationen abrufbar sind.
- VKE benötigt zur **Umformung in RDF** einen Schritt weniger als der Ansatz von Usery und Varanka, die ihre Daten erst in das GML und dann in das RDF-Format konvertieren, wohingegen VKE direkt aus den ursprünglichen Geodaten die RDF-Tripel generiert. Es werden also weniger Operationen benötigt, um die jeweiligen Formate umzuschreiben.
- Wie dem OGC-Vorschlag für den GeoSPARQL-Standard (Perry u. Herring, 2011) zu entnehmen ist, sind bisher keine Rasterformate oder Möglichkeiten zur Rasterprozessierung vorgesehen. Es existiert also noch kein offizieller Vorschlag für die Repräsentation von Rastern, weshalb der Datenanbieter eigene Lösungen finden muss (wie beispielsweise Usery und Varanka, 2011). Dadurch könnte das Risiko steigen, dass jeder Anbieter seine eigene Lösung präsentiert, was den einfachen Datenaustausch erschweren würde. Die Vektorisierung des Rasters bei VKE bringt den Vorteil, dass die Daten im aufkommenden GeoSPARQL-Standard repräsentiert werden, welcher Vektoren unterstützt.

³ <http://www.w3.org/wiki/LargeTripleStores>, 29.11.2011

⁴ <http://www.franz.com/>

⁵ http://semanticweb.com/what-one-trillion-means-for-the-semantic-web_b22623, 29.11.2011

5.3 Modellgrundlagen VKE

5.3.1 Implementierung der räumlichen Beziehungen

Eine zentrales Element von Geodaten sind der Raumbezug (Georeferenzierung) und die räumlichen (topologischen) Beziehungen zu anderen Objekten. Diese räumlichen Beziehungen (z. B. Nachbarschaften) müssen vor der Umformung zu RDF abgeleitet und dann als Tripel abgelegt werden. Dies ist grundsätzlich solange nötig, bis geometrische Berechnungen bei den Abfragen miteinbezogen werden oder die Daten ähnlich einem topologischen Modell abgelegt werden können.

Auch der Ordnance Survey erstellte solche Beziehungen bei der Repräsentation der administrativen Grenzen von Grossbritannien als Linked Data (Goodwin u. a., 2008). Dabei beziehen sie sich auf vier Eigenschaften des RCC8 Modells⁶. Diese bezeichnen sie als: `completelySpatiallyContains` (entspricht NTPPi), `tangentiallySpatiallyContains` (entspricht TPPi), `isSpatiallyEqualTo` (entspricht EQ) und `borders` (entspricht EC) (Goodwin u. a., 2008). Eine administrative Region kann somit in einer anderen enthalten sein (NTPPi, TPPi), direkt an eine andere angrenzen (EC) und/oder die gleiche Grenze und Fläche wie eine andere (EQ) aufweisen (Cohn u. a., 1997).

Aufgrund der Fragestellung des Anwendungsfalles “Waldfläche und Höhe simpel”, begrenzte sich das Interesse an den Daten auf die Perimeter der Jagdbanngebiete, womit vor allem die GeoSPARQL Beziehungen `ogc:equals`, `ogc:within` und `ogc:contains` im Fokus standen. Da `ogc:contains` die Inversion von `ogc:within` ist und im Datenmodell jeweils nur eine gerichtete räumliche Beziehung definiert wurde, konnte `ogc:contains` ausgeschlossen werden. Um `ogc:equals` einzubinden hätten die kongruenten Geometrien ermittelt werden müssen. Die Wahrscheinlichkeit, dass kongruente Geometrien vorhanden sind wurde als gering eingeschätzt, weshalb `ogc:equals` nicht implementiert wurde. Die Eigenschaft `ogc:within` entspricht den RCC8 Eigenschaften TPP (tangential proper part) und NTPP (non-tangential proper part) (Perry u. Herring, 2011). Das heisst, ein Objekt liegt vollständig innerhalb eines anderen ungleichen und grösseren Objektes, wobei sich die Ränder der Objekte tangential berühren (TPP) oder nicht (NTPP) (Cohn u. a., 1997).

Quelltext 5.1 zeigt ein Beispiel-Tripel aus dem Ansatz VKE für ein Waldstück, das in einem bestimmten Jagdbanngebiet liegt (N3-Format). Die Primärfläche mit der Identifikationsnummer 3 liegt also innerhalb des Jagdbanngebietes mit der eindeutigen Nummer 476968.

```
1 <http://www.wsl.example/JB/Primaerflaeche/3>
2   <http://www.opengis.net/rdf#within>
3   <http://www.wsl.example/JB/Jagdbanngebiet/476968> .
```

Quelltext 5.1: Waldobjekt im N3-Format.

5.3.2 Datenmodell

Gemäss dem vereinfachten Use Case “Waldfläche und Höhe simpel” und den konzeptionellen Grundlagen, wurde ein passendes Datenmodell (Vokabular/Ontologie) entwickelt. Es bildet eine wichtige Basis für die Umformung der Daten in das RDF-Format. Die definierten Klassen mit den zugehörigen Attributen und Beziehungen zu den anderen Klassen werden aufgezeigt, womit klar wird, welche Daten in das RDF-Format konvertiert werden müssen.

Der letzte Stand dieses Modells ist in Abbildung 5.1 ersichtlich. Die in Rechtecken umrahmten Begriffe

⁶ vgl. Cohn u.a., 1997

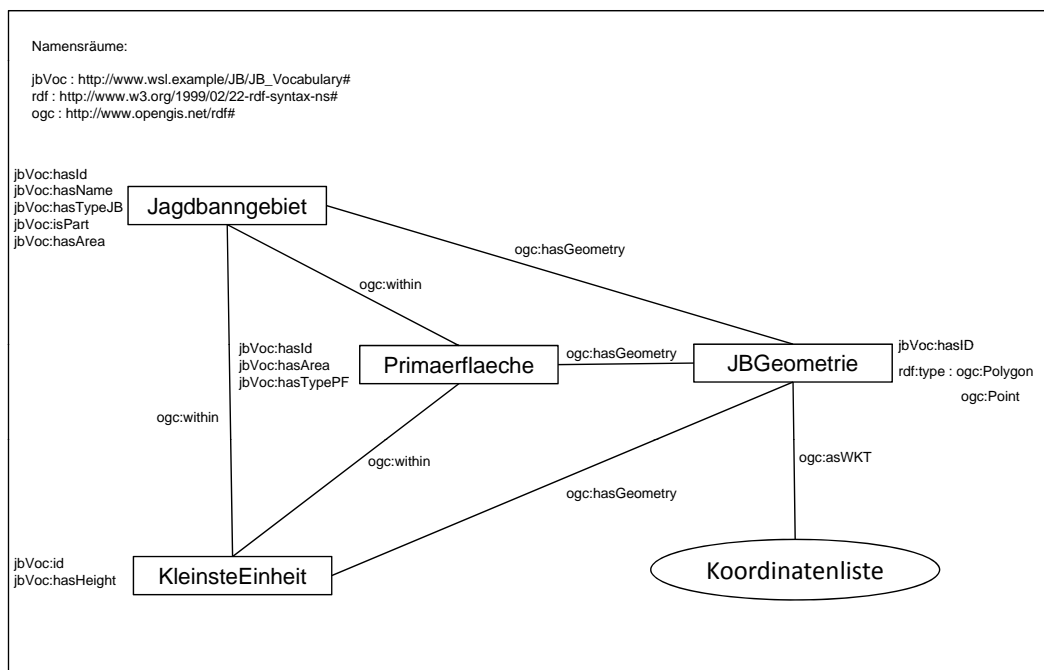


Abbildung 5.1: Datenmodell des Ansatzes VKE.

stellen die Klassen dar und daneben stehend befinden sich die zugehörigen Attribute. Grundsätzlich stehen die ovalen Formen für Literale. In diesem Modell ist aber aufgrund der Übersichtlichkeit nur die Koordinatenliste explizit als Literal ausgewiesen, da sie eine spezielle Syntax besitzt. Bei den Attributen handelt es sich, bis auf die Beziehungen und die Eigenschaften der Klasse **Geometrie**, allesamt um Literale. Im Modell werden die URIs mit Namensräumen abgekürzt.

Die Beziehungen zwischen den Klassen sind genau genommen ebenfalls als Attribute einer Klasse abgelegt, werden aber wegen ihrer verbindenden Bedeutung separat aufgeführt (vgl. Kapitel 2). Beispielsweise die Beziehung `ogc:hasGeometry` zwischen den Klassen **Jagdbanngebiet** und **JBGeometrie** bedeutet, dass **Jagdbanngebiet** das Attribut `ogc:hasGeometry` besitzt. Zudem wird kein Literal, sondern eine andere Klasse bzw. eines ihrer Objekte referenziert. Die Beziehung ist gerichtet, d. h. es besteht keine Verbindung von **JBGeometrie** zurück zu **Jagdbanngebiet**.

Die zentrale Klasse im Modell ist **Jagdbanngebiet**, welche in der Regel auch den Ausgangspunkt bei Abfragen stellt. Sie verfügt über eine Identifikationsnummer `jbVoc:hasID`, die aus dem ursprünglichen Datensatz übernommen wurde. Zusätzlich zur Fläche (`jbVoc:hasArea`), dem Namen (`jbVoc:hasName`) und dem Schutztypus eines Gebietes (`jbVoc:hasTypeJB`), wurde die Eigenschaft `jbVoc:isPart` hinzugenommen. Sie gibt an, ob es sich um ein Teilgebiet eines grösseren Jagdbanngebietes handelt.

Den Jagdbanngebieten sind die Klassen **Primaerflaeche** und die aus der Verschneidung sämtlicher Daten hervorgegangene **KleinsteEinheit** untergeordnet (Zustandekommen der Kleinsten Einheiten wird im Abschnitt 5.4 erläutert). Beide besitzen eine eindeutige ID (`jbVoc:hasId`), topologische Beziehungen und spezifische Attribute. Bei den Primärflächen sind diese die Fläche (`jbVoc:hasArea`) und der Typ (`jbVoc:hasTypePF`) wie beispielsweise Wald, Fels, Sumpf usw. Die kleinsten Einheiten werden lediglich durch die Höhenangabe (`jbVoc:hasHeight`) ergänzt. Die topologische Information `ogc:within` stammt aus der Vorprozessierung und stellt die räumlichen Beziehungen zwischen den Klassen dar (vgl. Abschnitt 5.4).

Die Geometrien wurden mittels einer zusätzlichen Klasse (`JBGeometrie`) abgelegt (gemäss GeoSPARQL-Standard). Sie besitzt als Attribute `ogc:asWKT`, welches die Koordinaten der Geometrie aufnimmt, und eine eindeutige Identifikationsnummer. Zusätzlich wird `JBGeometrie` eine weitere Klasse aus dem `ogc` Namensraum durch das Prädikat `rdf:type` zugewiesen. Das heisst, die Geometrie ist vom Typ her entweder ein `ogc:Polygon` oder ein `ogc:Point`, welche beide durch das GeoSPARQL-Vokabular des OGC definiert werden (vgl. Perry u. Herring, 2011).

5.4 Datenprozessierung

Bevor die Daten in das RDF-Format konvertiert werden konnten, waren einige Vorprozessierungen nötig, was mehrheitlich mit Hilfe des Geoinformationssystems ArcGIS⁷ (Version 10.0) von ESRI bewerkstelligt wurde. Entsprechend dem vereinfachten Use Case “Waldfläche und Höhe simpel” beschränken sich die benötigten Daten auf die Jagdbanngebiete, die thematische Ebene Primärflächen aus VECTOR25 und auf das Höhenmodell DHM25. Diese Datensätze mussten so weit aufbereitet werden, dass sämtliche Informationen, welche als Tripel abgelegt werden sollen, vorhanden und abrufbar waren. Der Datensatz Jagdbanngebiete konnte weitgehend unverändert übernommen werden. Lediglich nicht relevante Attribute wurden aufgrund der Übersichtlichkeit entfernt.

Die thematische Ebene Primärflächen wurde auf das Gebiet der Jagdbanngebiete mittels einer “Clip” Operation begrenzt. Die Abbildungen 5.2 und 5.3 zeigen rot umrandet einen Ausschnitt des Jagdbanngebietes “Urirotstock”, welches aus mehreren Teilen mit jeweils unterschiedlichen Schutztypen besteht. Die grünen Flächen entsprechen dem Wald, die roten den Siedlungsgebieten und die grauen stellen felsige Gebiete dar.

Aus Abbildung 5.3 wird ersichtlich, dass nach dem Ausschneiden die Primärflächen auf die Grenzen der Jagdbanngebiete reduziert wurden. Es gibt also keine Primärflächen mehr, die ausserhalb der Jagdbanngebiete liegen.

Das digitale Höhenmodell DHM25 wurde, wie die Primärflächen, gemäss den Jagdbanngebieten mit dem Werkzeug “Extract by Mask” extrahiert. Dabei gilt es aber zu beachten, dass die extrahierten Zellen nicht genau mit den Jagdbanngebietsgrenzen übereinstimmen. Dies ist auf die GIS Operation zurückzuführen, da sie die einzelnen Rasterzellen nicht weiter zerteilt und Vektoren (welche als Schnittmasken dienen) bei der Prozessierung ebenfalls zu Rastern konvertiert werden.⁸ Diese Ungenauigkeit konnte vorerst ausser Acht gelassen werden: Zum einen stand beim Ansatz VKE das Testen möglicher Arten der Datenkonvertierung im Vordergrund und zum anderen sind die problematischen Flächen relativ klein in Relation zur gesamten Fläche der Jagdbanngebiete.

Das Problem ist in Abbildung 5.4 illustriert. Dieses Mal ist das Gebiet Urirotstock ganz abgebildet und innerhalb dieses Gebietes ist das extrahierte digitale Höhenmodell zu erkennen. Rechts daneben befindet sich ein vergrösserter Ausschnitt des weiss umrahmten Gebietes, der die beschriebene Ungenauigkeit aufzeigt: Die Rasterzellen (in schwarz) über- oder unterschreiten die Jagdbanngebietsgrenzen. An den Stellen, wo das Raster die Grenze unterschreitet, entsteht eine Lücke ohne Werte des Rasters, da durch die Extraktion diese Rasterzellen “verloren” gegangen sind.

Das extrahierte Höhenmodell wurde danach mit dem Werkzeug “Raster to Polygon” vektorisiert. Das heisst, jede einzelne Rasterzelle wurde gemäss ihrer Form (hier Quadrat) zu einem Polygon. Das resultierende “Fischnetz” wurde nicht vereinfacht, lediglich benachbarte Zellen mit den gleichen Werten (hier

⁷ <http://www.esri.com/software/arcgis/index.html>

⁸ <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/001w0000001t000000, 24.10.2011>

http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Extract_by_Mask/009z00000002n000000/, 23.10.2011

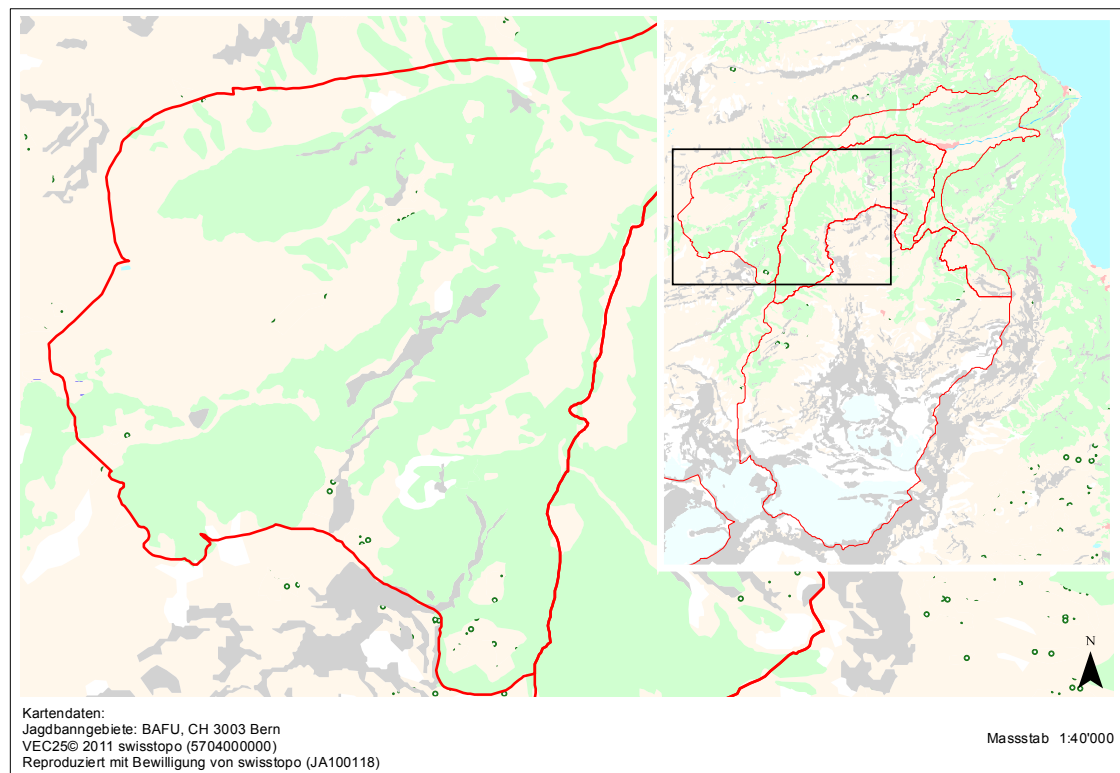


Abbildung 5.2: Primärflächen im Originalzustand mit dem rot umrandeten Jagdbanngebiet “Urirotstock”.

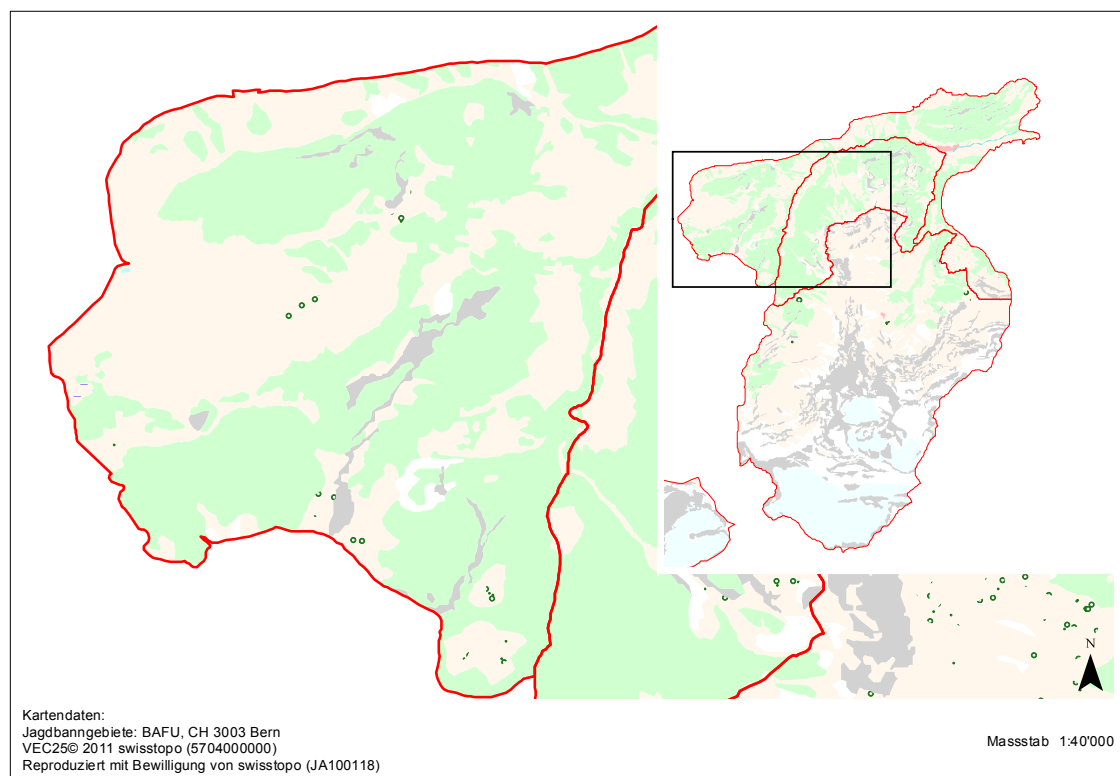


Abbildung 5.3: Primärflächen auf das Jagdbanngebiet “Urirotstock” zugeschnitten (nach Durchführung der “Clip” Operation).

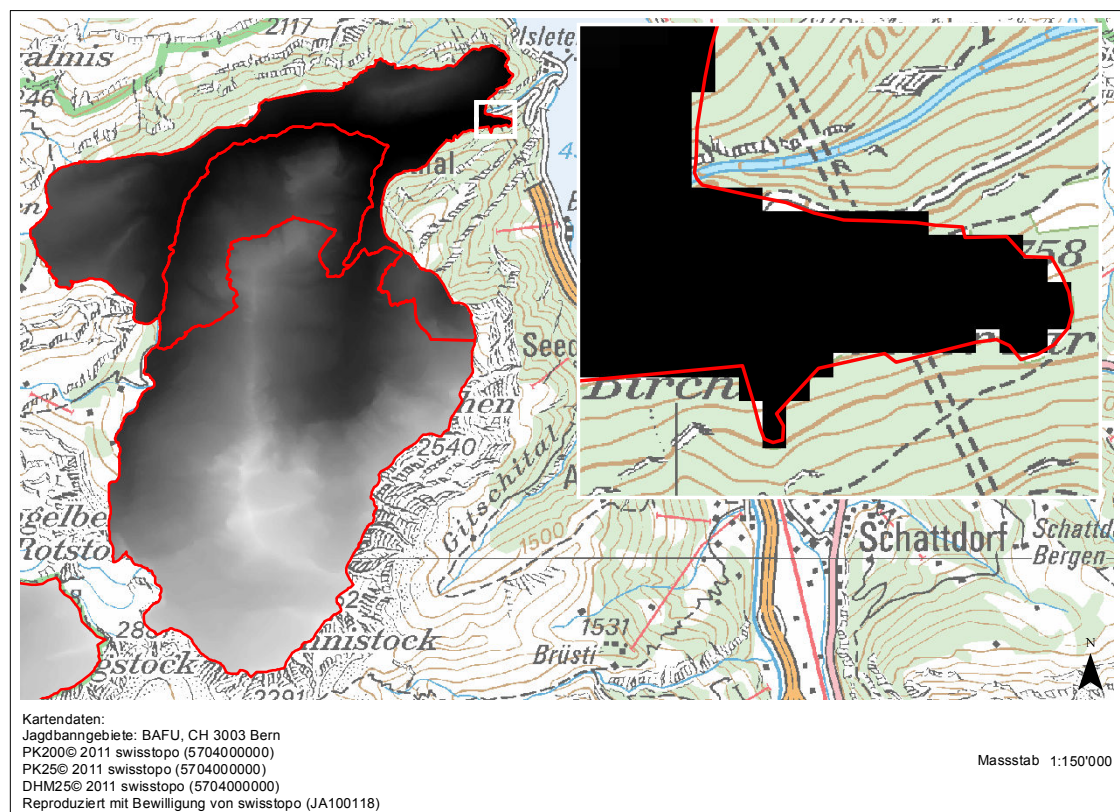


Abbildung 5.4: Digitales Höhenmodell DHM25 auf die Jagdbanngebiete zugeschnitten und Problematik der Rasterextraktion (kleine Karte).

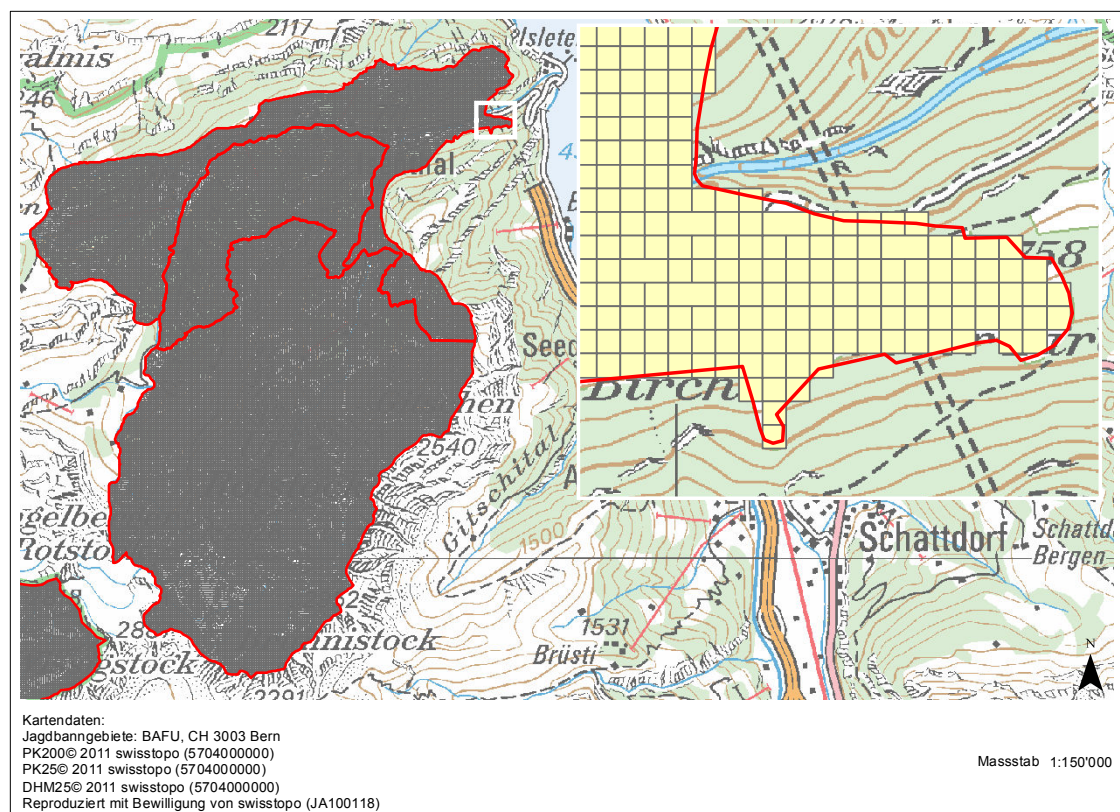


Abbildung 5.5: Digitales Höhenmodell DHM25 vektorisiert (Fischnetzstruktur).

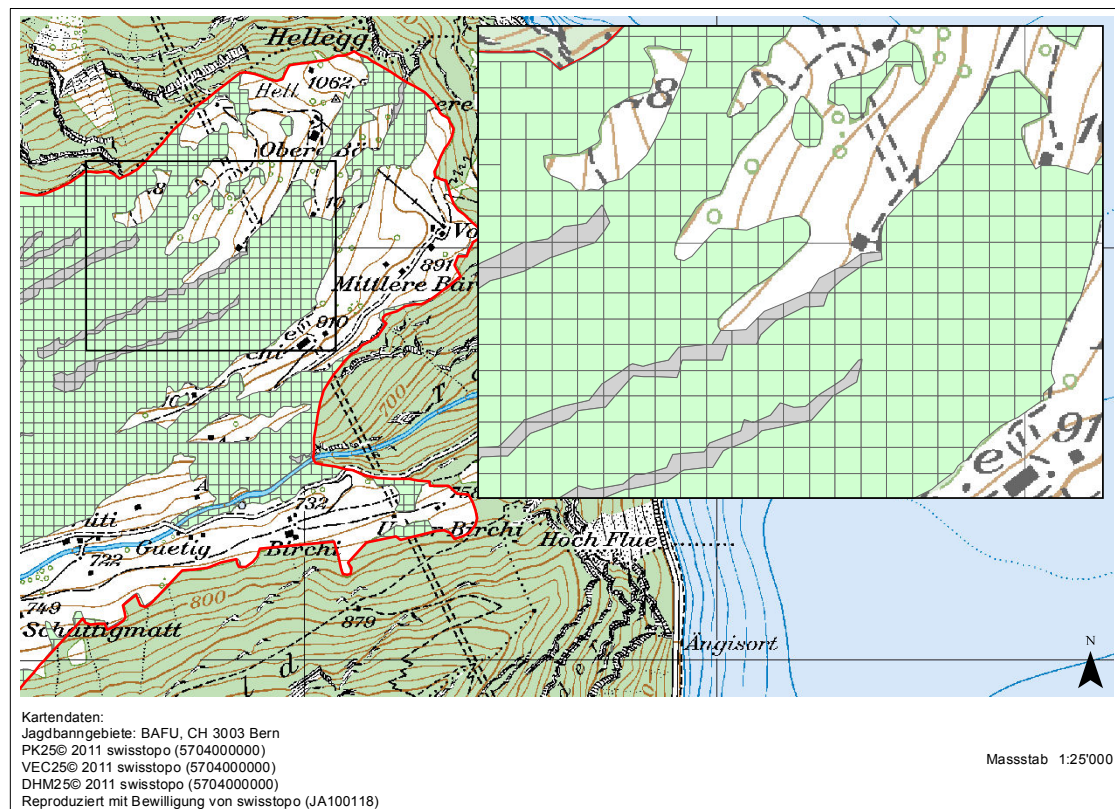


Abbildung 5.6: Union aus Jagdbanngebieten, Primärflächen und vektorisiertem DHM25. Die daraus entstandenen Kleinsten Einheiten sind grau umrandet.

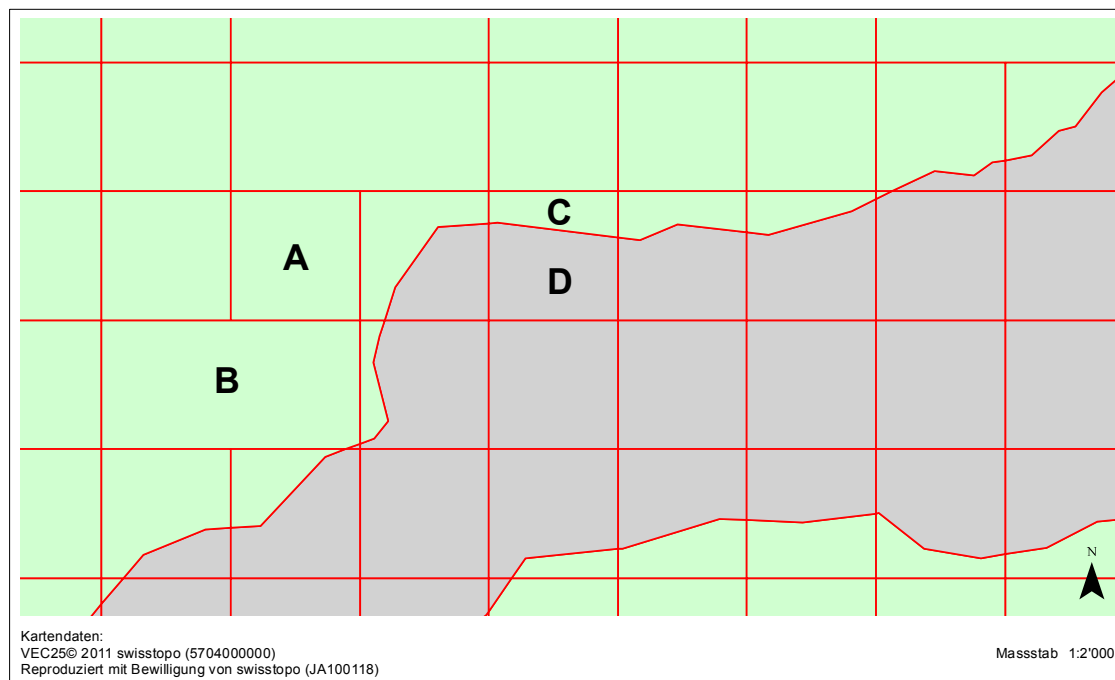


Abbildung 5.7: Vergrösserte Darstellung der Kleinsten Einheiten (rot umrandet).

Höhe) wurden zu einem Polygon zusammengefasst. Das ist möglich, da dieses Werkzeug zur Konversion ganzzahlige Rasterwerte benötigt, weshalb die Werte vorher gerundet werden müssen.⁹

Die vorgängig beschriebene Ungenauigkeit hätte durch Vorziehen des Einsatzes des Werkzeugs “Raster to Polygon” umgangen werden können. Also erst die Vektorisierung des gesamten Rasters, gefolgt vom Ausschneiden nach den Jagdbanngebieten. Jedoch ist die Vektorisierung des gesamten Rasters sehr rechenintensiv.

Aus der Abbildung 5.5 wird die Fischnetzstruktur ersichtlich und es lässt sich die grosse Datenmenge erahnen, welche bereits in dieser Prozessierungsphase anfiel.

Somit waren alle Ausgangsdatensätze für sich gesehen aufbereitet. Diese wurden mit dem Werkzeug “Union” miteinander verschnitten. Das Ergebnis sind Objekte, die von keinen anderen Objekten (aus allen Layern) geschnitten werden. Die namensgebenden “Kleinsten Einheiten” (VKE) entstanden. Jede Kleinste Einheit trägt die Informationen aller überlagernden Datensätze. Das heisst, sie verfügen über Attribute, in welchem Jagdbanngebiet, in welcher Primärfläche und in welcher Rasterzelle (ausser in den problematischen Randbereichen) sie liegen. Daraus konnten die räumlichen (topologischen) Beziehungen abgeleitet werden.

Abbildung 5.6 zeigt, dass die Fischnetzstruktur als Folge der Verschneidungen teilweise aufgelöst wurde. Abbildung 5.7 zeigt eine vergrösserte Darstellung der Kleinsten Einheiten, die zusätzlich durch die rote Umrandung hervorgehoben werden. Viele Rasterzellen wurden unverändert in Polygone umgewandelt (A). Benachbarte Zellen mit den gleichen Höhenwerten wurden zu einem Polygon zusammengenommen (B). Rasterzellen, die beispielsweise zu einem Drittel im Wald und zu den restlichen zwei Dritteln im Felsgebiet liegen, werden gemäss den Grenzen der beiden Primärflächen getrennt (C und D).

Der Quelltext 5.2 zeigt die Tripel einer Kleinsten Einheit (ohne die Geometrie) mit der ID 14116. Die ersten beiden Tripel verorten die Kleinste Einheit im Jagdbanngebiet 476968 und in der Primärfläche 8. Das letzte Tripel definiert die Höhe mit 2078 Meter als Datentyp `double`.

```

1 <http://www.wsl.example/JB/SmallestUnit/14116>
2   <http://www.opengis.net/rdf#within>
3   <http://www.wsl.example/JB/Jagdbanngebiet/476968> .
4
5 <http://www.wsl.example/JB/SmallestUnit/14116>
6   <http://www.opengis.net/rdf#within>
7   <http://www.wsl.example/JB/Primaerflaeche/8> .
8
9 <http://www.wsl.example/JB/SmallestUnit/14116>
10  <http://www.wsl.example/JB/JB_Vocabulary#hasHeight>
11  "2078"^^<http://www.w3.org/2001/XMLSchema#double> .

```

Quelltext 5.2: Kleinste Einheit im Format-N3.

⁹ <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//001200000008000000>, 10.01.2012

5.5 Umformung in RDF

Mit den vorprozessierten Datensätzen und der Ontologie waren alle Grundlagen zur Umformung in das RDF-Format vorhanden. Konkret wurden die unveränderten Jagdbanngebiete, die ausgeschnittenen Primärflächen und die erzeugten Kleinsten Einheiten mit den zugehörigen Geometrien und Sachattributen umgeformt.

Umformung oder Konvertierung bezeichnet die Formulierung von Tripeln auf Grundlage der bestehenden Daten. Das Datenmodell gibt an, was für Tripel und wie sie erstellt werden sollen. Das folgende Tripel im Quelltext 5.3 weist dem Jagdbanngebiet "Schwarzhorn" dessen Fläche in m² zu:

```
1 <http://www.wsl.example/JB/Jagdbanngebiet/476927>
2   <http://www.wsl.example/JB/JB_Vocabulary#hasArea>
3   '71020201.13'^^<http://www.w3.org/2001/XMLSchema#double> .
```

Quelltext 5.3: Tripel für die Fläche eines Jagdbanngebietes.

Die Tripel wurden auf zwei verschiedene Arten erstellt:

- Mapping Programm: Das Programm D2RQ¹⁰ (Version 0.7) erlaubt ein Mapping auf relationalen Datenbanken. Das bedeutet, es ermöglicht den Zugriff auf Nicht-RDF Datenbanken aus Sicht eines RDF-Graphen.¹¹

Zudem kann D2RQ zur halb-automatisierten Erstellung von Tripeln verwendet werden. Dazu wurde mittels einer eigenen Mapping-Sprache definiert, wie die Daten aus der Datenbank in Tripeln ausgedrückt werden müssen.

Im Quelltext 5.4 findet sich ein Beispiel für das Mapping der Namen der Jagdbanngebiete:

```
1 map:jbName a d2rq:PropertyBridge;
2           d2rq:belongsToClassMap map:Jagdbanngebiet;
3           d2rq:property jbVoc:hasName;
4           d2rq:column "jb_2010.JB_NAME";
5           d2rq:datatype xsd:string;
```

Quelltext 5.4: Mapping für die Namen der Jagdbanngebiete.

Die erste Zeile bestimmt, dass es sich bei der folgenden Definition um ein Attribut einer Klasse handelt. In der zweiten und dritten Zeile werden die Klasse (`map:Jagdbanngebiet`) und das Prädikat (`jbVoc:hasName`) entsprechend dem eigenen Vokabular festgelegt. Das Attribut (Objekt) gehört also zur Klasse `Jagdbanngebiet` (Subjekt) und wird durch das Prädikat `jbVoc:hasName` verbunden. In der vierten Zeile ist die entsprechende Tabellenspalte in der Datenbank angegeben, während die fünfte den Datentyp (`xsd:string`) bestimmt.

Ein mögliches Tripel aus dem Mapping ist im Quelltext 5.5 aufgeführt:

```
1 <http://www.wsl.example/JB/Jagdbanngebiet/476898>
2   <http://www.wsl.example/JB/JB_Vocabulary#hasName>
3   'Mythen'^^<http://www.w3.org/2001/XMLSchema#string> .
```

Quelltext 5.5: Benennung eines Jagdbanngebietes.

¹⁰<http://www4.wiwiw.fu-berlin.de/bizer/d2rq/>

¹¹<http://www4.wiwiw.fu-berlin.de/bizer/d2rq/spec/index.htm>, 25.10.2011

- Mapping Skript: Mit der Programmiersprache Python¹² (Version 2.6) wurden Skripte geschrieben, welche ebenfalls die Daten aus den räumlichen Datenbanken auslesen und zusätzlich prozessieren können. Python wird von ArcGIS aktiv unterstützt und erlaubt durch die Einbindung der ArcPy-Erweiterung den Zugriff auf Werkzeuge und Funktionen, um mit Geodaten zu arbeiten.¹³ Somit konnten beispielsweise die Geometriedaten auf einfache Weise extrahiert werden.

Das Resultat der Umformungen waren mehrere RDF-Dateien im N3-Format, welche all die gewünschten Informationen enthielten. Diese Dateien wurden schlussendlich in den Triple Store Sesame geladen.

5.6 Resultate des vereinfachten Use Case

Der zu Beginn des Ansatzes VKE gestellte Use Case “Waldfläche und Höhe simpel” benötigt ein Jagdbanngebiet als Input, anhand dessen die gefragten Werte ermittelt werden sollen. Das Gebiet “Trescolmen” setzt sich aus mehreren Teilgebieten zusammen, vereint unterschiedliche Primärflächen (z. B. Fels, Siedlung, Wald usw.) in sich und ist somit ein repräsentatives Jagdbanngebiet. Der vereinfachte Use Case zur Erinnerung: “Wie gross ist die Waldfläche innerhalb eines bestimmten Jagdbanngebietes und wie viel betragen durchschnittliche, maximale und minimale Höhe?”

Der Anwendungsfall kann in zwei Teile separiert werden. Der erste fragt nach der Waldfläche, der zweite nach den statistischen Werten der Höhen.

Die SPARQL-Abfrage für die gesamte Waldfläche im Gebiet Trescolmen (Quelltext 5.6), wurde über die Sesame Workbench abgesetzt.

```

1 SELECT DISTINCT (SUM(?flaeche) AS ?Waldflaeche)
2 WHERE {
3   ?Jagdbanngebiet jbVoc:jbName "Trescolmen"^^<http://www.w3.org/2001/
4     XMLSchema#string>.
5   ?Primaerflaeche ogc:within ?Jagdbanngebiet.
6   ?Primaerflaeche jbVoc:pftype "Z_Wald"^^<http://www.w3.org/2001/
7     XMLSchema#string>.
8   ?Primaerflaeche jbVoc:hasArea ?flaeche .
9 }
```

Quelltext 5.6: SPARQL-Abfrage für die Waldfläche.

Unter **SELECT** wird festgelegt, dass die Abfrageergebnisse aufsummiert werden sollen (**SUM**), was zur Ausgabe eines einzelnen Wertes führt. In der **WHERE**-Klausel werden erst alle Teilgebiete abgerufen, die dem Jagdbanngebiet Trescolmen angehören. Dann folgt der Aufruf sämtlicher Primärflächen in diesem Gebiet (Zeile 5) und die Beschränkung auf die Waldflächen (Zeile 6 und 7). Zuletzt werden die Flächenwerte der Waldgebiete abgefragt (Zeile 8).

Das Resultat der Abfrage illustriert Abbildung 5.8 mit einem Ausschnitt der Benutzeroberfläche von Sesame (Workbench). Die gesamte Waldfläche im Jagdbanngebiet Trescolmen beträgt 5'612'458.341 m². Das Resultat wurde mit einer äquivalenten Abfrage im GIS verifiziert. Jedoch trat eine kleine Abweichung von 3 cm² auf. Es kann angenommen werden, dass die Abweichung zu gering ist, um auf einen Fehler in der RDF-Umformung oder SPARQL-Formulierung hinzudeuten. Vermutlich kann sie auf Rundungsfehler

¹²<http://www.python.org>

¹³http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/What_is_Python/002z00000001000000/, 25.10.2011
http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Essential_Python_vocabulary/002z000000022000000/, 25.10.2011

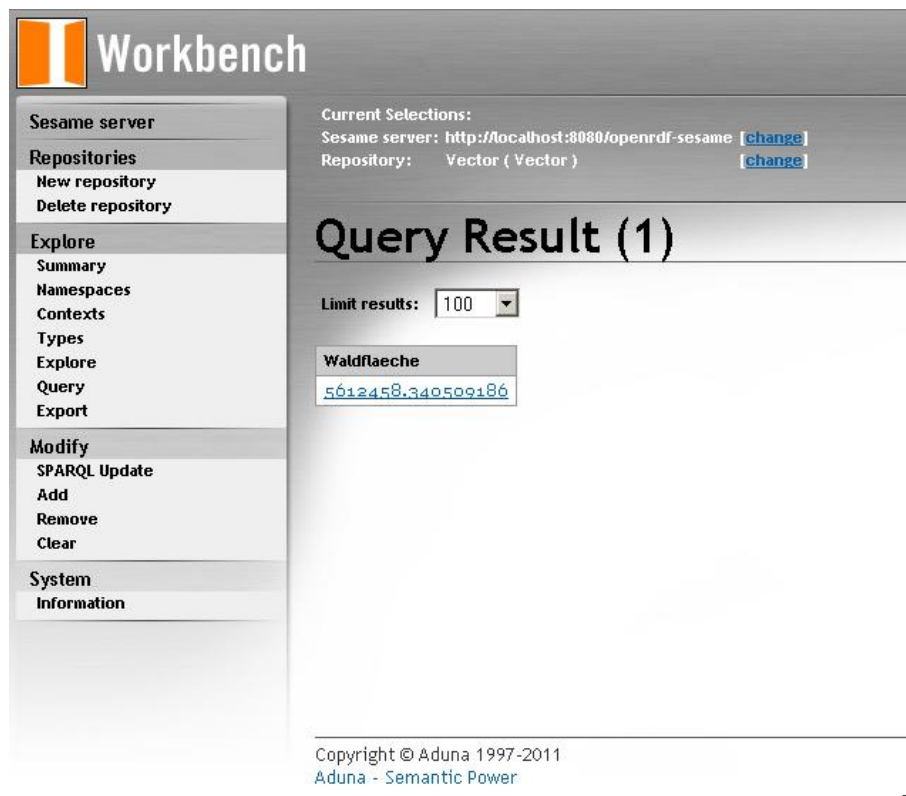


Abbildung 5.8: Resultat der SPARQL-Abfrage in m^2 nach der Waldfläche im Jagdbanngebiet “Trescolmen” (Ausschnitt der Benutzeroberfläche Sesame Workbench).

bei der Aufsummierung der Flächen zurückgeführt werden.

Der zweite Teil des Use Case bezieht sich auf die Höhenwerte und damit auf die Kleinsten Einheiten, da sie die Höheninformationen aus den Rasterzellen beinhalten. Das hatte einen rapiden Anstieg der zu verarbeitenden Datenmenge zur Folge. Auf das Gebiet Trescolmen bezogen bedeutet dies eine Erhöhung von 161 Primärflächen (aus dem ersten Teil) auf 12'084 Kleinste Einheiten, die jeweils im Waldgebiet liegen. Die Abfragen wurden mit einem Python Skript abgesetzt.

Die durchschnittliche Höhe des Waldes ergab sich durch die SPARQL-Abfrage im Quelltext 5.7:

```

1 SELECT DISTINCT (AVG(?Hoehe) AS ?DurHoehe)
2 WHERE {
3   ?Jagdbanngebiet jbVoc:jbName "Trescolmen"^^<http://www.w3.org/2001/
4     XMLSchema#string>.
5   ?Primaerflaeche ogc:within ?Jagdbanngebiet.
6   ?Primaerflaeche jbVoc:pfType "Z_Wald"^^<http://www.w3.org/2001/
7     XMLSchema#string> .
8   ?KlEinheit ogc:within ?Primaerflaeche.
9   ?KlEinheit jbVoc:hasHeight ?Hoehe.
10 }
```

Quelltext 5.7: SPARQL-Abfrage der Durchschnittshöhe.

Die Kleinsten Einheiten tragen die Höheninformationen und weisen räumliche Beziehungen zu den Jagdbanngebieten und Primärflächen auf. Da die Höhen der bewaldeten Flächen von Interesse sind, muss die Abfrage die Primärflächen einbeziehen, um zu ermitteln, welche Kleinsten Einheiten ins Waldgebiet fallen. Deshalb entsprechen die ersten drei Tripelmuster der vorangegangenen Abfrage zur Ermittlung der Waldfläche (Zeile 3 - 7). Zusätzlich müssen die Kleinsten Einheiten, die in den betroffenen Primärflächen liegen (Zeile 8), und deren Höhenwert bestimmt werden (Zeile 9). In der **SELECT**-Klausel änderte sich **SUM** zu **AVG**, da nicht mehr summiert, sondern der Durchschnitt berechnet wird.

Die durchschnittliche Höhe beträgt gerundet 1'375 Meter. Das Resultat ist aber nicht vollends korrekt. Wie bereits im Kapitel 5.4 erwähnt wurde, entspricht das extrahierte Raster nicht genau den Jagdbanngebietsgrenzen. Die so entstandenen Lücken haben bei der Verschneidung ("Union") aller Datenebenen den Höhenwert 0 zugewiesen bekommen. Auf die Waldflächen in Trescolmen bezogen, sind ca. 26'071 m² von insgesamt 5'586'387 m² Fläche betroffen. Deshalb wurden diese Flächen von der Auswertung ausgeschlossen.

In der SPARQL-Abfrage im Quelltext 5.8 wurde zusätzlich ein Filter hinzugenommen, der nur Höhen grösser als 0 zum Berechnen zulässt. Das ist legitim, weil der offiziell tiefste Punkt der Schweiz auf 194 Metern über Meer liegt¹⁴ und der tiefste Höhenwert im digitalen Höhenmodell DHM25 der gesamten Schweiz 157.6 Meter beträgt.

Im Rahmen der korrigierten Abfrage werden gleich alle gefragten Werte auf einmal ermittelt, also durch eine einmalige Abfrage.

```

1 SELECT DISTINCT (AVG(?Hoehe) AS ?DurHoehe) (MIN(?Hoehe) AS ?MinHoehe)
2               (MAX(?Hoehe) AS ?MaxHoehe)
3 WHERE {
4   ?Jagdbanngebiet jbVoc:jbName "Trescolmen"^^<http://www.w3.org/2001/
5     XMLSchema#string>.
6   ?Primaerflaeche ogc:within ?Jagdbanngebiet.
7   ?Primaerflaeche jbVoc:pfType "Z_Wald"^^<http://www.w3.org/2001/
8     XMLSchema#string> .
9   ?KlEinheit ogc:within ?Primaerflaeche.
10  ?KlEinheit jbVoc:hasHeight ?Hoehe.
11  FILTER (?Hoehe > 0 )
12 }
```

Quelltext 5.8: SPARQL-Abfrage der durchschnittlichen, minimalen und maximalen Höhe des Waldes.

¹⁴<http://www.eda.admin.ch/eda/de/home/doc/infoch/chgeog.html>, 23.11.2011

Die neu berechnete Durchschnittshöhe des Waldes beträgt gerundet 1379 Meter. Der höchste Punkt liegt auf 2054 Metern und der tiefste auf 756 Metern. Auch diese Ergebnisse wurden mit Hilfe des GIS überprüft. Entgegen dem ersten Teil des Anwendungsfalles konnten keine Abweichungen festgestellt werden.

Der vereinfachte Use Case “Waldfläche und Höhe simpel” kann mit dem Ansatz VKE beantwortet werden.

5.7 Diskussion und Fazit

In diesem Kapitel wurde der anhand des vereinfachten Use Case “Waldfläche und Höhe simpel” entwickelte Ansatz VKE vorgestellt. Zentrale Inhalte sind die vollumfängliche Konvertierung von Rasterdaten in das RDF-Format und die Erfassung der räumlichen Beziehungen. Die Datensätze wurden vorprozessiert, damit sämtliche benötigten Informationen vorlagen und direkt als Tripel ausgedrückt werden konnten. Die Grundanforderungen des vereinfachten Use Case wurden erfüllt. Trotzdem stellt der Ansatz aus diversen Gründen keine optimale Lösung dar, wie nachfolgend erläutert wird:

- Dem Ansatz VKE gehen einige manuelle Verarbeitungsschritte voraus, um die gewünschten RDF-Tripel generieren zu können. Hierbei wäre eine gewisse Automatisierung wünschenswert. Kommt es beispielsweise zu einer Aktualisierung der Ausgangsdaten, müsste die Vorprozessierung (oder zumindest Teile davon) wiederholt werden. Da sich in diesem Fall an der Struktur des Datenmodells und somit am Vorgehen nichts ändert, bietet es sich an, gewisse Schritte zu automatisieren.
- Ursprünglich bestand die Absicht zugunsten einer hohen Flexibilität möglichst alle Daten umzuformen. Das heisst, basierend auf den generierten Tripeln sollten unterschiedliche Fragestellungen beantwortet und neue Tripel einfach angefügt werden können.
Die kleinsten Einheiten, der grösste Datenbestand im Triple Store, basieren auf dem letzten Teil der Vorprozessierung. Diese Tatsache wirkt sich jedoch negativ auf die Flexibilität aus. Sollten neue Datenebenen hinzugefügt werden (unabhängig ob Raster oder Vektoren), könnte dies zur Entstehung neuer kleinsten Einheiten führen. Das heisst, die Vorprozessierung müsste unter Einbezug der neuen Daten wiederholt werden. Es sei denn, man verzichtet auf die topologische Einordnung der hinzugekommenen Daten, was aber den Ideen von VKE widerspricht.
- Der vereinfachte Anwendungsfall “Waldfläche und Höhe simpel” kann mit dem Ansatz VKE beantwortet werden. Betrachtet man VKE aber hinsichtlich der komplexen Use Cases, wird klar, dass aufgrund der fehlenden räumlichen Prozessierungsmöglichkeiten eine aufwändige Vorprozessierung nötig ist, um die Daten soweit aufzubereiten, damit basierend auf den generierten Tripeln die Anwendungsfälle gelöst werden können.
- Die Vektorisierung des Rasters führte zu einer beachtlichen Datenzunahme im Triple Store. Die Tabelle 5.1 zeigt eine Aufstellung des Speicherbedarfs der RDF-Daten. In der ersten Tabellenzeile ist zum Vergleich der benötigte Speicherplatz des ursprünglichen, auf die Jagdbanngebiete zugeschnittenen Rasterdatensatzes zu sehen. Es fällt sofort auf, dass der Speicherbedarf für die RDF-Daten, welche Geometrien beinhalten, durchwegs sehr hoch ist. Diese starke Datenzunahme darf nicht unbeachtet bleiben, besonders angesichts der geringen Menge der konvertierten Rasterdaten und Jagdbanngebiete. Das Höhenmodell hat zwar mit 25 Metern eine relativ feine Auflösung, doch sind die Daten auf die Jagdbanngebiete begrenzt, d. h. es handelt sich um kleine Ausschnitte.
Die mit VKE generierten Tripel, welche sich auf Geometrien beziehen, nehmen in Bezug auf die

Daten	Anzahl Tri- pel	Effektiver Speicher- bedarf	Speicherbedarf pro 1 Mil- lion Tripel
Ursprünglicher auf Jagdbanngebiete zuge- schnittener Rasterdatensatz	-	290 MB	-
RDF-Datei mit Beschreibung der Kleinsten Einheiten (Attribute, Beziehungen und Geo- metrien)	3.4 Mio.	3.6 GB	ca. 1 GB
RDF-Datei mit Geometrien der Jagdbannge- biete	342	2.2 MB	6.4 GB
RDF-Datei mit Geometrien der Primärflächen	126'831	80 MB	631 MB
RDF-Datei mit Sachattributen der Jagdbann- gebiete und Primärflächen (ohne Geometrien)	212'071	32 MB	151 MB

Tabelle 5.1: Speicherbedarf der RDF-Daten.

Tripel der Sachattribute wesentlich mehr Speicher in Anspruch. Vergleicht man diese Werte mit Arbeiten, die Triple Stores anhand von Benchmarks¹⁵ untersucht haben, bestätigt sich dieses Bild. In der Arbeit von Rohloff u. a. (2007) nahmen eine Million Tripel maximal 175 MB in Anspruch. Auch bei Bizer u. Schultz (2009) wurden für 100 Millionen Tripel in Sesame 17 GB beansprucht, was 170 MB pro Million entspricht.

VKE hat eine moderate Anzahl Tripel produziert, jedoch ist der Speicherbedarf vergleichsweise hoch. Dies kann auf die Umformung der Geometrien in das RDF-Format zurückgeführt werden, da jeweils sämtliche Stützpunkte konvertiert und somit gespeichert werden. Dadurch kann der Geometrie-String des Tripel, welcher die Stützpunkte enthält, sehr lang werden. Bereits jene Kleinsten Einheiten, welche ihre ursprüngliche Zellenform behalten haben, werden durch fünf Koordinatenpaare (Eckpunkte, wobei Anfangs- bzw. Startpunkt doppelt) beschrieben.

Fazit

Versucht man die Anwendungsfälle aus einer allgemeinen Perspektive zu betrachten, stellt sich die Frage, inwiefern sich der momentane Entwicklungsstand von Linked Data für Geodaten eignet. Ein Grossteil der Ansätze (auch VKE) versuchen einerseits die Daten in das RDF-Format zu konvertieren, andererseits kann das Potential von räumlichen Abfragen (attributiv oder geometrisch) nicht ausgeschöpft werden. Im Bereich der attributiven Abfragen müssten im Voraus alle möglichen topologischen Beziehungen ermittelt werden. Dennoch können keine quantitativen Aussagen, wie beispielsweise die Entfernung zwischen zwei Orten, getroffen werden. Die dazu benötigten geometrischen Abfragen sind noch nicht implementiert und es ist offen, ob sie die gleiche Effizienz wie mit einem GIS erreichen können. Der aufkommende Standard GeoSPARQL würde in diesem Bereich eine wichtige Lücke schliessen. Obwohl die Standardisierung praktisch sicher ist, gibt dies keine Garantie dafür, dass der Standard auch im beabsichtigten Umfang implementiert wird. Zudem deckt der Vorschlag zum GeoSPARQL-Standard die Thematik um Raster nicht ab (Perry u. Herring, 2011).

Zwar konnten die Ursprungsdaten im Ansatz VKE mit geringem Informationsverlust übernommen werden, doch ist der Nutzen fraglich. Ohne implementierte Möglichkeiten, die Daten im Triple Store zu prozessieren und somit neue Informationen zu generieren, kann nur das direkt abgefragt werden, was explizit abgelegt wurde. Sollte beispielsweise die durchschnittliche Steigung eines Jagdbanngebietes er-

¹⁵<http://www4.wiwi.fu-berlin.de/bizer/berlinparqlbenchmark/>
<http://swat.cse.lehigh.edu/projects/lubm/>

mittelt werden, müssen diverse Rechenschritte miteinbezogen werden und da stellt sich die Frage, ob SPARQL dafür geeignet ist.

Ein weiterer zentraler Aspekt von Linked Data ist die Verknüpfung mit fremden Datenquellen. Dabei ist fraglich, wie viel die kleinsten Einheiten dazu beitragen können. Isoliert betrachtet sind sie für den Benutzer semantikarm, da sie über keine aussagekräftigen Attribute verfügen. Zudem fragt sich, womit sie sinnvoll verknüpft werden können, wohingegen die Verknüpfung des gesamten Höhenrasters (z. B. zu einer Dokumentation des Datenanbieters) wieder mehr Sinn machen würde. Diese Problematik wird im Kapitel 6 noch genauer ausgeführt.

Ansatz “Ausgegliederte Räumliche Operationen” (ARO)

Der Ansatz der Vektorisierten Kleinsten Einheiten (VKE) bietet einen möglichen Weg mit Rasterdaten im Linked Data Kontext umzugehen. Jedoch vermochte er angesichts der Anforderungen der Use Cases nicht vollends zu überzeugen. Dies ist vor allem auf die fehlenden räumlichen Prozessierungsmöglichkeiten zurückzuführen, weshalb basierend auf den gewonnenen Erkenntnissen ein neuer Ansatz entwickelt wurde.

Der Ansatz der Ausgegliederten Räumlichen Operationen (ARO) hält nicht mehr strikt an der vorrangigen Verwendung von Linked Data Technologien fest, wie dies bei VKE gefordert wird. Das heisst, die Use Cases können nun auch unter Zuhilfenahme Linked Data fremder Technologien beantwortet werden. Die Umformung von konkreten Geodaten in das RDF-Format rückt bei ARO in den Hintergrund, während die benutzerfreundliche Bereitstellung von Daten und deren Verknüpfung im Vordergrund stehen.

Der Ansatz ARO wurde nicht im Hinblick auf einen einzigen spezifischen Anwendungsfall entwickelt, sondern sollte imstande sein, alle fünf Use Cases beantworten zu können.

In diesem Kapitel wird der Ansatz ARO mit seinen konzeptionellen Grundlagen, den modellierten räumlichen Beziehungen und dem allgemeinen Datenmodell erläutert. Zudem wird die Umformung der sogenannten Grundlagendaten in das RDF-Format und die Verknüpfung mit externen Datenquellen beschrieben.

6.1 Konzeptionelle Grundlagen

In einigen Ansätzen wird im Verlaufe der Datenprozessierung/-aufbereitung nicht auf den Gebrauch von räumlichen Operationen oder Processing Services (wie in GIS angeboten) verzichtet. So hängt die erfolgreiche Verwendung des Ansatzes VKE unter anderem von der vorausgehenden Prozessierung der Daten mit einem GIS ab. Dadurch können die Datensätze spezifisch vorbereitet werden, um daraus die benötigten Tripel zur Lösung der Use Cases zu generieren. Dazu zählen beispielsweise die topologischen Beziehungen. Auch der Ordnance Survey hat ein GIS verwendet, um die topologischen Beziehungen zu ermitteln (Goodwin u. a., 2008). Der USGS (Usery u. Varanka, 2011) setzt parallel zu Linked Data auf die Verwendung von GIS zur Prozessierung und grafischen Darstellung der Geodaten, indem die räumlichen RDF-Daten zusätzlich als eigenes Tripel im GML-Format abgelegt werden. Auch die komplexen Use Cases legen im Verlauf des Lösungsprozesses den Einsatz von räumlichen Operationen nahe.

Der Ansatz ARO versucht bewusst die Stärken unterschiedlicher Technologien zu vereinen, d. h., die räumlichen Prozesse werden an Systeme oder Operatoren ausgelagert, die speziell dafür erstellt wurden. Dadurch kann das Potential räumlicher Daten optimal ausgeschöpft werden. Folglich wird am grundsätzlichen Prinzip, die Geodaten mit räumlichen Operationen zu prozessieren, festgehalten. Ein mögliches alternatives Prinzip zur Prozessierung könnte die Definition einer detaillierten Ontologie sein, was in dieser Arbeit aber nicht weiter untersucht wird.

Durch die Auslagerung sollte ein wesentlicher Gewinn an Flexibilität entstehen. Denn mit ARO ist es nicht mehr nötig die Daten vor der Konvertierung in das RDF-Format zu prozessieren, sondern je nach Anwendungsfall können die spezifischen Operationen aufgerufen und aus den Ergebnissen die RDF-Daten generiert werden. Auch bei der Hinzunahme neuer Datensätze müssen lediglich deren Metadaten im Triple Store ergänzt werden, da sie bei der Prozessierung dynamisch eingebunden werden können.

Ein wesentlicher Grund für die Ausgliederung der räumlichen Prozesse ist das Fehlen von ausgereiften Lösungen aus dem Umfeld von Linked Data zur Prozessierung und Abfrage von Geodaten. Deshalb wird bei ARO ein Grossteil der räumlichen Funktionalität an ein GIS bzw. dessen Operatoren ausgelagert. Dies betrifft sowohl den Bereich der Datenhaltung, als auch den der Datenprozessierung. Sämtliche Geodaten, inklusive den Rastern, werden in räumlichen Datenbanken abgelegt (Personal Geodatabase bzw. File Geodatabase). Nur ausgewählte Daten sollen komplett in das RDF-Format umgeformt werden. Dazu zählen beispielsweise die Jagdbanngebiete, welche inklusive den Geometrien in den Triple Store gespeichert werden. Für die anderen Daten werden lediglich die Metainformationen als Tripel abgelegt, sowie eine Referenz zum Speicherort. Somit hat ARO nicht den Anspruch, sämtliche Daten möglichst getreu dem Original in das RDF-Format zu überführen und in einem Triple Store abzulegen. Trotzdem kann aber auf alle Daten zugegriffen werden, wovon jedoch nur ein Teil direkt im RDF-Format verfügbar ist. Welche Daten (besonders Geodaten) in RDF repräsentiert werden, entscheidet sich am erwarteten Mehrwert für Benutzer und ob, gemäss dem dritten Linked Data Designprinzip, nützliche Informationen als Tripel bereitgestellt werden können (Berners-Lee, 2009). Das folgende Beispiel erläutert, weshalb die einzelnen Jagdbanngebiete in RDF umgeformt und warum die Objekte der Primärflächen im ursprünglichen Format behalten werden:

Jagdbanngebiete haben eine grössere Fläche, basieren auf einer gesetzlichen Grundlage und deren Namen werden oft in anderen Zusammenhängen verwendet (Toponyme). So steht der Name “Mythen” nicht nur für das gleichnamige Jagdbanngebiet, sondern auch für eine Region¹, ein Bergmassiv², ein Einkaufszentrum³ usw. Dadurch können Menschen den Begriff “Mythen” in einen Kontext einordnen. Primärflächen hingegen haben keine Namen, sind lediglich nach ihrer Objektart (z. B. Wald, Sumpf) klassiert und haben als Einzelobjekte für den Menschen eine geringe Aussagekraft, da die Datenklassierung nicht “menschenfreundlich” ist. Sie besitzen aus menschlicher Perspektive wenig semantische Informationen. Eine Verknüpfung zu fremden Ressourcen gestaltet sich schwierig. Beispielsweise kann ein durchschnittlicher Benutzer mit dem Waldgebiet, das die Nummer “50934” trägt und eine Fläche von 460 m² hat, nur wenig anfangen. Gleich verhält es sich mit der Suche nach passenden Verknüpfungen. Wohingegen ein Objekt, das den gesamten Wald des Schweizerischen Nationalparks repräsentiert, wieder von grösserem Interesse ist und somit mit den Geometrien und Sachattributen als Tripel abgelegt werden würde. Insgesamt entsteht dadurch eine relativ schlanke Struktur, mit nur wenig konkreten Geodaten im RDF-Format.

Die Abbildung 6.1 zeigt eine schematische Darstellung des Ansatzes ARO. Gemäss dem jeweiligen Use Case (1) wird ein Python-Skript erstellt (2), das die benötigten Arbeitsschritte durchführt. Dadurch lassen sich das GIS bzw. dessen räumliche Operationen (3) und der Triple Store (3) steuern, sowie eigene Tripel erstellen (4). Ein exemplarischer Ablauf von ARO könnte folgendermassen aussehen:

In den Triple Store werden die Grundlagendaten (vgl. Kapitel 6.4) gespeichert. Sie bilden den Ausgangspunkt für die Use Cases. Darauf werden mittels SPARQL-Abfragen zum einen die betroffenen Jagdbanngebiete (Objekte) und zum anderen die benötigten Metadaten ermittelt (besonders wichtig sind die Speicherpfade zu den Geodaten). Dadurch können nun die ursprünglichen Geodaten in den räumlichen

¹ <http://www.mythenregion.ch/>

² <https://de.wikipedia.org/wiki/Mythen>

³ <http://www.mythen-center.ch/>

Datenbanken abgerufen und die benötigten räumlichen Operationen mit den geforderten Parametern initiiert werden. Danach folgt die Prozessierung der Daten, bis aus den Ergebnissen die erforderlichen Tripel generiert werden können. Diese Tripel werden im Triple Store gespeichert, in die bestehende Datenstruktur eingegliedert und allenfalls mit fremden Datenquellen verknüpft.

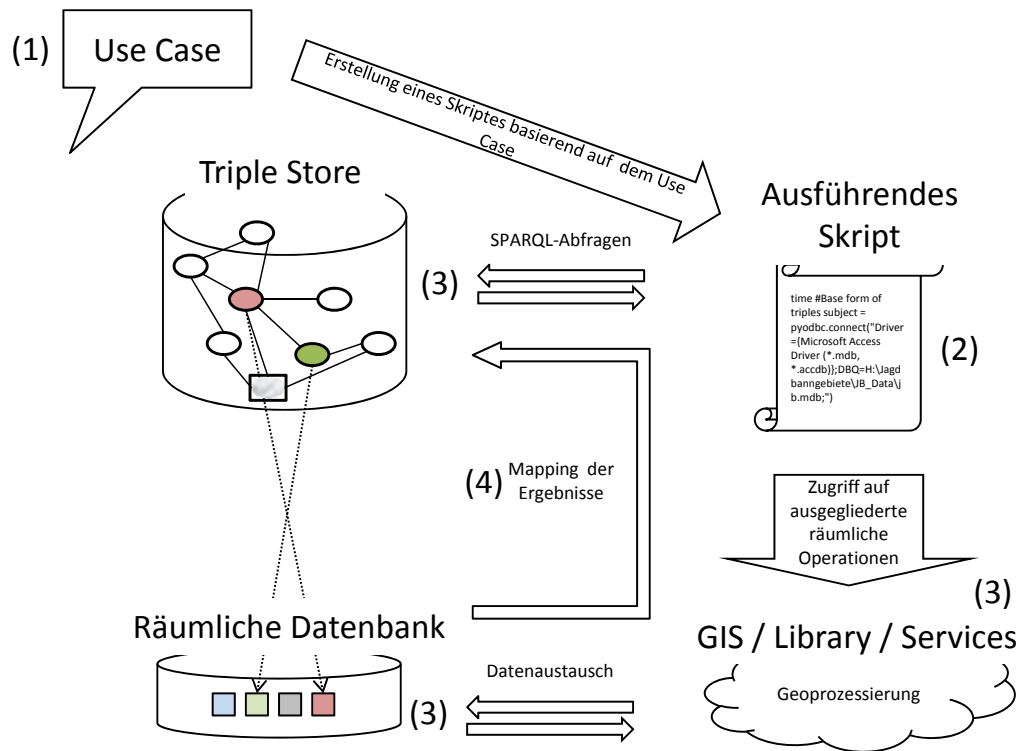


Abbildung 6.1: Schematische Darstellung des Ansatzes ARO.

Beim Ansatz ARO hat Linked Data nicht mehr die Funktion komplexe räumliche Abfragen zu ermöglichen, sondern Informationen (oft auf übergeordneter Ebene) bereitzustellen und zu verknüpfen. Der Benutzer soll durch Linked Data in Erfahrung bringen können, was für Datensätze existieren und über welche Metadaten sie verfügen. Zusätzlich werden einige wenige konkrete Geodaten (Objekte) im RDF-Format gespeichert, die für den Benutzer von breiterem Interesse sind und die sinnvoll mit fremden Ressourcen verknüpft werden können.

6.2 Vergleich VKE und ARO

Der folgende Vergleich zwischen den Ansätzen VKE und ARO zeigt die wesentlichsten Unterschiede und die beabsichtigten Verbesserungen auf:

- Der Umfang zur Vorprozessierung der Daten ist bei ARO wesentlich kleiner. Die Ausgangsdaten bedürfen einer geringen Bearbeitung und können direkt als Tripel abgelegt werden. Die topologischen Beziehungen werden wie bei VKE mit Hilfe eines GIS ermittelt.
- Je nach Use Case werden verschiedene räumliche Operationen benötigt. Für den Ansatz VKE be-

deutet dies, dass für jeden Anwendungsfall die Daten spezifisch aufbereitet werden müssen. Somit wird der Triple Store jeweils mit diesen spezifisch benötigten Daten ergänzt.

Bei ARO hingegen besteht ein Grundstock an RDF-Daten im Triple Store, der unabhängig vom Use Case bestehen bleibt. Nur die entsprechenden Resultate werden allenfalls zusätzlich in den Triple Store geladen. Dies dient vor allem der Verknüpfung und zur Entdeckung neuer Daten, beispielsweise mit einem Linked Data Browser. Daher weist ARO eine vielseitigere Verwendbarkeit (Flexibilität) auf. Wird beispielsweise ein weiteres digitales Höhenmodell berücksichtigt, können einfach die Metadaten im Triple Store hinzugefügt werden. Je nachdem, ob der Datensatz für die aktuelle Fragestellung relevant ist oder nicht, kann er in die Geoprozessierung miteinbezogen werden.

- ARO kann grundsätzlich sämtliche Use Cases ohne explizit angegebene topologische Beziehungen lösen. Soll beispielsweise die Felsfläche eines Jagdbanngebietes bestimmt werden, kann dies unter Verwendung von räumlichen Operationen durchgeführt werden, ohne dass räumliche Beziehungen zwischen diesen Objekten im Triple Store definiert sind. Diese Unabhängigkeit von den räumlichen Beziehungen erlaubt die Erstellung eines schlankeren Datenmodells.
Beim Ansatz VKE hingegen muss die Topologie explizit abgelegt sein, damit die als “Fels” gekennzeichneten Objekte den jeweiligen Jagdbanngebieten räumlich zugeordnet werden können.
- Da die Geodaten jeweils vollständig in einer räumlichen Datenbank abgelegt und die räumlichen Prozessierungen ausgelagert werden, können beliebige räumliche Operationen durchgeführt werden. Diese reichen weit über das vorgesehene Spektrum von GeoSPARQL hinaus. Dadurch ist es möglich Use Cases zu beantworten, die eine komplexe Geoprozessierung erfordern.
- Bei ARO werden Geodaten mit ihren Geometrien und Sachattributen zurückhaltend in das RDF-Format umgeformt. Dadurch kann Speicherplatz eingespart werden, besonders auch weil die RDF-Tripel der Rastergeometrien wegfallen.
ARO speichert nur Geodaten mit Sachattributen und Geometrien ab, wenn diese als potentiell nützlich für den Benutzer angesehen werden. Beispielsweise kann ein Benutzer mit einem einzelnen Objekt aus den kleinsten Einheiten des Ansatzes VKE wenig anfangen.

6.3 Modellgrundlagen ARO

6.3.1 Modellierung der räumlichen Beziehungen

Für die Lösung der Anwendungsfälle ist der Ansatz ARO auf keine explizite Angaben räumlicher Beziehungen im RDF-Format angewiesen. Ein solcher Verzicht würde aber die Linked Data Komponente stark vernachlässigen und ist daher nicht im Sinne von ARO. Die Angabe der topologischen Beziehungen beschränkt sich auf Klassen, die konkrete räumlich referenzierte Objekte repräsentieren. Zwar könnten auch ganze Geodatenätze räumlich zueinander verortet werden, doch wären diese Beziehungen eher diffus. Werden beispielsweise die Primärflächen auf die Jagdbanngebiete zugeschnitten, könnten sie als “innerhalb eines Jagdbanngebietes” (`ogc:within`) verortet werden. Jedoch bleibt unklar welche Objekte genau innerhalb welcher Jagdbanngebiete liegen. Hingegen können bei der Verknüpfung der einzelnen Objekte diese zu anderen Objekten explizit in Beziehung gesetzt werden. So können beispielsweise die Jagdbanngebiete innerhalb einer biogeographischen Region liegen (`ogc:within`) oder mindestens zwei überlappen

(`ogc:overlaps`). Dies ist auch für potentielle Benutzer interessant, da somit über einfache SPARQL-Abfragen die Position eines Objektes zu anderen ermittelt werden kann. Die räumlichen Beziehungen werden, wie schon bei VKE, gemäss dem vorgeschlagenen GeoSPARQL-Standard definiert.

6.3.2 Datenmodell

Das entwickelte Datenmodell (Abbildung 6.2) deckt alle fünf Use Cases ab und enthält alle relevanten Klassen. Es kann bei Bedarf einfach erweitert werden, indem neue Klassen integriert werden. Das Datenmodell stellt eine Weiterentwicklung des Modells aus VKE dar. Die Jagdbanngebiete bilden ebenfalls die zentrale Klasse, sind aber hierarchisch untergeordnet.

An hierarchisch höchster Stelle steht die Klasse **ProJBKomponenten**. Sie beschreibt jegliche Ressourcen (im Rahmen der Anwendungsfälle), die im Zusammenhang mit der Thematik um die Jagdbanngebiete stehen. Dementsprechend sind die Attribute wenig spezifisch und sollen die Ressource auf allgemeine Weise repräsentieren, z. B. durch eine kurze Beschreibung. Dafür wurde auf das Vokabular “Dublin Core Metadata Element Set” (`dc`) zurückgegriffen. **ProJBKomponenten** werden festgelegte Typen zugeordnet. So kann es sich bei der Ressource um Geodaten, um eine Dokumentation, oder um Konzeptionelle Grundlagen handeln. Damit findet eine erste Spezifikation statt.

Der Klasse **Geodaten** gehören sämtliche Ressourcen an, die Gegenstände, Geländeformen und Infrastrukturen an der Erdoberfläche mit räumlichem Bezug repräsentieren und über Geometrien verfügen (Bill u. Zehner, 2001). Die Inhalte der Datensätze werden in dieser Klasse aber nicht gespeichert, sondern eine Referenz aus dem Triple Store zur räumlichen Datenbank. Das heisst, die Attribute verweisen zum einen auf den Namen und zum anderen auf den Speicherpfad der Datensätze. Dadurch können über SPARQL-Abfragen die benötigten Datensätze ermittelt und die entsprechenden Parameter an die räumlichen Operationen übergeben werden. Diese Klasse bildet einen zentralen Bestandteil von ARO zur Nutzung der externen Technologien. Sie übernimmt eine Brückenfunktion vom Triple Store zum GIS.

Die Klasse **Dokumentationen** ist analog zu den Geodaten aufgebaut, jedoch werden statt räumliche Daten, Ressourcen mit beschreibendem Charakter referenziert. Die Beschreibungen müssen sich auf Ressourcen innerhalb des Datenmodells beziehen, können aber von unterschiedlicher Art sein, wie schriftliche Dokumente, Bilder, Tonaufnahmen etc. So können beispielsweise den Geodaten die Datenbeschreibungen oder den Jagdbanngebieten die Objektbeschreibungen angehängt werden.

Die Klasse **KonzeptionelleGrundlagen** repräsentiert Ressourcen, die von Bedeutung für die Lösung einzelner Use Cases sind, aber keine Geometrien oder Ressourcen mit dokumentierendem Charakter beinhalten. Sie enthalten Informationen, die besonders in Kombination mit anderen Ressourcen verwendet werden. Im Datenmodell gehört ihr lediglich die Klasse **Vegetationshoehestufen** an. Die Höhenstufen können zwar räumlich den **BiogeographischeRegionen** zugeordnet werden, sind aber nicht diskret. Deren genaue räumliche Abgrenzung ergibt sich erst in Kombination mit den **Geodaten** (Höhenmodell). Denn je nach digitalem Höhenmodell nehmen die einzelnen Höhenstufen unterschiedliche Ausdehnungen an.

Den **Geodaten** sind weitere Klassen untergeordnet, die konkrete Objekte aus den Geodatenätzen repräsentieren. Dazu zählt die zentrale **Jagdbanngebiete** Klasse, welche den Ausgangspunkt für die Abfragen der Anwendungsfälle darstellt. Sie besitzt die gleichen Attribute wie bei VKE, bis auf die Fläche und das neue Attribut `jbVoc:aehnlichWie`. Das letztere Attribut wird jeweils dynamisch in Abhängigkeit des komplexen Use Case 3 “Ähnliche Gebiete” generiert. Das heisst, dieses Attribut ist temporär und es besitzen dieses nur, die durch den Anwendungsfall betroffene Objekte. Zu den Geodaten gehört auch die Klasse **BiogeographischeRegionen**, welcher neben den Attributen für die ID und den Namen auch eines für die Beschreibung der einzelnen Objekte angehört (`dc:description`). Die Geometrien werden wie bei VKE

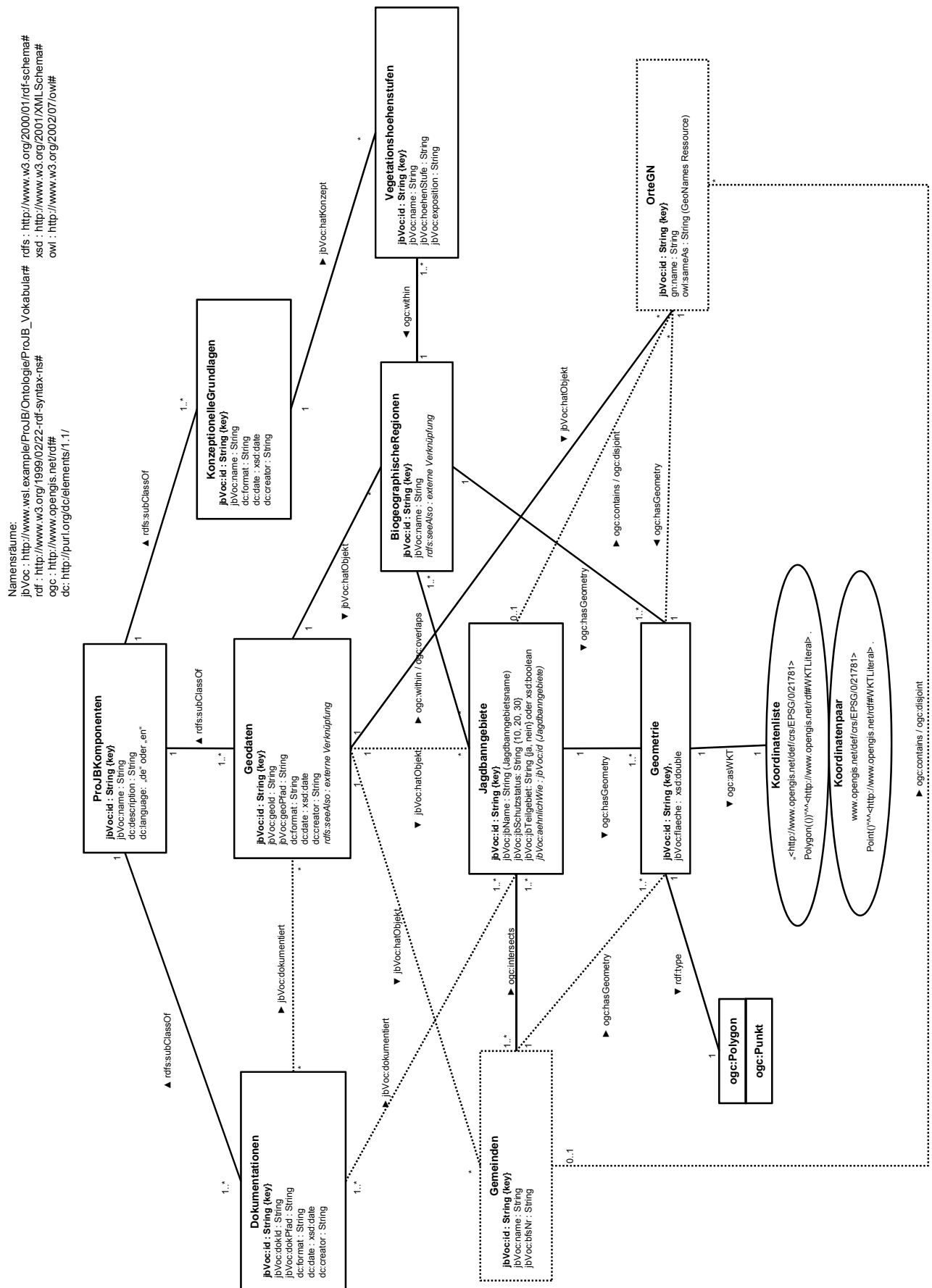


Abbildung 6.2: Datenmodell des Ansatzes ARO.

abgelegt, nach dem vorgeschlagenen GeoSPARQL-Standard, jedoch mit einem kleinen Unterschied: Die Fläche einer Geometrie ist nun ein Attribut der Klasse `Geometrie`. Da im Modell Punkte oder Polygone vorgesehen sind, ist dieses Attribut nur bei Polygonen vorhanden.

Die vorgestellten Klassen sind beständig im Triple Store vorhanden. Daneben existieren Klassen wie `OrteGN` und `Gemeinden`, die temporär sind und sich abhängig von den Use Cases zusammensetzen. `OrteGN` und `Gemeinden` stellen die Abfrageergebnisse aus dem zweiten komplexen Use Case "Betroffene Gemeinden" dar. Je nachdem welche Jagdbanngebiete abgefragt werden, ändert sich deren Inhalt, da jeweils unterschiedliche Gemeinden betroffen sein werden.

6.4 Umformung der Grundlagendaten

Der Begriff Grundlagendaten bezeichnet die im Triple Store permanent gespeicherten RDF-Daten, welche unabhängig von den Anwendungsfällen vorhanden sind. Sie umfassen RDF-Tripel zu den Datensätzen VECTOR25/VECTOR200, DHM25/DHM RIMINI, Jagdbanngebiete und Biogeographische Regionen. Die Tripel werden gemäss den Klassen `ProJBKomponenten`, `Dokumentationen`, `Geodaten`, `Biogeographische Regionen` und `Jagdbanngebiete` generiert. Jedoch werden nur bei den beiden letzteren konkrete Geoobjekte (Sachattribute + Geometrien) in RDF umgeformt, weshalb die Klassen einem eigenen Datensatz entsprechen. Die Landschafts- und Höhenmodelle werden lediglich mit ihren entsprechenden Metadaten abgelegt. Ein Grossteil der Tripel wird mit Mapping-Skripten generiert, wobei stellenweise auch das Mapping Programm D2RQ zum Einsatz kommt. Die Quelltexte 6.1 und 6.2 zeigen jeweils einen Auszug der generierten Tripel, gemäss den Klassen `ProJBKomponenten` bzw. `Geodaten`, zum Datensatz VECTOR25.

```

1 <http://www.wsl.example/ProJB/Daten/DT2>
2   <http://www.wsl.example/ProJB/Ontologie/ProJB_Vokabular#name>
3   "VECTOR25 Primaerflaechen"^^<http://www.w3.org/2001/XMLSchema#string> .
4 <http://www.wsl.example/ProJB/Daten/DT2>
5   <http://purl.org/dc/elements/1.1/creator>
6   "Bundesamt fuer Landestopographie, swisstopo"^^<http://www.w3.org/
7       2001/XMLSchema#string> .
8 <http://www.wsl.example/ProJB/Daten/DT2>
9   <http://purl.org/dc/elements/1.1/format>
10  "ESRI Feature Class"^^<http://www.w3.org/2001/XMLSchema#string> .

```

Quelltext 6.1: Auszug der generierten Tripel der Primärflächen von VECTOR25 gemäss der Klasse `ProJBKomponenten`.

```

1 <http://www.wsl.example/ProJB/Daten/RD/RD2>
2   <http://www.wsl.example/ProJB/Ontologie/ProJB_Vokabular#id>
3   "RD2"^^<http://www.w3.org/2001/XMLSchema#string> .
4 <http://www.wsl.example/ProJB/Daten/RD/RD2>
5   <http://www.wsl.example/ProJB/Ontologie/ProJB_Vokabular#geoID>
6   "vec25_primflae"^^<http://www.w3.org/2001/XMLSchema#string> .
7 <http://www.wsl.example/ProJB/Daten/RD/RD2>
8   <http://www.wsl.example/ProJB/Ontologie/ProJB_Vokabular#geoPfad>
9   "H:/Jagdbanngebiete/ JB_Data/priflae.mdb/vec25_primflae"^^<http://
10  www.w3.org/2001/XMLSchema#string> .

```

Quelltext 6.2: Auszug der generierten Tripel der Primärflächen von VECTOR25 gemäss der Klasse `Geodaten`.

Den drei Tripeln der Klasse `ProJBKomponenten` (Quelltext 6.1) kann entnommen werden, dass es sich um eine Ressource mit dem Namen `VECTOR25 Primaerflaechen` (Zeilen 1 - 3), mit dem Urheber `swisstopo` (Zeilen 4 - 7) und dem Format `ESRI Feature Class` (Zeile 8 - 10) handelt. Bereits daraus lässt sich schliessen, dass es sich um räumliche Daten handelt. Die Tripel der Klasse `Geodaten` (Quelltext 6.2) vergeben die ID `RD2` (Zeile 1 - 3) sowie den Namen (Zeilen 4 - 6) und Speicherpfad (Zeilen 7 - 10) des originalen Geodatensatzes.

6.5 Verknüpfung der Grundlagendaten mit externen Quellen

Mit dem Ansatz ARO wird die Verknüpfung mit fremden Datenquellen implementiert, was bei VKE eher im Hintergrund stand. Dadurch wird einer der Kerngedanken von Linked Data realisiert. Durch die Verknüpfung unterschiedlichster Datenquellen soll ein globaler Datenraum entstehen, der die Entdeckung von neuen zusätzlichen Daten ermöglichen soll (Heath u. Bizer, 2011).

Die Grundlagendaten werden manuell mit fremden Quellen verknüpft. Das heisst, es wird gezielt nach passenden Ressourcen aus fremden Quellen gesucht. Der Service `<sameAs>`⁴ erweist sich dabei als hilfreich. Er ermöglicht einen URI-basierten bzw. eine textbasierte Suche nach ähnlichen oder gleichen Ressourcen. Die Abbildung 6.3 zeigt einen Ausschnitt der Resultate aus der textbasierten Suche nach dem Begriff “Jagdbanngebiet”. Bereits die ersten Suchergebnisse verweisen auf DBpedia Ressourcen (in unterschiedlichen Sprachen), die sich auf die Jagdbanngebiete beziehen. Sie stellen eine ideale Verknüpfung in den Linked Data Raum bereit, da sehr viele andere Datenbestände mit DBpedia verknüpft sind. Der Quelltext 6.3 enthält ein Tripel, das die Klasse der Jagdbanngebiete mit einer entsprechenden englischsprachigen DBpedia Ressource verknüpft. Dazu wurde nicht das Prädikat `owl:sameAs` verwendet, sondern `rdfs:seeAlso`, da die Ressourcen nicht das Gleiche repräsentieren. Die Klasse bezieht sich auf die realen Jagdbanngebiete, während die DBpedia Ressource eine Liste dieser Gebiete verkörpert.

Auf diese Weise wurden diverse Ressourcen, wie einzelne Jagdbanngebiete oder biogeographische Regionen, anhand der Namen mit DBpedia verknüpft.

```

1 <http://www.wsl.example/ProJB/Ontologie/ProJB_Vokabular#Jagdbanngebiet>
2   <http://www.w3.org/2000/01/rdf-schema#seeAlso>
3   <http://dbpedia.org/resource/List_of_federal_hunting_
4     reserves_in_Switzerland> .

```

Quelltext 6.3: Verknüpfung der Klasse `Jagdbanngebiet` mit DBpedia.

⁴ <http://www.sameas.org/>

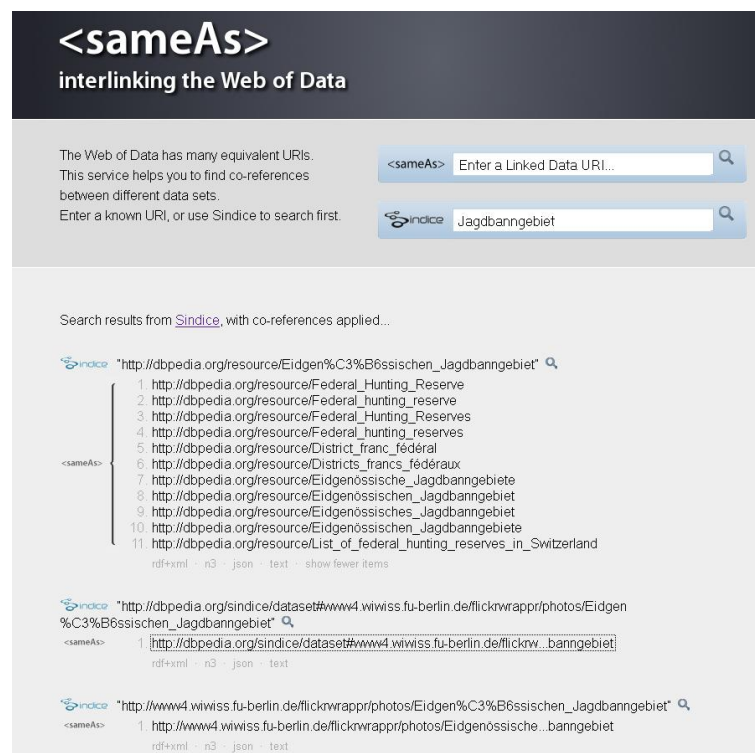


Abbildung 6.3: Resultate der Suche nach “Jagdbanngebiet” über den Service <sameAs>.

6.6 Diskussion

Das Kapitel hat den Ansatz der Ausgliederten Räumlichen Operationen (ARO) beschrieben. Er wurde basierend auf den Erfahrungen von VKE entwickelt. ARO bedient sich unterschiedlicher Technologien und bewegt sich daher nicht mehr strikte im Linked Data Umfeld. Um einen möglichst grossen Nutzen aus den Geodaten zu ziehen, werden räumliche Operationen einbezogen und teilweise die Datenhaltung an dafür geeignete Komponenten ausgelagert. Linked Data dient somit vor allem der Verknüpfung und Beschreibung von Daten. Nur ausgewählte Geodaten werden in das RDF-Format umgeformt. Die Einzelobjekte dieser Daten müssen für den Benutzer aussagekräftig und sinnvoll verknüpfbar sein. Geodatenätze mit weniger relevanten Objekten werden über ihre Metadaten beschrieben, wobei die eigentlichen Daten in räumlichen Datenbanken gehalten werden. Mit dem Ansatz ARO sollten alle Anwendungsfälle beantwortet werden können.

Der Ansatz ARO basiert auf einer grundlegenden Annahme, dass Benutzer von Linked Data tendenziell an allgemeinen Informationen interessiert sind. Beispielsweise im Sinne von: Was für Daten existieren? Wo kann ich sie finden? Welche weiteren Daten sind damit verknüpft? Liegt das Interesse auf detaillierten Daten und deren Auswertung (wie bei den komplexen Use Cases), kann der Benutzer wohl als Spezialist bezeichnet werden. Dieser würde für solche Prozesse wahrscheinlich geeignetere Komponenten, wie GIS oder generell räumliche Operationen und somit auch Geodatenformate statt dem RDF-Format bevorzugen. Der Ansatz VKE beispielsweise hat die Geodaten in das RDF-Format konvertiert, jedoch bleibt offen, ob ein Spezialist nicht lieber die Originaldaten (besonders bei den Rastern) gehabt hätte, eventuell im GML-Format. ARO zielt bereits in diese Richtung und könnte modifiziert werden, so dass die Referenz zum Speicherort der Daten durch einen Downloadlink ersetzt wird, worüber die Daten direkt im GML-Format bezogen werden könnten.

Im Rahmen von ARO wird auch die Frage aufgeworfen, in welchem Ausmass Daten in RDF konvertiert und als Linked Data bereitgestellt werden sollen. Dies kann als eine Grundsatzfrage verstanden werden und betrifft unterschiedlichste Datenarten. Geodaten können über ihre Metadaten in Linked Data integriert und verknüpft werden, ohne den konkreten Inhalt (Objekte, Rasterzellen) miteinzubeziehen. Das gleiche Prinzip kann beispielsweise auf Textdateien übertragen werden. Auch sie können alleine über Metadaten in Linked Data integriert werden, wobei eine Referenz auf den Speicherort des eigentlichen Dokumentes verweist. Würde der ganze Text als ein String innerhalb eines Tripel abgelegt, könnte dies gleichgesetzt werden, wie wenn ein Raster im GML Coverage Format in einem String abgelegt wird. Dadurch wird der Inhalt direkt zugreifbar, ist aber unstrukturiert vorhanden. Wird der Text auseinandergenommen und in Fragmenten abgelegt, ist dies vergleichbar mit dem Ablegen einzelner Objekte oder gar einzelner Zellen eines Rasters (wie bei VKE). Dabei stellt sich wiederum die Frage nach dem Mehrwert. Denn solche Fragmente haben tendenziell für Menschen weniger Bedeutung, sind also anonymer. So sind Verknüpfungen zu fremden Datenquellen auf einer solch anonymen Ebene schwieriger zu realisieren, als auf einer allgemeinen Ebene, wie es beispielsweise die Metadaten bieten.

Andererseits existieren durchaus gute Gründe auch den Inhalt von Daten in RDF zu repräsentieren: Werden Geodaten in das RDF-Format gemäss einer einheitlichen Syntax konvertiert, erleichtert dies die Austauschbarkeit und den Zugang. Falls im Zuge der Einführung von GeoSPARQL auch die räumlichen Prozessierungsmöglichkeiten implementiert werden, täten sich dem Benutzer viele Möglichkeiten auf. So könnten direkt über Linked Data räumliche Abfragen gemacht werden. Die Resultate wären somit bereits im Triple Store integriert und möglicherweise mit fremden Daten verknüpft.

Der Ansatz ARO versucht eine **Zwischenlösung** zu bieten, indem nur Geodaten in das RDF-Format konvertiert werden, die keinen allzu spezifischen Detaillierungsgrad aufweisen und dadurch für die Allgemeinheit der Benutzer von Interesse sein könnten. Die Rasterdaten werden bei ARO bisher nur über die Metadaten in Linked Data integriert, was beispielsweise durch den Ansatz des USGS (Usery u. Varanka, 2011) optimiert werden könnte. Denn gerade auch Rasterdaten können Objekte enthalten, die von breiterem Interesse sind. Dadurch werden auch die als wichtig erachteten Objekte aus dem Raster in Linked Data integriert.

In den folgenden Kapiteln wird die praktische Implementierung des Ansatzes ARO anhand der Use Cases vorgestellt.

Einfache Use Cases

Die einfachen Use Cases können basierend auf den Grundlagendaten im Triple Store (gemäss dem Datenmodell ARO) beantwortet werden. Dazu werden lediglich die Klassen “Jagdbanngebiete” und “Geometrie” benötigt. Das heisst, es werden keine zusätzliche räumliche Operationen oder Daten verwendet. Die einfachen Use Cases lauten:

- Einfacher Use Case 1 “Attribute abfragen”:
Welche Jagdbanngebiete gehören einer bestimmten Schutzkategorie an bzw. beinhalten Teilgebiete, die dieser Schutzkategorie zugeordnet werden können?
- Einfacher Use Case 2 “Aggregation”:
Wie gross ist die Fläche aller Jagdbann- bzw. Teilgebiete, die einer bestimmten Schutzkategorie angehören?

In diesem Kapitel werden das Vorgehen zur Lösung der Anwendungsfälle erläutert, sowie die Resultate präsentiert. Zum Schluss folgt eine kurze Diskussion der Ergebnisse.

7.1 Vorgehen und Resultate

Da sämtliche benötigten Daten bereits im Triple Store vorliegen, gestaltet sich das Vorgehen einfach. Die SPARQL-Abfragen können über die Sesame Workbench abgesetzt werden. Der Use Case “Attribute abfragen” benötigt lediglich ein einfaches Graph-Muster, bestehend aus zwei Tripelmustern, wie im Quelltext 7.1 dargestellt. Im Beispiel werden sämtliche Jagdbanngebiete bzw. deren Teilgebiete mit der Schutzkategorie 10 gesucht (Zeile 6). Also alle Objekte, die integral geschützte Gebiete repräsentieren. Zuerst werden die entsprechenden Objekte aus der Klasse Jagdbanngebiete mit dem Schutzstatus “10” ausgeschieden. Dabei handelt es sich aber je nachdem um ganze Jagdbanngebiete oder um Teilgebiete. Das Attribut `jbVoc:jbName` ist unabhängig von dieser möglichen Aufteilung und beinhaltet den Namen des Gebietes, d. h. die Teilgebiete werden über den Namen den entsprechenden Jagdbanngebieten zugeordnet. Deshalb muss mit dem zweiten Tripelmuster lediglich noch der Name abgefragt werden (Zeile 7). Da ein Jagdbanngebiet mehrere Teilgebiete mit dem gleichen Schutzstatus haben kann, wird im SELECT-Abschnitt der Abfrage `DISTINCT` eingefügt. Dadurch lassen sich Duplikate im Ergebnis vermeiden.

```

1 PREFIX jbVoc:<http://www.wsl.example/ProJB/Ontologie/ProJB_Vokabular#>
2 PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
3
4 SELECT DISTINCT ?Jagdbanngebiet
5 WHERE{
6 ?Teilgebiet jbVoc:jbSchutzstatus "10"^^xsd:string .
7 ?Teilgebiet jbVoc:jbName ?Jagdbanngebiet .
8 }

```

Quelltext 7.1: SPARQL-Abfrage für den einfachen Use Case “Attribute abfragen”.

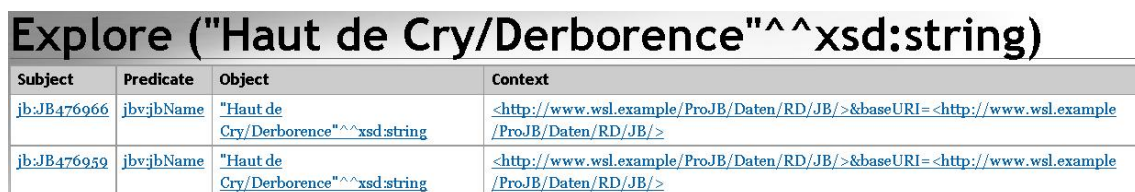
Das Resultat der Abfrage ist eine Liste sämtlicher Jagdbanngebiete, die dieser Schutzkategorie angehören, bzw. die Teilgebiete dieser Kategorie beinhalten. Die Abbildung 7.1 stellt einen Ausschnitt des Ergebnisses aus der Sesame Workbench dar. Demnach sind 36 von 41 Jagdbanngebieten integral geschützt bzw. beinhalten integral geschützte Teilgebiete.

Die Abbildung 7.2 zeigt eine Auflistung der Teilgebiete des Jagdbanngebietes “Haut de Cry/Derborence”. Diese erscheint, wenn auf die entsprechende Ressource in Abbildung 7.1 geklickt wird. Klickt man nun eines dieser Teilgebiete an, erscheint eine Darstellung mit allen zu dieser Ressource zugehörigen Tripeln. In Abbildung 7.3 wurde das Teilgebiet “JB476966” ausgewählt.



Query Result (36)	
Limit results:	100
Jagdbanngebiet	
"Dixence"^^xsd:string	
"Turtmantal"^^xsd:string	
"Leukerbad"^^xsd:string	
"Haut de Cry/Derborence"^^xsd:string	
"Val Ferret / Combe de l'A"^^xsd:string	
"Mauvoisin"^^xsd:string	
"Bietschhorn"^^xsd:string	

Abbildung 7.1: Ausschnitt der Ergebnisse der Abfrage aus Quelltext 7.1 in der Benutzeroberfläche Sesame Workbench.



Explore ("Haut de Cry/Derborence"^^xsd:string)			
Subject	Predicate	Object	Context
jb:JB476966	jbvjbName	"Haut de Cry/Derborence"^^xsd:string	<http://www.wsl.example/ProJB/Daten/RD/JB/>&baseURI=<http://www.wsl.example/ProJB/Daten/RD/JB/>
jb:JB476959	jbvjbName	"Haut de Cry/Derborence"^^xsd:string	<http://www.wsl.example/ProJB/Daten/RD/JB/>&baseURI=<http://www.wsl.example/ProJB/Daten/RD/JB/>

Abbildung 7.2: Die Teilgebiete des Jagdbanngebietes “Haut de Cry/Derborence” in der Benutzeroberfläche Sesame Workbench.

Explore (jb:JB476966)

Subject	Predicate	Object	Context
jb:JB476966	rdf:type	jbv:Jagdbanngebiete	<http://www.wsl.example/ProJB/Daten/RD/JB/>&baseURI=<http://www.wsl.example/ProJB/Daten/RD/JB/>
jb:JB476966	jbv:id	"JB476966"^^xsd:string	<http://www.wsl.example/ProJB/Daten/RD/JB/>&baseURI=<http://www.wsl.example/ProJB/Daten/RD/JB/>
jb:JB476966	jbv:jbSchutzstatus	"20"^^xsd:string	<http://www.wsl.example/ProJB/Daten/RD/JB/>&baseURI=<http://www.wsl.example/ProJB/Daten/RD/JB/>
jb:JB476966	ogc:intersects	<http://www.wsl.example/ProJB/Daten/RD/BR/BR91928>	<http://www.wsl.example/ProJB/Daten/RD/JB/>&baseURI=<http://www.wsl.example/ProJB/Daten/RD/JB/>
jb:JB476966	jbv:jbTeilgebiet	true	<http://www.wsl.example/ProJB/Daten/RD/JB/>&baseURI=<http://www.wsl.example/ProJB/Daten/RD/JB/>
jb:JB476966	jbv:jbName	"Haut de Cry/Derborene"^^xsd:string	<http://www.wsl.example/ProJB/Daten/RD/JB/>&baseURI=<http://www.wsl.example/ProJB/Daten/RD/JB/>
jb:JB476966	ogc:hasGeometry	jb:JBGeom476966	<http://www.wsl.example/ProJB/Daten/RD/JB/>&baseURI=<http://www.wsl.example/ProJB/Daten/RD/JB/>
rd:RD1	jbv:hatInstanz	jb:JB476966	<http://www.wsl.example/ProJB/Daten/RD/>&baseURI=<http://www.wsl.example/ProJB/Daten/RD/>

Abbildung 7.3: Das Teilgebiet "JB476966" in detaillierter Ansicht in der Benutzeroberfläche Sesame Workbench.

Die Abfrage des Use Case "Aggregation" ist im Quelltext 7.2 festgehalten. Sie baut auf dem ersten einfachen Use Case auf und besteht aus drei Tripelmustern. Es wird die Gesamtfläche aller partiell geschützten Gebiete (Schutzstatus "20") abgefragt. Hier gilt bei geteilten Jagdbanngebieten, dass nur die effektiv zutreffenden Teilflächen zusammengerechnet werden. Das erste Tripelmuster (Zeile 7) ermittelt die betroffenen Gebiete, während das zweite und dritte (Zeilen 8 und 9) die zugehörige Geometrie aufrufen, da die Flächeninformation ein Attribut der Klasse Geometrie ist (nur bei Polygonen).

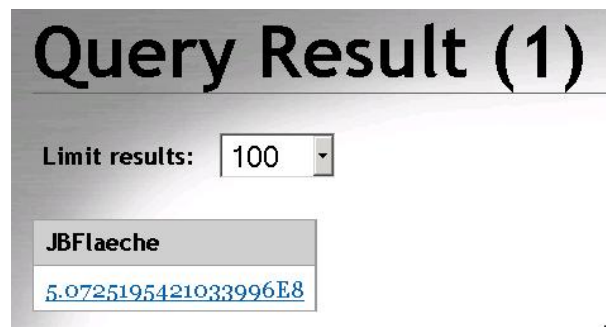
```

1 PREFIX jbVoc:<http://www.wsl.example/ProJB/Ontologie/ProJB_Vokabular#>
2 PREFIX ogc:<http://www.opengis.net/rdf#>
3 PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
4
5 SELECT (SUM(?Flaeche) AS ?JBFlaeche)
6 WHERE{
7 ?Teilgebiet jbVoc:jbSchutzstatus "20"^^xsd:string .
8 ?Teilgebiet ogc:hasGeometry ?Geometrie .
9 ?Geometrie jbVoc:flaeche ?Flaeche .
10 }

```

Quelltext 7.2: SPARQL-Abfrage für den einfachen Use Case "Aggregation".

Die Gesamtfläche der partiell geschützten Gebiete beträgt 507 km² (5.07·10⁸ m²). In der Abbildung 7.4 ist die Ausgabe der Sesame Workbench ersichtlich. Die Resultate der beiden einfachen Anwendungsfälle wurden mit Hilfe des GIS (ArcGIS) überprüft und es wurde das identische Ergebnis erzielt.



JBFlaeche
5.0725195421033996E8

Abbildung 7.4: Ergebnis der Abfrage aus Quelltext 7.2 in der Sesame Workbench.

7.2 Fazit

In diesem Kapitel wurden die SPARQL-Abfragen der beiden einfachen Use Cases, basierend auf dem Datenmodell von ARO, vorgestellt. Entscheidend für die erfolgreiche Lösung ist in erster Linie die Umformung der Originaldaten in das RDF-Format, da die Anwendungsfälle mit den bereits vorhandenen Daten gelöst werden können. Das heisst, die generierten Tripel müssen die Informationen aus den Ursprungsdaten korrekt wiedergeben. So müssen beispielsweise die entsprechenden Flächeninformationen den passenden Gebieten zugeordnet werden. Die beiden einfachen Use Cases stellen also nicht mehr als einfache Abfragen über den konvertierten RDF-Bestand dar. Die mit räumlichen Daten verbundenen Herausforderungen werden hier nicht behandelt, da weder die Geometrien noch topologische Beziehungen für die Lösung benötigt werden.

Komplexer Use Case “Waldflächen und Höhenstufen”

Der komplexe Use Case “Waldflächen und Höhenstufen” lautet: *Wie verteilen sich die Waldflächen einzelner oder sämtlicher Jagdbanngebiete auf die verschiedenen ökologischen Höhenstufen, unter Berücksichtigung der regionalen Variabilität dieser Höhenstufen innerhalb der Schweiz?*

Der Anwendungsfall beinhaltet mehrere Schritte zur Lösung. So müssen erst die Waldflächen ermittelt werden, welche räumlich innerhalb der gefragten Jagdbanngebiete liegen. Danach müssen diese Waldflächen den Höhenstufen zugeteilt und für die jeweilige Stufe aufsummiert werden. Die Unterteilung der Höhenstufen basiert auf Angaben aus der Literatur. Die regionale Variabilität der Stufen wird anhand der biogeographischen Regionen unterschieden. Das heisst, die Höhenunterteilung wird jeweils innerhalb dieser Regionen festgelegt. Zusätzlich ist mit dem Use Case auch eine Massstabsentscheidung verbunden. Sie ergibt sich aufgrund der vorhandenen Ausgangsdaten. Die Geodaten zu den Primärflächen und dem Höhenmodell liegen jeweils doppelt, aber mit unterschiedlichen Massstäben (damit auch Genauigkeit) vor.

In diesem Kapitel wird das Vorgehen zur Lösung des Use Case erläutert. Dazu gehören neben der konkreten Umsetzung auch noch einige konzeptionelle Überlegungen sowie zusätzliche Informationen, die aufbereitet werden müssen. Zudem wird an einem konkreten Beispiel ein mögliches Ergebnis des Anwendungsfalles “Waldflächen und Höhenstufen” gezeigt. Eine kurze Diskussion am Ende schliesst das Kapitel ab.

8.1 Use Case spezifische Grundlagen

Der Use Case “Waldflächen und Höhenstufen” kann nicht alleine mit den vorhandenen Datensätzen gelöst werden. Es müssen zusätzliche Informationen hinzugezogen und teilweise in das RDF-Format umgeformt werden. Zudem müssen für die Datensätze in unterschiedlichen Massstäben (VEC25/VEC200 bzw. DHM25/DHM RIMINI) Kriterien festgelegt werden, wann welche eingesetzt werden. Namentlich umfassen die zusätzlichen Grundlagen die Höhenstufen und eine Massstabsentscheidung.

8.1.1 Höhenstufen

Das Aussehen und die Zusammensetzung einer Vegetation¹ ist abhängig von der geographischen Lage und den Lebensbedingungen (Klima, Boden usw.). So bewirken ähnliche Lebensbedingungen eine ähnliche Vegetation. Diese ändert sich in der Schweiz von Norden nach Süden, sowie auch mit der Höhe, so dass in Gebirgen ein vertikale Veränderung der Vegetation auszumachen ist (Landolt, 2003).

Es lassen sich sogenannte Höhenstufen unterscheiden, die das Gebirge ringförmig umschliessen und sich

¹ Die Vegetation ist die Gesamtheit der in einem Gebiet vorkommenden Pflanzengemeinschaften und -individuen (Pflanzen-decke)(Landolt, 2003).

Höhenstufe	Region	Höhenbereich (nord- /südexponierte Lage)	Durchschnittlicher Höhenbereich
collin	Jura: Mittelland: Nordalpen: Zentral- und Südalpen	-500/600m -600m -600/700m -800/900m	-600m
montan	Jura: Mittelland: Nordalpen: Zentralalpen Südalpen	500/600 - 1200m über 600m 600/700 - 1200/1500m 800/900 - 1200m 800/900 - 1600/1700m	600 - 1200m
subalpin	Nordalpen: Zentralalpen Südalpen	1200/1500 - 1800/2100m 1200 - 1700/2100m 1600/1700 - 1900/2200m	1200 - 2000m
alpin	Nordalpen: Zentralalpen Südalpen	1800/2100 - 2400/2600m 1700/2100 - 2700/3000m 1900/2200 - 2500/2800m	über 2000m

Tabelle 8.1: Höhenstufenunterteilung in m.ü.M. (verändert nach: <http://www.agroscope.admin.ch/ziel-leitarten/00631/02626/index.html?lang=de>, 29.12.2011).

anhand bestimmter Vegetationstypen, Wuchsformen und Artenzusammensetzungen ausscheiden lassen. Diese Höhenstufen weisen keine scharfen Abgrenzungen auf, sondern laufen in breiten Übergangszonen ineinander. Die Vegetation in Gebirgen wird vor allem durch die Temperatur und Kontinentalität beeinflusst. Kontinentalität bezeichnet eine grössere Temperaturamplitude, mehr Frost, wenig Niederschläge, eine hohe Ein- und Ausstrahlung sowie eine geringe Luftfeuchtigkeit. Die Kontinentalität nimmt in der Regel von äusseren Gebieten eines Gebirges zu den inneren zu. Zudem nimmt sie auch von mittleren Höhen in die unteren und höheren Lagen zu. Mittlere Höhen haben oft mehr Niederschläge und Bewölkung als untere Lagen, was ein ozeanischeres Klima erzeugt (Landolt, 2003).

Je nach Autor und den von ihnen definierten Kriterien werden die Höhenstufen unterschiedlich eingeteilt. Ein Grund dafür ist die grosse Variation innerhalb der Höhenstufen. Dennoch lassen sich durchgehende Charakterzüge anhand der Artengefüge der Baumschicht identifizieren (Ellenberg, 1996).

Im Rahmen dieser Arbeit wird eine angepasste Unterteilung der Forschungsanstalt Agroscope verwendet² (vgl. Tabelle 8.1). Die eigene Klassifikation weicht insofern vom Original der Agroscope ab, dass die Höhengrenzen schlüssig aneinander angepasst wurden, damit keine Lücken oder Überlappungen entstehen. In dieser Einteilung existieren die colline, montane, subalpine und alpine Stufe. Zusätzlich unterscheiden sich die Höhenbereiche der einzelnen Stufen, je nach biogeographischer Lage und Exposition. Jedoch sind im Original die subalpine und alpine Stufe für die Regionen Jura und Mittelland nicht definiert, da gemäss Agroscope diese Gebiete gar nicht oder nur vereinzelt in die vorgegebenen Höhenbereiche fallen.

Die erste Zahl in der Tabelle 8.1 steht jeweils für nördlich exponierte Gebiete, die zweite für die südlich exponierten. Die Stufen sind diskret unterteilt, obwohl sie in der Realität kontinuierliche Übergänge aufweisen.

Eine kurze Charakterisierung der Höhenstufen:

- Die natürliche Vegetation der collinen Stufe besteht aus sommergrünen Laubwäldern. Die vorherrschenden Laubbaumarten unterscheiden sich je nach biogeographischer Region (Landolt, 2003).

² <http://www.agroscope.admin.ch/ziel-leitarten/00631/02626/index.html?lang=de>, 29.12.2011

- Die obere Grenze der montanen Stufe wird durch das Vorkommen der Buche begrenzt. Die montane Stufe zeichnet sich durch seltenen Frost in der Vegetationszeit (Abfliessen der Kaltluft in der Nacht), moderate Temperaturschwankungen, hohe Niederschläge und Luftfeuchtigkeit sowie einer grösseren Bewölkung aus (Landolt, 2003; Ellenberg, 1996).
- Die obere Grenze der subalpinen Stufe wird durch die Rottanne bestimmt, dementsprechend herrschen Nadelwälder vor. Das Klima definiert sich durch grössere Trockenheit, Temperaturschwankungen und Einstrahlung als die montane Stufe (Landolt, 2003). Mit zunehmender Höhe werden die Lebensbedingungen für Bäume ungünstiger, so dass die Waldgrenze ebenfalls auf dieser Stufe liegt (Ellenberg, 1996).
- Auf der alpinen Stufe finden sich vor allem kurze Rasen. Höhere Wuchsformen (z. B. Sträucher, Hochstauden) sind nur an günstigen Stellen zu finden. Sie wird durch das Vorkommen zusammenhängender Rasenflecken begrenzt (Landolt, 2003).

Neben den süd- und nordexponierten Gebieten existieren auch ebene Flächen, die keiner der beiden Expositionen zugeordnet werden können. In diesen Fällen wird auf die Durchschnittswerte (vierte Tabellenspalte) zurückgegriffen. Da die Durchschnittswerte unabhängig von der Exposition sind, vertreten diese auch ebene Gebiete.

Eine Ungenauigkeit besteht in der fehlenden subalpinen Höhenbereichsdefinition für den Jura. Gemäss der Aufstellung von Agroscope werden diese als “nur in Fragmenten vorhanden” angegeben. Jedoch hat eine Überprüfung ergeben, dass ca. 10% (457 km²) der Gesamtfläche dieser biogeographischen Region über der definierten montanen Höhenstufe liegen. Dies müsste allenfalls bei späteren Weiterentwicklungen zusätzlich berücksichtigt werden, so dass auch für den Jura eine subalpine Stufe definiert wird.

8.1.2 Massstab

Da teilweise Daten in unterschiedlichen Massstäben vorhanden sind, sollen diese sinnvoll eingesetzt werden können. Informationen zu den Primärflächen und den Höhenstufen bilden einen zentralen Bestandteil des komplexen Use Case “Waldflächen und Höhenstufen” und liegen als jeweils zwei unterschiedliche Datensätze vor (VEC25 bzw. VEC200 und DHM25 bzw. DHM RIMINI). Der Entscheid welche Daten wann verwendet werden, soll anhand der benötigten Zeit zur Prozessierung des Anwendungsfalles getroffen werden. Dabei wurde angenommen, dass der Zeitaufwand der räumlichen Operationen mit der Grösse der zu prozessierenden Fläche zusammenhängt.

Um zu ermitteln ab wie vielen Jagdbanngebieten bzw. ab welcher Flächengrösse ein Wechsel zwischen den Datensätzen Sinn macht, wurden Zeitmessungen durchgeführt. Dazu wurden die Teilgebiete der Jagdbanngebiete nach ihrer Fläche geordnet und dann beginnend mit dem kleinsten Gebiet die benötigte Zeit der zur Beantwortung der Abfrage nötigen räumlichen Operationen (z. B. *intersect*) gemessen. Dieses Gebiet wurde um das nächst grössere Gebiet ergänzt und wiederum die Prozessierungszeit gemessen. Das wurde solange fortgesetzt, bis alle Gebiete in der Messung berücksichtigt wurden. Somit nahm die zu prozessierende Gesamtfläche stetig, aber in unterschiedlich grossen Schritten zu. Das gleiche Vorgehen wurde zusätzlich in absteigender Reihenfolge, d. h. beginnend mit den grössten Gebieten wiederholt. Die Messungen wurden jeweils für das DHM RIMINI zusammen mit VEC200 und DHM25 mit VEC25 durchgeführt. Aus den Abbildungen 8.1 und 8.2 für das DHM25 kann entnommen werden, dass es jeweils keine signifikante Veränderung in der benötigten Zeit gegeben hat. Jeweils die erste Messung mit nur einem Gebiet hat am längsten gedauert. Danach ist die Zeit relativ konstant geblieben. Das gleiche Bild hat



Abbildung 8.1: Zeitmessung von den kleinsten zu den grössten Teilgebieten mit DHM25 und Vec25.

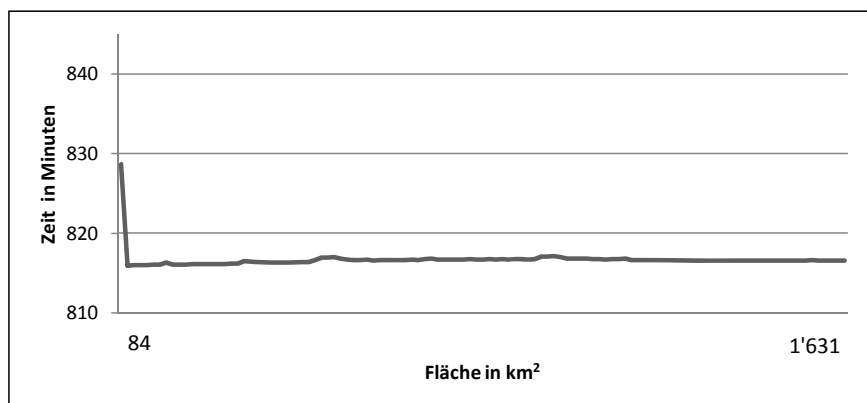


Abbildung 8.2: Zeitmessung von den grössten zu den kleinsten Teilgebieten mit DHM25 und Vec25.

sich bei Verwendung der DHM RIMINI Daten ergeben, wobei die benötigte Prozessierungszeit generell um ein vielfaches tiefer lag.

Da somit keine Grenze festgestellt werden konnte, ab welcher Flächengrösse sich die benötigte Prozessierungszeit erheblich verändert, wurde die Grenze “willkürlich” bei 100 km² gesetzt.³ Das heisst, ist die Fläche der abgefragten Gebiete kleiner, werden die grossmassstäbigen Daten (VEC25 und DHM25) verwendet, ist die Fläche grösser, ist es umgekehrt. Der Grenzwert kann einfach angepasst werden, so dass allfällige Änderungen der Grenze schnell umgesetzt werden können.

8.2 Datenaufbereitung der Höhenstufen

Zu den bereits im Triple Store gespeicherten Grundlagendaten, müssen noch die Informationen zu den Höhenstufen ergänzt werden. Erst damit ist die für den Use Case benötigte Datengrundlage vorhanden. Die Höhenstufen werden gemäss der Klasse *Vegetationshoehenstufen* im Datenmodell generiert und als Tripel hinzugefügt. Jedes Tripel beschreibt dabei eine Höhenstufe mit einer bestimmten Exposition und regionalen Zugehörigkeit.

³ Zum Vergleich: Der Median der Flächen der Jagdbannteilgebiete beträgt 9 km².

Das Beispiel im Quelltext 8.1 repräsentiert die subalpine Stufe für nordexponierte Lagen in der biogeographischen Region der Südalpen.

```

1 @Prefix  vh:<http://www.wsl.example/ProJB/Konzepte/VH/>
2 @Prefix  jbVoc:<http://www.wsl.example/ProJB/Ontologie/ProJB_Vokabular#>
3 @Prefix  ogc:<http://www.opengis.net/rdf#>
4 @Prefix  xsd:<http://www.w3.org/2001/XMLSchema#>
5
6 vh:VH27  jbVoc:id           "VH27"^^xsd:string .
7 vh:VH27  jbVoc:name         "subalpin"^^xsd:string .
8 vh:VH27  jbVoc:hoehenStufe  "1600, 1900"^^xsd:string .
9 vh:VH27  jbVoc:exposition   "nordexponiert"^^xsd:string .
10 vh:VH27  ogc:within         <http://www.wsl.example/ProJB/Daten/RD/BR/BR91930> .

```

Quelltext 8.1: Tripel der subalpinen Stufe (nordexponiert) in den Südalpen.

Wie bei allen Klassen wird eine eindeutige ID vergeben. Die Attribute `jbVoc:name`, `jbVoc:hoehenstufe` und `jbVoc:exposition` enthalten die relevanten Informationen zur Höhenstufe, wobei `jbVoc:hoehenstufe` den Höhenbereich, also von 1600 bis 1900 m.ü.M, angibt. Da die Höhenstufen je nach biogeographischer Region unterschiedlich ausfallen, werden sie topologisch durch `ogc:within` verortet. Damit im Verlaufe des Lösungsprozesses die Exposition der Gebiete ermittelt werden kann, wurde ein entsprechender Datensatz vorbereitet. Jeweils für das DHM25 und DHM RIMINI wurde die Exposition mit dem räumlichen Operator “Aspect”⁴ berechnet und vektorisiert (in ArcGIS). Das heisst, jeder Rasterzelle wurde neu ein Wert zwischen 0 und 359.9 zugewiesen, der für die Ausrichtung in Grad steht. Mittels einer Reklassifikation (*reclassify*) wird nur noch zwischen nord- und südexponierten, sowie ebenen Zellen unterschieden. Ebene Zellen werden automatisch bereits durch die Operation ausgewiesen. Zu den nordexponierten gehören alle Zellen mit den Werten von 0 bis 89.9 Grad und von 270 bis 359.9 Grad. Zu den südexponierten Stellen gehören alle Zellen mit den Werten 90 - 269.9 Grad. Danach wurde das reklassifizierte Raster vektorisiert.

8.3 Datenprozessierung

Die Abbildung 8.3 zeigt die Abfolge der Prozesse zur Lösung des Anwendungsfalles. Nach Bestimmung der zu bearbeitenden Jagdbanngebiete werden mit SPARQL-Abfragen auf den Triple Store die Pfade zu den benötigten Geodatensätzen ermittelt. Die Pfade der Jagdbanngebiete und der Biogeographischen Regionen können direkt abgefragt werden. Zusätzlich wird auch die Identifikationsnummer `jbVoc:id` der betroffenen Jagdbanngebiete für die späteren Geoprozessierungen in Erfahrung gebracht. Darauf folgt die Massstabsentscheidung, also die Festlegung aus welchen Datensätzen die Informationen zu den Primärflächen und Höhen genommen werden. Damit werden vorerst keine weiteren Informationen aus dem Triple Store benötigt und es folgt die ausgelagerte Prozessierung mit Hilfe unterschiedlicher räumlicher Operationen

Die Pfade zu den Geodatensätzen sind bekannt, womit sie direkt mit den räumlichen Operationen verwendet werden können. Zuerst werden die Waldflächen aus dem Primärflächen Datensatz mit einer Verschneidungs-Operation (*intersect*) auf die betroffenen Objekte aus dem Jagdbanngebietsdatensatz begrenzt. Das Ergebnis ist ein Datensatz, der nur noch die als “Wald” gekennzeichneten Objekte innerhalb der betroffenen Jagdbanngebietsflächen enthält. Dieser Datensatz wird wiederum mit der bereits vorprozessierten Exposition überschritten. Dadurch werden die Waldobjekte allenfalls weiter zerteilt, da

⁴ <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009z000000tr000000.htm>

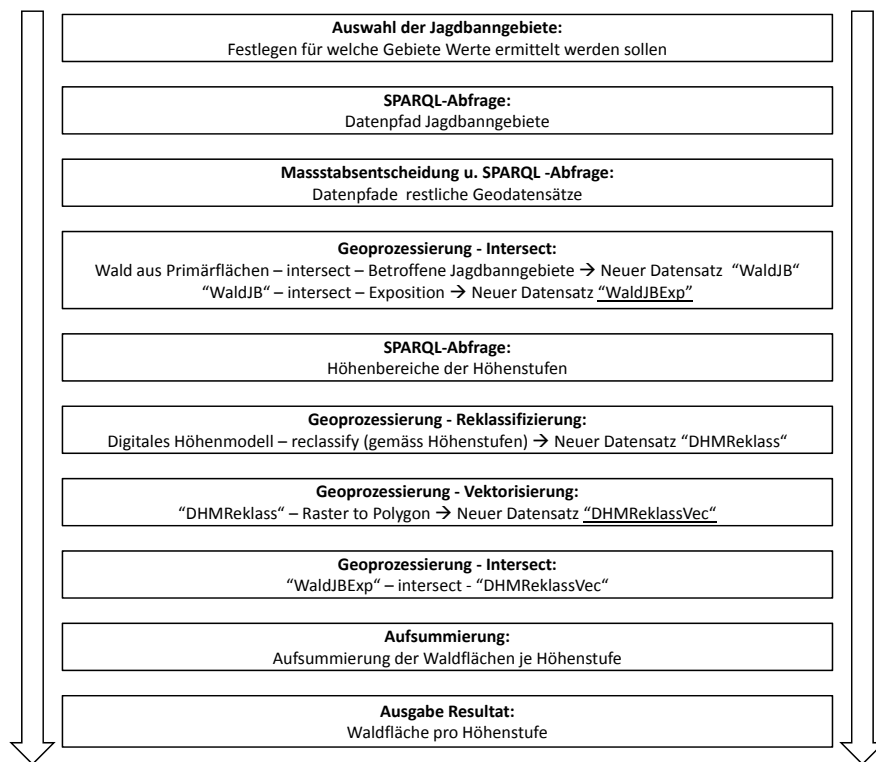


Abbildung 8.3: Ablaufdiagramm des komplexen Use Case “Waldfläche und Höhenstufen”.

sie unterschiedliche Expositionen aufweisen können. Die somit entstandenen Objekte sind Waldflächen mit einer bestimmten Exposition. Der gesamte Datensatz wird als “WaldJBExp” bezeichnet. Die Abbildung 8.4 zeigt im linken Kartenausschnitt das Ergebnis der Verschneidung des Waldes mit den Jagdbannggebieten. Der Wald ist also nur noch innerhalb der Jagdbannggebiete vorhanden. Der rechte Ausschnitt zeigt den neu generierten Datensatz “WaldJBExp”. Die Waldflächen teilen sich nun in südexponierte (gelb) und nordexponierte (blau) Gebiete auf. Als nächstes muss das Höhenraster prozessiert werden, um die benötigte Einteilung in die unterschiedlichen Höhenstufen zu erreichen. Dazu muss erst wieder auf den Triple Store zurückgegriffen werden, da die Informationen zu den Höhenstufen dort gespeichert sind. Durch SPARQL-Abfragen werden die Unterteilungen der Höhenstufen ermittelt und in Listen abgelegt. Danach wird wieder mit den (ausgelagerten) räumlichen Operationen weitergearbeitet. Das Höhenraster wird, aufgeteilt nach den biogeographischen Regionen, gemäß den Höhenstufen reklassifiziert (*reclassify*). Das heisst, die konkreten Höhenangaben gehen verloren und werden durch die Höhenstufen collin, montan, subalpin und alpin ersetzt. Jede Rasterzelle kann also einer Höhenstufe zugeordnet werden. Danach wird das reklassifizierte Raster, wie schon bei VKE, vektorisiert. Dieser neue Vektordatensatz wird als “DHMRklassVec” bezeichnet. Die Abbildung 8.5 zeigt einen direkten Vergleich zwischen dem digitalen Höhenmodell DHM RIMINI (linker Kartenausschnitt) und dessen Reklassifizierung (rechter Kartenausschnitt). Als letzte Geooperation werden die Datensätze “WaldJBExp” und “DHMRklassVec” miteinander überschritten (*intersect*). Dies wird jeweils separat für die südexponierten und nordexponierten Objekte aus dem Datensatz “WaldJBExp” durchgeführt. Daraus resultieren Objekte, die den Wald in den Jagdbannggebieten repräsentieren und über Informationen zu deren Exposition und zugehörigen Höhenstufe verfügen. Schlussendlich werden die einzelnen Objekte gemäß den Höhenstufen aufsummiert, wodurch man die jeweilige Waldfläche auf den unterschiedlichen Höhenstufen erhält.

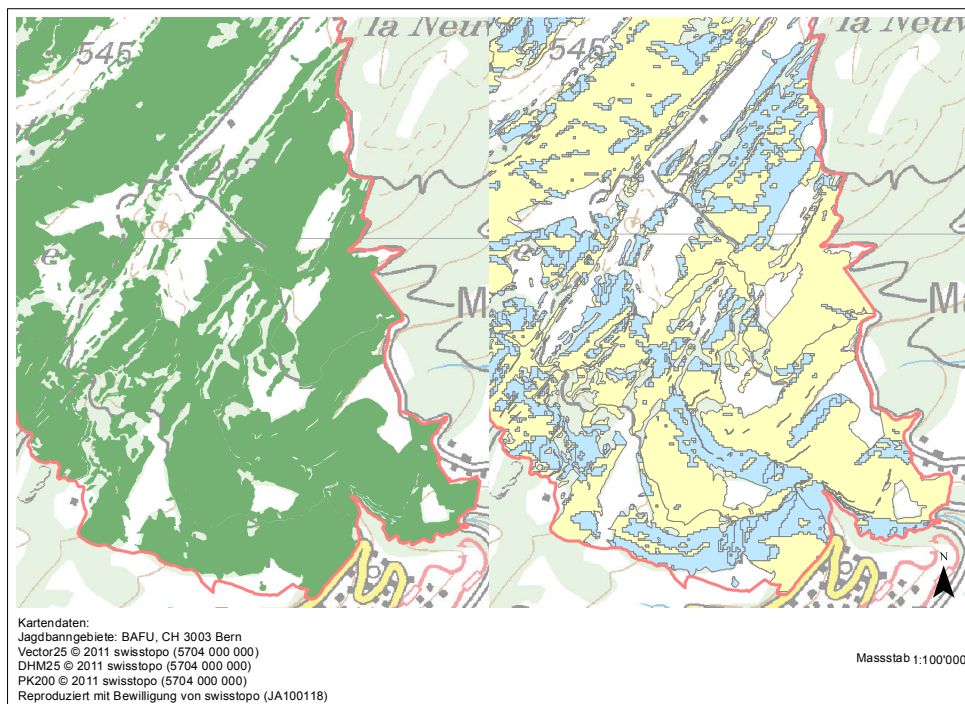


Abbildung 8.4: Ergebnisse der zwei Intersect-Operationen.

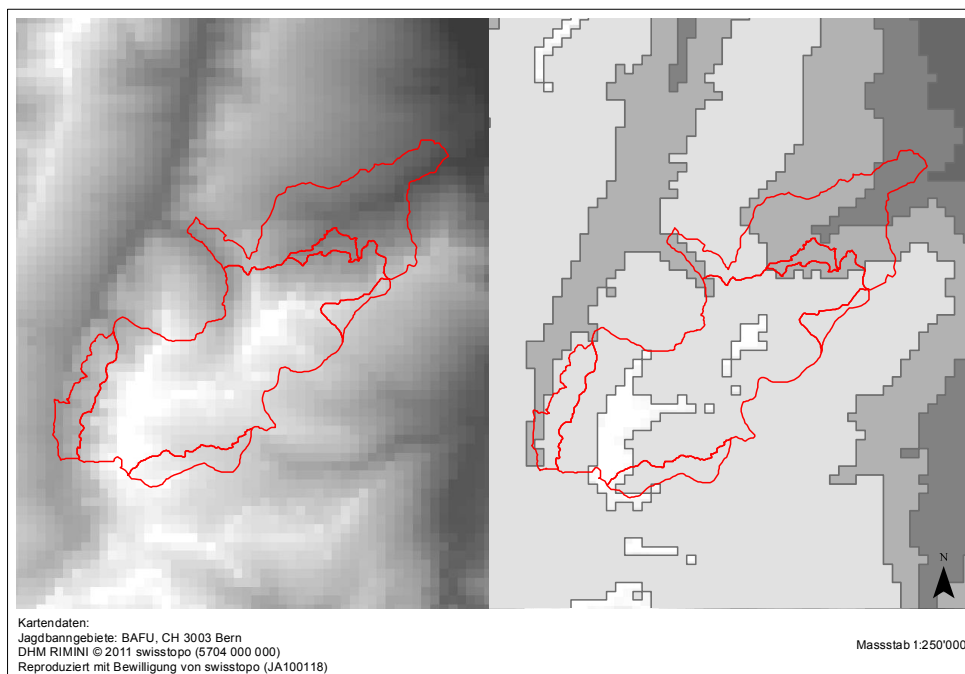


Abbildung 8.5: Ursprüngliches digitales Höhenraster und dessen Reklassifizierung gemäss den Höhenstufen (Jagdbanngebiet "Beverin").

8.4 Resultate

Der folgende Abschnitt stellt die Resultate der konkreten Abfrage nach den zufällig ausgewählten Jagdbanngebieten “Dent de Lys” und “Bernina-Albris” vor. Dieses Beispiel kann natürlich beliebig variiert werden, indem jegliche Jagdbanngebiete oder auch nur Teilgebiete abgefragt werden.

Die Jagdbanngebiete “Dent de Lys” und “Bernina-Albris” haben zusammen eine Gesamtfläche von ca. 93 km², wovon die Waldflächen ca. 15 km² einnehmen. Die Waldflächen pro Höhenstufe (beider Gebiete zusammen) sind in Tabelle 8.2 aufgelistet. Auf der collinen Stufe befinden sich gar keine Waldflächen, wohingegen auf der alpinen Stufe ein vergleichsweise grosser Waldanteil auszumachen ist.

Höhenstufe	Fläche in km ²
collin	0
montan	1.688
subalpin	3.742
alpin	9.605

Tabelle 8.2: Ergebnis - Waldfläche je Höhenstufe.

8.5 Diskussion

In diesem Kapitel wurde das grundsätzliche Vorgehen zur Lösung des komplexen Use Case “Waldflächen und Höhenstufen” erläutert, sowie die Resultate einer konkreten Abfrage vorgestellt. Die Lösung des Anwendungsfalles stützt sich stark auf die ans GIS ausgegliederten räumlichen Operationen, da eine Reihe von Geoprozessierungen nötig sind, um an die zusätzlichen Informationen zu kommen. Die Abfragen auf dem Triple Store liefern in diesem Fall vor allem Angaben für die Durchführung der räumlichen Operationen, wie die Pfade zu den Geodatensätzen oder die Höheninformationen.

Der komplexe Use Case “Waldflächen und Höhenstufen” erlaubt einen breiten Spielraum zur Umsetzung, weshalb immer bedacht werden sollte, dass die Resultate nur im Rahmen der angegebenen Parameter gelten. So könnte beispielsweise auf eine andere Einteilung der Höhenstufen oder der Regionen zurückgegriffen werden, woraus andere Werte resultieren würden.

Das Ergebnis der Abfrage besteht aus Zahlenwerten, die sich auf eines bzw. mehrere konkrete Jagdbanngebiete sowie Waldflächen beziehen und dadurch einen räumlichen Bezug haben. Somit könnte beispielsweise im Datenmodell den entsprechenden Objekten der Klasse Jagdbanngebiete eines oder mehrere Attribute hinzugefügt werden, die diese Verteilung der Waldflächen repräsentieren. Besser wäre die direkte Einbindung bei den entsprechenden Waldflächen, also den Objekten aus dem Datensatz Primärflächen. Von den Primärflächen werden jedoch keine konkreten Objekte abgelegt. Ein wichtiger Grundgedanke von Linked Data, Daten miteinander zu Verknüpfen, kommt bei diesem Anwendungsfall weniger zum Zug. Dies liegt vor allem am Ergebnis des Anwendungsfalles, da es sich um numerische Werte handelt, die nur schwierig mit anderen Daten verknüpft werden können. Auf eine Speicherung der Resultate im Triple Store wird verzichtet, da diese als Produkt einer komplexen Abfrage gesehen werden und weniger als neu generierte Ressourcen, die in den Triple Store integriert und verknüpft werden sollten.

Der Anwendungsfall zeigt die hohe Flexibilität des Ansatzes ARO auf und die breite Einsatzmöglichkeit. Denn es können Operationen durchgeführt werden, die mit SPARQL bzw. GeoSPARQL nicht möglich bzw. vorgesehen sind. Dies trifft besonders auch auf die Rasterprozessierung zu (z. B. Reklassifizierung der Raster).

Wie eventuell bereits aus den Abbildungen der Zeitmessung (Abbildungen 8.1 und 8.2) aufgefallen ist, benötigt die gesamte Prozessierung mit den grossmassstäbigen Daten (DHM25 und VEC25) unter den gegebenen Voraussetzungen (wie z. B. Hardware) sehr viel Zeit (über 800 Minuten). Das ist auf die räumlichen Operationen zurückzuführen und nicht auf die Linked Data Technologien (z. B. Triple Store). Das heisst, eine Bearbeitung des Anwendungsfalles mit einem GIS würde ähnlich lange dauern. Ein Grossteil der Zeit wird für die Verschneidungs-Operationen (*intersect*) der Waldflächen mit den Primärflächen bzw. mit der Exposition benötigt. Dies könnte allenfalls durch eine effizientere Implementierung verbessert werden. Eine weitere Möglichkeit zur Verkürzung der Bearbeitungszeit ist die Bereitstellung von bereits vorprozessierten Datensätzen. Beispielsweise könnten die Verschneidungs-Operationen ausgelassen und stattdessen durch einmalig für alle Jagdbanngebiete generierte Datensätze, die die Ergebnisse der Verschneidung enthalten, ersetzt werden. Jedoch werfen sich dadurch (wie schon bei VKE) Fragen auf, inwieweit Daten im Voraus aufbereitet werden sollen. Denn bei einer weitreichenden Vorprozessierung kommt man wieder in den Bereich des Ansatzes VKE. Generell sollten nur Daten vorprozessiert werden, die häufig in gleicher Form verwendet werden. Beispielsweise würde die Vorprozessierung der reklassifizierten Raster Sinn machen, da die Höhenstufen relativ beständig sind und sich sicherlich nicht innerhalb kurzer Zeitabstände ändern. Gleiches gilt für die Aktualität der digitalen Höhenmodelle.

Komplexer Use Case “Betroffene Gemeinden”

Der komplexe Use Case “Betroffene Gemeinden” lautet: *Auf welchen Gemeinden liegt ein Jagdbanngebiet? Wie heissen die Gemeinden gemäss GeoNames und das zugehörige Jagdbanngebiet?*

Der Anwendungsfall basiert weniger auf komplexen räumlichen Operationen, wie dies bei “Waldflächen und Höhenstufen” der Fall ist, sondern konzentriert sich auf die Einbindung einer externen Datenquelle. Die Grundidee von Linked Data, ein Web der Daten zu schaffen (durch Verknüpfung von Ressourcen), steht im Zentrum. Da es sich bei den Daten aus GeoNames um nutzergenerierte Inhalte (*user generated content*) handelt, muss die Vertrauenswürdigkeit, Qualität und Relevanz der Daten mit Vorsicht genossen werden. Allgemein werden diese Qualitätskriterien in der zukünftigen Entwicklung von Linked Data von grösserer Bedeutung sein (Bizer u. a., 2009a).

Durch die Bereitstellung von Verknüpfungen zu GeoNames werden weitere Zugänge in das Web der Daten geschaffen, da von GeoNames aus wiederum Verknüpfungen zu anderen Datenquellen existieren.

In diesem Kapitel wird ein möglicher Lösungsweg des Anwendungsfalles vorgestellt. Zuerst folgen einige konzeptionelle Grundlagen zur Einbindung der GeoNames-Daten. Danach wird die Datenprozessierung mit den grundsätzlichen Arbeitsschritten erläutert. Zum Schluss werden die Resultate einer konkreten Abfrage vorgestellt, gefolgt von einer Diskussion der Implementierung.

9.1 Einbindung von GeoNames

Der Einbezug von Daten aus fremden Quellen beinhaltet unter anderem die sinnvolle Integration in die eigene Datenstruktur. Dazu können beispielsweise Übersetzungen zwischen den jeweiligen Vokabularen bereitgestellt werden. Eine einfache Möglichkeit auf Ebene der Instanzen bietet `owl:sameAs`¹, womit Ressourcen, die das Gleiche repräsentieren, verknüpft werden können (Bizer u. a., 2009a).

Die Einbindung von GeoNames in das eigene Datenmodell soll insofern realisiert werden, dass die entsprechenden GeoNames-Objekte, welche eine Gemeinde repräsentieren, gefunden und Verknüpfungen von den eigenen Ressourcen zu GeoNames bereitgestellt werden. Schlussendlich sollen die GeoNames-Objekte selektiert sein, die Gemeinden repräsentieren, auf welchen ein Jagdbanngebiet vollständig oder teilweise liegt. Bevor die Daten aus GeoNames in den Triple Store integriert werden können, müssen einige Probleme gelöst werden:

Eine Gemeinde ist ein flächiges Objekt, wird aber in GeoNames als Punkt repräsentiert. Dadurch ist die genaue räumliche Ausdehnung der Gemeinde unbekannt. Es kann nicht direkt ermittelt werden, welche dieser GeoNames-Punkte eine Gemeinde repräsentieren, die innerhalb eines Jagdbanngebietes sind oder mit dem Jagdbanngebiet eine gemeinsame Schnittmenge teilen. So können beispielsweise Punkte ausserhalb der Jagdbanngebietsgrenzen eine betroffene Gemeinde sein, aber ebenso gut eine Gemein-

¹ <http://www.w3.org/TR/owl-ref/#sameAs-def>

de, die von keinem Jagdbanngebiet tangiert wird. Deshalb wird zusätzlich ein Polygon-Datensatz mit den Gemeinden der Schweiz hinzugezogen. Dadurch können erst die betroffenen Gemeinden aus dem Gemeindedatensatz ausgewählt werden und anhand deren (flächigen) Geometrien die potentiell passenden GeoNames-Objekte gefunden werden. Der zusätzliche Geodatensatz der digitalen Gemeindegrenzen (swissBOUNDARIES3D²) stammt von der swisstopo. Um eine klare Unterscheidung von den Geonames-Daten zu gewährleisten, wird der swisstopo Datensatz als “Gemeindegrenzen” bezeichnet.

Wie bereits erwähnt beinhaltet GeoNames nutzergenerierte Daten, was zu gewissen Unsicherheiten führt, die beachtet werden müssen. Die Daten können unvollständig sein, d. h. es könnte zu lückenhaften Resultaten kommen, indem nicht alle betroffenen Gemeinden ausgegeben werden. Zudem können falsche Einträge vorhanden sein, wie beispielsweise eine falsche räumliche Referenzierung oder falsch geschriebene Namen.

Ein weiteres Problem ist die Beschränkung des GeoNames Webservice³ auf die Rückgabe von maximal 1000 Einträgen pro Abfrage, über welchen die aktuellsten Daten von GeoNames bezogen werden. Somit können die Punktdaten der Schweiz nicht auf einmal abgefragt werden. Deshalb werden die Objekte pro Kanton bzw. Halbkanton abgefragt. Jedoch besteht hier die Gefahr, dass zum einen Einträge ohne Angabe des Kantons nicht gefunden und falls ein Kanton mehr als 1000 Objekte hat, diese nicht vollständig erfasst werden. Zusätzlich zur der nach Kantonen aufgeteilten Abfrage, werden gezielt nur Einträge gesucht, die als “third-order administrative division” codiert sind. Das heisst, es werden nur Orte zurückgegeben, die der dritten administrativen Unterteilung, also den Gemeinden, entsprechen.⁴ Durch diese zusätzliche Einschränkung ist die Wahrscheinlichkeit relativ klein, dass für einen Kanton mehr als 1000 Einträge existieren. Zudem werden dadurch die effektiven GeoNames-Objekte zurückgegeben, die eine Gemeinde repräsentieren und nicht beispielsweise einen gleichnamigen Tierpark.

9.2 Datenprozessierung

Die Abbildung 9.1 zeigt die Abfolge der Prozesse zur Lösung des Anwendungsfalles auf.

Zuerst werden die aktuellen Daten von GeoNames über den “Geonames Search Webservice”⁵ abgefragt (vgl. Abschnitt 9.1). Die zurückgegebenen Daten werden in einer neu erstellten Objektklasse (*feature class*) in einer räumlichen Datenbank gespeichert. Nachdem die Jagdbanngebiete, für welche die Auswertung durchgeführt werden soll, ausgewählt sind, werden mittels SPARQL-Abfragen die Pfade zu den Geodatensätzen ermittelt. Dazu zählen die Jagdbanngebiete, die Gemeindegrenzen und die über den Webservice abgefragten GeoNames-Daten. Danach folgt eine kurze Sequenz von Geoprozessierungen mit der räumlichen Operation “Spatial Join”. In einem ersten Schritt werden die Gemeinden aus dem Datensatz Gemeindegrenzen ermittelt, auf welchen ein gefragtes Jagdbanngebiet liegt. Damit hat man die räumliche Ausdehnung dieser Gemeinden und kann wiederum mit einem “Spatial Join” die entsprechenden GeoNames-Objekte auswählen, d. h. jene GeoNames-Punkte, die innerhalb einer dieser betroffenen Gemeinden (aus Gemeindegrenzen) liegen.

Die so selektierten GeoNames-Objekte müssen aber nicht eine betroffene Gemeinde repräsentieren, da lediglich eine räumliche Selektion durchgeführt, aber die Bedeutung dieser Objekte nicht überprüft wird. Deshalb müssen die Daten zusätzlich gefiltert werden:

² <http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/landscape/swissBOUNDARIES3D.html>, 11.02.2012

³ <http://www.geonames.org/export/geonames-search.html>

⁴ <http://www.geonames.org/export/codes.html>, 11.01.2012

⁵ <http://www.geonames.org/export/geonames-search.html>

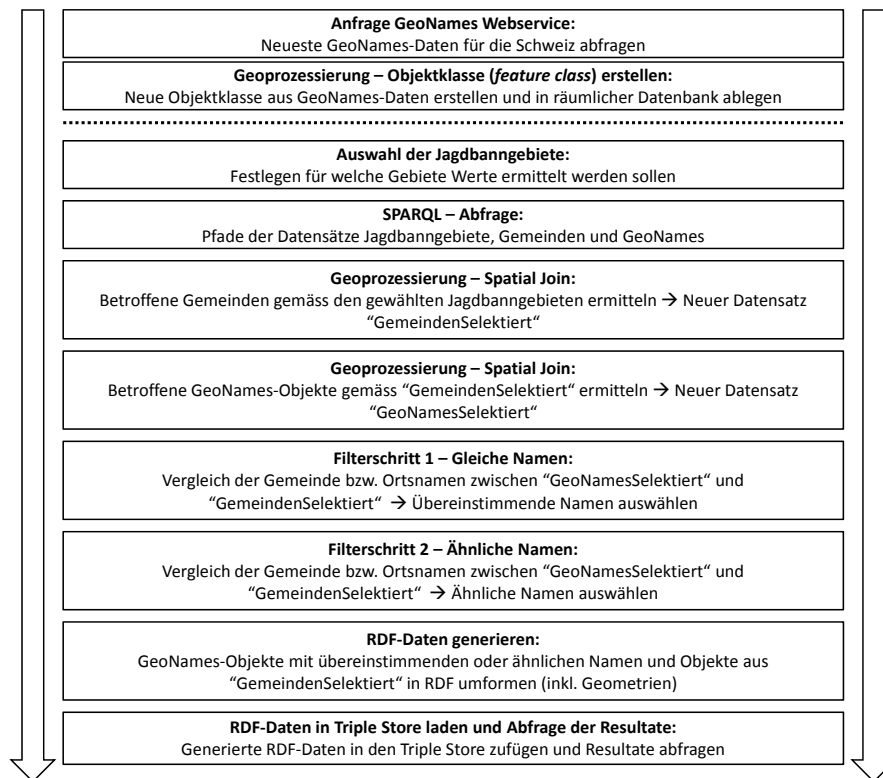


Abbildung 9.1: Ablaufdiagramm des komplexen Use Case “Betroffene Gemeinden”.

Der erste Filterschritt führt einen Abgleich der Namen aus GeoNames und den Gemeindennamen aus Gemeindegrenzen durch. Stimmen diese überein, wird angenommen, dass das entsprechende GeoNames-Objekt diese Gemeinde repräsentiert. Der zweite Filterschritt prüft, ob der Name des einen Objektes im Namen des anderen Objektes vorkommt, wobei die häufig verwendete Abkürzung “St.” für “Sankt” ebenfalls berücksichtigt wird. So ist beispielsweise der Name einer Gemeinde im Gemeindegrenzen-Datensatz der swisstopo “Schwanden (GL)”, wohingegen er bei GeoNames lediglich “Schwanden” lautet. Die beiden Filterschritte versuchen auf einfache Weise die passenden GeoNames-Objekte zu ermitteln, welche eine betroffene Gemeinde verkörpern. Dies soll für den Use Case ausreichen, obwohl im Bereich der *name disambiguation* viele Arbeiten existieren.

Mittels den genannten Arbeitsschritten können die selektierten GeoNames-Objekte und die selektierten Gemeinden aus dem Geodatensatz Gemeindegrenzen in das RDF-Format umgeformt werden. Die Tripel aus den GeoNames-Daten werden als Objekte der Klasse `OrteGN` und die Tripel der Gemeindegrenzen-Daten als Objekte der Klasse `Gemeinden` in den Triple Store gespeichert. Der Quelltext 9.1 zeigt einen Auszug der Tripel der Gemeinde “Collonges” der Klasse `OrteGN` (ohne RDFS Definitionen).

```

1 @Prefix orteGN:<http://www.wsl.example/ProJB/Daten/GD/OrteGN/>
2 @Prefix jbVoc:<http://www.wsl.example/ProJB/Ontologie/ProJBVokabular#>
3 @Prefix ogc:<http://www.opengis.net/rdf#>
4 @Prefix xsd:<http://www.w3.org/2001/XMLSchema#>
5 @Prefix owl:<http://www.w3.org/2002/07/owl#>
6
7 orteGN:GN2661146 jbVoc:id "GN2661146"^^xsd:string .
8 orteGN:GN2661146 jbVoc:name "Collonges"^^xsd:string .

```

```

9  orteGN:GN2661146 owl:sameAs <http://sws.geonames.org/2661146/about.rdf> .
10
11 orteGN:GN2661146 ogc:hasGeometry orteGN:OrteGNGeom2661146 .
12 orteGN:OrteGNGeom2661146 ogc:asWKT
13     "<http://www.opengis.net/def/crs/EPSG/0/21781>
14     Point(568769.5808 113477.94))"^^ogc:WKTLiteral .
15
16 orteGN:GN2661146
17     jbVoc:wirdTangiert
18     <http://www.wsl.example/ProJB/Daten/GD/JB/JB476960> .

```

Quelltext 9.1: Tripel der Gemeinde “Collonges” der Klasse `OrteGN`.

Die ersten und zweiten Tripel (Zeilen 7/8) legen ID und Namen fest. Durch `owl:sameAs` (Zeile 9) wird die Verknüpfung von der eigenen Ressource (Objekt der Klasse `OrteGN`) zur fremden Ressource in GeoNames hergestellt. Da die Daten aus GeoNames nach administrativen Einheiten gefiltert wurden, repräsentieren beide Ressourcen das gleiche Objekt in der Wirklichkeit, nämlich eine Gemeinde. Die Tripel der Zeilen 11 - 14 definieren die Geometrie im vorgesehenen GeoSPARQL-Standard. Dadurch können sie bei einer späteren Implementierung dieses Standards direkt verwendet werden. Dies ist auch der Grund, weshalb die Objekte aus GeoNames zusätzlich im eigenen Triple Store abgelegt und nicht direkt referenziert werden. So würde bei einer direkten Verknüpfung nur das letzte Tripel (Zeile 16-18) in abgeänderter Form verwendet:

```

<http://sws.geonames.org/2661146/about.rdf>
    jbVoc:wirdTangiert
    <http://www.wsl.example/ProJB/Daten/GD/JB/JB476960> .

```

9.3 Resultate

Im Folgenden werden die Resultate für die Abfrage nach den Jagdbanngebieten “Kärpf”, “Grand Muveran” und “Piz Ela” vorgestellt. Nachdem die nötigen Prozessierungsschritte (gemäss Abbildung 9.1) durchlaufen und die generierten RDF-Daten in den Triple Store geladen wurden, konnte die Abfrage aus dem Quelltext 9.2 abgesetzt werden.

```

1  SELECT DISTINCT ?GeoNamesName ?JagdbanngebietName ?Verknuepfung
2  WHERE{
3  {?Jagdbanngebiet jbVoc:jbName "Kärpf"^^xsd:string .}
4  UNION {?Jagdbanngebiet jbVoc:jbName "Grand Muveran"^^xsd:string .}
5  UNION {?Jagdbanngebiet jbVoc:jbName "Piz Ela"^^xsd:string .}
6  ?orteGN jbVoc:wirdTangiert ?Jagdbanngebiet .
7  ?orteGN owl:sameAs ?Verknuepfung .
8  ?orteGN jbVoc:name ?GeoNamesName .
9  ?Jagdbanngebiet jbVoc:jbName ?JagdbanngebietName .
10 }

```

Quelltext 9.2: Abfrage der betroffenen Gemeinden der Jagdbanngebiete “Kärpf”, “Grand Muveran” und “Piz Ela”.

Es sollen die Namen der Gemeinden gemäss GeoNames, das zugehörige Jagdbanngebiet und die Verknüpfung zu GeoNames zurückgegeben werden. In den ersten drei Tripelmustern (Zeilen 3 - 5) werden die gefragten Jagdbanngebiete festgelegt. Der Befehl `UNION` legt fest, dass mindestens eines dieser drei Tri-

pelmuster passen muss. Danach werden in den Zeilen 6 - 9 die entsprechenden Attribute gesucht, welche zurückgegeben werden sollen.

Die Tabelle 9.1 zeigt die Resultate dieser Abfrage (ohne die Verknüpfungen). Insgesamt wurden 26 Gemeinden gefunden. Zum Vergleich sind in der rechten Spalte die Resultate einer äquivalenten Abfrage über den in RDF umgeformten Gemeindegrenzen-Datensatz, welche 28 Einträge zurückgegeben hat, zu sehen. Konkret fehlen beim Jagdbanngebiet Kärpf zwei Gemeinden - nämlich "Diesbach (GL)" und "Hätzingen". Bei einer manuellen Suche bei geonames.org wird klar, dass Einträge zu diesen beiden Gemeinden existieren, aber nicht als administrative Einheiten, sondern als "bewohnter Ort" codiert sind. Anhand dieser Abweichung wird die bereits angesprochene Problematik mit nutzergenerierten Daten ersichtlich.

Jagdbanngebiet	Ort aus GeoNames	Ort aus Gemeindegrenzen
Kärpf	Andiast	Andiast
	Betschwanden	Betschwanden
	Elm	Diesbach (GL)
	Engi	Elm
	Haslen	Engi
	Linthal	Haslen
	Luchsingen	Hätzingen
	Matt	Linthal
	Mitlödi	Luchsingen
	Rüti (GL)	Matt
	Schwanden	Mitlödi
	Sool	Rüti (GL)
	Waltensburg/Vuorz	Schwanden (GL)
		Sool
		Waltensburg/Vuorz
Grand Muveran	Bex	Bex
	Chamoson	Chamoson
	Collonges	Collonges
	Conthey	Conthey
	Fully	Fully
	Gryon	Gryon
	Lavey-Morcles	Lavey-Morcles
	Leytron	Leytron
	Ollon	Ollon
	Ormont-Dessus	Ormont-Dessus
Piz Ela	Bergün/Bravuogn	Bergün/Bravuogn
	Filisur	Filisur
	Tinizong-Rona	Tinizong

Tabelle 9.1: Resultat der Abfrage nach den Jagdbanngebieten "Kärpf", "Grand Muveran" und "Piz Ela".

9.4 Diskussion

In diesem Kapitel wurden die Grundlagen zur Lösung des komplexen Use Case "Betroffene Gemeinden" und die Resultate einer konkreten Abfrage erläutert. Im Vordergrund dieses Anwendungsfalles steht die Verknüpfung mit dem fremden Datensatz von GeoNames und weniger die Geoprozessierung.

Eine Hauptproblematik besteht in der Ermittlung der GeoNames-Objekte, welche eine Gemeinde repräsentieren, auf denen das entsprechende Jagdbanngebiet liegt. Zum einen fehlt die räumliche Ausdehnung

der GeoNames-Objekte (da Punkte) und zum anderen sollen diese Objekte die administrative Einheit “Gemeinde” repräsentieren und nicht etwa ein gleichnamiges Restaurant oder einen gleich benannten Wald. Bereits bei der Abfrage der Daten von GeoNames über den Webservice werden nur Objekte zurückgegeben, die einer administrativen Einheit entsprechen. Durch einfache Filterungen, indem die Namen der GeoNames-Objekte und der Gemeindegrenzen-Objekte verglichen werden, sollen die entsprechenden GeoNames-Objekte selektiert werden, welche eine von den Jagdbanngebieten betroffene Gemeinde repräsentieren. In zukünftigen Entwicklungen könnte in Betracht gezogen werden, dass jeweils für sämtliche Gemeinden der Schweiz die entsprechenden GeoNames-Einträge ermittelt werden. Dazu müssten die Filterschritte ausgebaut werden, um möglichst korrekte Verknüpfungen zu GeoNames bereitzustellen. So wären mit jeder Abfrage über die Gemeindegrenzen auch automatisch die Verknüpfungen zum entsprechenden GeoNames-Objekt vorhanden.

Die Resultate des Anwendungsfalles sind wesentlich von der Qualität der GeoNames-Daten abhängig. Da es sich um nutzergenerierte Inhalte handelt, müssen gewisse Ungenauigkeiten und Fehler akzeptiert werden.

Würde in Zukunft GeoSPARQL so implementiert werden, dass geometrische Abfragen möglich sind, könnte der Anwendungsfall ohne Hinzunahme der GIS Operationen gelöst werden. Dazu müsste ein SPARQL bzw. GeoSPARQL Zugang zu GeoNames erstellt werden und die Daten könnten so direkt abgefragt werden.

Komplexer Use Case “Ähnliche Gebiete”

Der komplexe Use Case “Ähnliche Gebiete” lautet: *Können Teilgebiete bzw. ganze Jagdbanngebiete aus-
geschieden werden, welche Ähnlichkeiten zu einem anderen Teilgebiet bzw. Jagdbanngebiet aufweisen?*

Dieser Anwendungsfall erlaubt einen breiten Interpretationsspielraum bezüglich der Ähnlichkeit, denn diese kann aufgrund unterschiedlichster Kriterien definiert werden. Im Rahmen dieser Arbeit wurden vier Kriterien bestimmt, anhand derer die Ähnlichkeit ermittelt werden soll:

- Prozentualer Wald- und Felsanteil bezüglich der Gesamtfläche eines Objektes.
- Prozentualer Anteil des Verbreitungsgebietes von Steinbockkolonien bezüglich der Gesamtfläche eines Objektes.
- Prozentuale Verteilung des Waldes auf die vier verschiedenen Höhenstufen eines Objektes.
- Durchschnittliche, maximale und minimale Steigung eines Objektes.

Eine Besonderheit in Bezug auf die Implementierung dieses Ansatzes ist die Bereitstellung von bereits vorprozessierten Datensätzen. Dies wurde vor allem aus Zeitgründen gemacht, damit die räumlichen Operationen nicht so zeitintensiv sind, wie beispielsweise beim ersten komplexen Use Case “Waldflächen und Höhenstufen”.

In diesem Kapitel wird die Implementierung des Anwendungsfalles vorgestellt. Dazu gehören besonders die Kriterien, anhand deren die Ähnlichkeit bestimmt wird. Zudem werden die Resultate einer konkreten Ähnlichkeitsanalyse präsentiert und zum Schluss folgt eine kurze Diskussion der Implementierung.

10.1 Bestimmung der Ähnlichkeit

Insgesamt wird die Ähnlichkeit anhand von vier Kriterien bestimmt, hinter welchen eine ganze Reihe von Prozessierungen stehen. Das Objekt, für welches ähnliche Gebiete ermittelt werden soll, wird als Ausgangsobjekt bezeichnet, während die zu überprüfenden Objekte als Vergleichsobjekte bezeichnet werden. Im folgenden werden diese Kriterien genauer erläutert:

Der prozentuale **Wald- und Felsanteil** eines Gebietes kann auf einfache Weise ermittelt werden, da bereits zwei grundlegende Datensätze vorprozessiert bereitgestellt werden. Diese entstanden durch Verschneidung (*intersect*) der Primärflächen mit den Jagdbanngebieten. Dadurch wurden die einzelnen Objekte den Jagdbann- bzw. Teilgebieten zugeordnet. Auf die Prozessierung bezogen, müssen lediglich die als “Wald” bzw. “Fels” codierten Objekte in den gefragten Gebieten selektiert und verrechnet werden. Über den Triple Store wird der Pfad zum Geodatenatz der Jagdbanngebiete abgefragt, sowie die ID des Ausgangsobjektes. Zusätzlich könnte auch die Gesamtfläche eines Gebietes, anhand derer die prozentualen Anteile berechnet werden, über den Triple Store abgefragt werden, jedoch wurde die Gesamtfläche

hier direkt aus dem Geodatenatz gelesen.

Die Ermittlung des prozentualen **Anteils des Verbreitungsgebietes der Steinbockkolonien** zieht einen weiteren Geodatenatz in den Anwendungsfall mit ein. Dieser wird aber nicht in den Triple Store gespeichert – weder die Metadaten, noch die einzelnen Objekte. Die Analyse verläuft ähnlich zum ersten Kriterium, indem die Steinbockkolonien mit den Jagdbanngebieten überschnitten (*intersect*) und daraus die prozentualen Anteile berechnet werden. Eine andere Möglichkeit wäre anstelle des Anteils, die Ermittlung eines Booleans. Das heisst, es wird lediglich überprüft, ob ein Jagdbanngebiet bzw. Teilgebiet mit den Verbreitungsgebieten der Steinbockkolonien überlappt. Jedoch besteht hier der Nachteil, dass eine messbare Grösse fehlt und dadurch beispielsweise auch Vergleichsobjekte als ähnlich gelten, die lediglich einen minimalen Teil der Steinbockkolonien beinhalten.

Beim Geodatenatz der Steinbockkolonien handelt es sich um einen vom BAFU veröffentlichten Datensatz.¹ Er beschreibt die maximalen Einstandsgebiete (Sommer- und Wintereinstände) der Steinbockkolonien.

Der Vergleich der **Verteilung der Waldfläche auf die verschiedenen Höhenstufen** lehnt sich an den ersten komplexen Anwendungsfall “Waldflächen und Höhenstufen” an und bedient sich dementsprechend derselben Datenprozessierung. Jedoch wurde die Prozessierung zeitlich stark verkürzt, indem bereits vorprozessierte Daten anstelle der räumlichen Operationen eingesetzt werden. So wurden die Verschneidung der Primärflächen mit den Jagdbanngebieten (inklusive deren Exposition) und die Vektorisierung des gemäss den Höhenstufen reklassifizierten Höhenrasters vorprozessiert. Damit müssen im Rahmen dieses Anwendungsfalles nur noch die beiden vorprozessierten Datensätze miteinander überschnitten werden, um die Höhenverteilung des Waldes ermitteln zu können.

Um die **durchschnittliche, maximale und minimale Steigung** eines Objektes zu ermitteln, wird auf einen vorprozessierten Rasterdatensatz zurückgegriffen. Dieser repräsentiert die Steigung, welche basierend auf dem digitalen Höhenmodell RIMINI berechnet wurde. Über den Triple Store wird der Pfad zum Geodatenatz der Jagdbanngebiete abgefragt. Danach wird das Steigungsraster, jeweils gemäss den gefragten Jagdbanngebieten bzw. Teilgebieten, extrahiert (*extract by mask*) und die wichtigsten statistischen Werte des extrahierten Rasters berechnet. Durch Auslesen der so berechneten durchschnittlichen, maximalen und minimalen Steigung können die Vergleichsobjekte direkt mit dem Ausgangsobjekt verglichen werden.

10.2 Datenprozessierung

Die Abbildung 10.1 zeigt die Abfolge der Prozesse zur Lösung des Anwendungsfalles auf. Zuerst wird festgelegt, für welches Gebiet ähnliche Gebiete ermittelt werden sollen. Da ein Jagdbanngebiet aus Teilgebieten bestehen kann, wird diese Analyse auf Ebene der Objekte durchgeführt. Ein Objekt kann also ein Teilgebiet oder ein ganzes Jagdbanngebiet sein (sofern es nur aus einem Gebiet besteht). Danach folgt der Vergleich der Wald- und Felsanteile zwischen dem Ausgangsobjekt und den Vergleichsobjekten. Die Anteile dürfen jeweils 20% über oder unter den Werten des Ausgangsobjektes liegen. Da es sich bei diesem Kriterium um den ersten Prozessierungsschritt handelt, werden die Werte für sämtliche Objekte des Datensatzes Jagdbanngebiete erhoben und mit dem Ausgangsobjekt verglichen. Die daraus selektierten Vergleichsobjekte werden nachfolgend nach dem Kriterium des Anteils der Verbreitungsgebiete der Steinbockkolonien weiter selektiert. Hier dürfen die Werte der Vergleichsobjekte um 40% abweichen. Das nächste Kriterium vergleicht die Verteilung der Waldflächen auf die Höhenstufen, wobei die Werte

¹ <http://www.bafu.admin.ch/gis/02911/07403/index.html?lang=de>, 14.01.2012



Abbildung 10.1: Ablaufdiagramm des komplexen Use Case “Ähnliche Gebiete”.

um 30% vom Ausgangsobjekt abweichen dürfen. Zuletzt erfolgt die Ermittlung der durchschnittlichen, maximalen und minimalen Steigung mit einer Toleranz von 5 Grad Abweichung. Als Schlussresultat der Prozessierungsreihe werden die Identifikationsnummern der als ähnlich geltenden Gebiete ausgegeben. Die Kriterien zur Bestimmung der Ähnlichkeit basieren nicht auf realen ökologischen Kriterien, sondern wurden willkürlich bestimmt, um einen möglichen Ablauf zur Ausscheidung ähnlicher Gebiete zu testen. Gleiches gilt für die Parameter der Kriterien. Sie wurden ebenfalls willkürlich festgelegt, damit schlussendlich mindestens ein ähnliches Gebiet ausgeschieden wird, wodurch die Funktionsfähigkeit aller Prozessierungsschritte bestätigt werden konnte. Falls später reale ökologische Kriterien hinzugezogen werden, können die bestehenden einfach ausgetauscht werden.

10.3 Resultate

Kriterium	Werte des Ausgangsobjektes "Kiental"	Anzahl ähnlicher Gebiete	Werte des ähnlichen Objek- tes "Teilgebiet Kärpf"
Wald- und Felsfläche	Anteil Waldfläche: 16% Anteil Felsfläche: 17%	67	Anteil Waldfläche: 6% Anteil Felsfläche: 14%
Steinbock- kolonien	Anteil Gebiete von Steinbockko- lonien: 6%	17	Anteil Gebiete von Steinbockko- lonien: 4%
Höhen- verteilung Wald	Anteil Wald colline Stufe: 0% Anteil Wald montane Stufe: 55% Anteil Wald subalpine Stufe: 31% Anteil Wald alpine Stufe: 14%	3	Anteil Wald colline Stufe: 0% Anteil Wald montane Stufe: 36% Anteil Wald subalpine Stufe: 39% Anteil Wald alpine Stufe: 25%
Steigung	Durchschnittliche Steigung: 24 Grad Maximale Steigung: 52 Grad Minimale Steigung: 1 Grad	1	Durchschnittliche Steigung: 24 Grad Maximale Steigung: 47 Grad Minimale Steigung: 1 Grad

Tabelle 10.1: Aufstellung der Resultate der Ähnlichkeitsanalyse mit dem Jagdbanngebiet "Kiental".

Im folgenden werden die ermittelten ähnlichen Gebiete für das Jagdbanngebiet "Kiental" vorgestellt, welches aus einem einzigen Objekt besteht. Die Tabelle 10.1 zeigt eine Aufstellung der Resultate. Schlussendlich wurde nur ein Vergleichsobjekt, ein Teilgebiet des Jagdbanngebietes "Kärpf", als ähnlich ermittelt. Dessen Werte sind in der letzten Spalte der Tabelle aufgeführt.

10.4 Diskussion

In diesem Kapitel wurden die Grundlagen zur Lösung des komplexen Use Case “Ähnliche Gebiete” und die Resultate einer konkreten Abfrage erläutert. Der Anwendungsfall greift auf eine Reihe von Prozessierungsschritten zurück, um die Ähnlichkeit von Gebieten zu ermitteln. Durch die offene Formulierung des Use Case entsteht eine breite Spanne an möglichen Implementierungen. So muss entschieden werden, anhand welcher Kriterien die Ähnlichkeit ermittelt werden soll. Diese Implementierung könnte durch weitere Kriterien ergänzt und/oder um bestehende Kriterien gekürzt werden.

Im Rahmen der Implementierung dieses Anwendungsfalles wurde auf einige vorprozessierte Datensätze zurückgegriffen, um die teilweise zeitintensiven räumlichen Operationen zu umgehen. Dabei wird wieder die Frage nach dem Grad der Vorprozessierung aufgeworfen. Bei einem hohen Grad würde es durchaus Sinn machen diese Datensätze mit ihren Metadaten im Triple Store zu speichern, damit die Benutzer ebenfalls Informationen zu diesen Datensätzen erhalten können. Sie könnten beispielsweise in einer eigenen Klasse als “Unterstützende Geodatensätze” abgelegt werden. Dadurch würde auch der Austausch zwischen dem Triple Store und den räumlichen Operationen verstärkt, indem beispielsweise die Pfade dieser Datensätze abgefragt werden.

Aus dem Use Case resultieren Werte, die für Jagdbanngebiete bzw. Teilgebiete stehen. Die Bedeutung von Linked Data liegt hier vor allem in der Bereitstellung der Metadaten. So könnte Linked Data im Vorfeld der Prozessierung interessante Informationen hervorbringen, indem beispielsweise ein Geodatensatz aus einer fremden Quelle entdeckt wird, der für ein weiteres Ähnlichkeitskriterium verwendet werden sollte. Durch die Angabe des Pfades kann dieser gefunden und mit den räumlichen Operationen verwendet werden. Linked Data würde also eher als Werkzeug zur Ermittlung potentiell interessanter Datensätze verwendet werden und weniger als direkte Datenquelle.

Schlussfolgerungen und Ausblick

Wie in dieser Arbeit bereits dargestellt wurde, sind Rasterdaten im Linked Data Kontext bisher eher im Hintergrund gestanden, weshalb entsprechend wenige Ansätze zu deren Integration existieren. Generell bestehen bezüglich Linked Data noch Lücken zwischen dem Stand der Entwicklung und den gewünschten Möglichkeiten im Umgang mit räumlichen Daten (Heath u. Goodwin, 2011). Entsprechend bewegt sich diese Arbeit im Umfeld der aktuellen Forschung und Entwicklung. Es wurden zwei Ansätze entwickelt, die einen Beitrag zur Schliessung der genannten Lücken leisten sollen.

11.1 Schlussdiskussion

Der Ansatz VKE

Der Ansatz der Vektorisierten Kleinsten Einheiten (VKE) wurde im Wesentlichen in drei Punkten kritisiert, die bei abschliessender Betrachtung in einen grösseren Kontext eingeordnet werden können:

- Damit der vereinfachte Anwendungsfall “Waldfläche und Höhe simpel” mit VKE gelöst werden konnte, mussten die **Daten vor der Umformung in RDF umfassend prozessiert werden**. Dies ist unter anderem auf die Ausscheidung der Kleinsten Einheiten zurückzuführen, um die Informationen topologisch verorten zu können. Beispielsweise muss eine bestimmte Rasterzelle nicht vollständig innerhalb einer Waldfläche liegen, sondern kann zu einem Viertel auch eine Felsfläche betreffen. Werden neue Datensätze hinzugenommen, könnten die Kleinsten Einheiten noch kleiner werden und in ihrer Anzahl steigen. Eine Möglichkeit um von diesen Kleinsten Einheiten wegzukommen, könnte die Änderung des Ansatzes zu einer approximierten Topologie sein. Statt der Ermittlung der Kleinsten Einheiten werden die Rasterzellen in ihrer Form belassen und als einzelne Zellen ungefähr den grösseren Objekten zugeordnet. So kann beispielsweise eine Zelle einem bestimmten Jagdbanngebiet zugeordnet werden, obwohl die Zelle nicht vollständig darin liegt.
- Obwohl die Rasterdaten mit dem Ansatz VKE relativ umfassend in den Triple Store gespeichert werden, können diese nur **in beschränktem Masse verwendet** werden. Dies ist vor allem auf zwei Gründe zurückzuführen: Es können nur Informationen attributiv abgefragt werden, die als Tripel vorhanden sind. Dadurch können auch nur topologische Beziehungen abgefragt werden, die explizit angegeben sind. Der zweite Grund ist die fehlende Möglichkeit geometrische Abfragen durchzuführen, wie z. B., ob ein bestimmter Punkt in einem bestimmten Polygon (*point in polygon*) liegt. Diese Kritik gilt nicht nur für VKE, sondern generell für räumliche Abfragen im Linked Data Kontext (Heath u. Goodwin, 2011).

Diese Tatsache führt zur grundsätzlichen Frage, ob an Linked Data ungeeignete Anforderungen gestellt werden? Ob die Erwartungen zu ähnlich sind, wie man sie an ein GIS stellt? Betrachtet

man die Use Cases, tendieren diese eher in eine Richtung, welche die Verwendung von räumlichen Operationen nahe legen. Linked Data ist ursprünglich für die Verknüpfung, Bereitstellung und Entdeckung von Daten gedacht (Heath u. Bizer, 2011). Andererseits wird aber mit dem aufkommenden GeoSPARQL-Standard suggeriert, dass einfache räumliche Operationen in Zukunft möglich sein sollen (Perry u. Herring, 2011). Dies wäre besonders für die Entdeckung von räumlichen Daten interessant. So könnten beispielsweise Daten abgefragt werden, die räumlich in einem bestimmten Gebiet liegen. Hierbei stellt sich aber wiederum die Frage, wie eine solche Abfrage über das Web der Daten umgesetzt werden könnte. Eine sicherlich wichtige Komponente für effiziente räumliche Abfragen bilden räumliche Indizes, die bezogen auf Linked Data noch in der Entwicklung stecken (Battle u. Kolas, 2011).

- Es wurde auf den verhältnismässig **hohen Speicherbedarf der RDF-Daten** aufmerksam gemacht, was zum Schluss noch etwas differenzierter betrachtet werden soll. Dieser erhöhte Speicherbedarf dürfte kein grösseres Problem darstellen, gerade hinsichtlich der heutzutage günstig verfügbaren und hohen Speicherkapazitäten. Ein extremes Beispiel für die Verarbeitung grosser Datenmengen sind die Experimente mit dem Teilchenbeschleuniger des CERN in Genf, welche jährlich 15 Petabytes (15 Millionen Gigabytes) an Daten produzieren.¹ Dies zeigt in Relation zu den generierten RDF-Daten, dass deren erhöhter Speicherbedarf nicht als ein primär limitierender Faktor angesehen werden kann. Wesentlich entscheidender sind die Kapazitäten der im Linked Data Bereich entwickelten Applikationen. Beispielsweise sollte ein Triple Store nicht nur mit einer hohen Anzahl Tripel, sondern auch mit Tripeln, die einen hohen Speicherbedarf haben, umgehen können.

Der Ansatz ARO

Der Ansatz der Ausgliederten Räumlichen Operationen (ARO) versucht die Lücken, welche mit räumlichen Daten in Kombination mit Linked Data bestehen, durch Einbindung von räumlichen Operationen und räumlichen Datenbanken zu überbrücken. Dadurch soll die Lösung der komplexen Use Cases ermöglicht werden. Die Rasterdaten werden ausserhalb des Triple Stores gespeichert und bei Bedarf in die räumlichen Operationen miteinbezogen. Dadurch gerät die Fragestellung, wie Rasterdaten in RDF repräsentiert werden können in den Hintergrund, da lediglich die Metadaten als Tripel beschrieben werden. Zudem muss die Ontologie, bezogen auf die topologischen Beziehungen, weniger ausgereift sein, da die räumlichen Operationen unabhängig von den darin definierten Beziehungen funktionieren.

Eine Erweiterungsmöglichkeit um Rasterdaten stärker einzubinden, könnte der Einbezug des Ansatzes von Uery u. Varanka (2011) sein, womit zumindest relevante Objekte aus dem Raster über Linked Data verfügbar wären. Dies könnte für Benutzer unterstützend wirken, um ein besseres Bild von den angebotenen Rasterdaten zu erhalten.

Der Ansatz ARO kann als eine einfache Vorstufe zu einem "räumlichen Triple Store" betrachtet werden (vergleichbar zu einer relationalen Datenbank mit einer räumlichen Erweiterung). Ein wesentlicher Vorteil von ARO ist die Flexibilität, da unterschiedlichste räumliche Operationen durchgeführt werden können.

¹ <http://www.lhc-facts.ch/index.php?page=datenverarbeitung>, 16.01.2012

11.2 Erreichtes und Erkenntnisse

Auf die in der Einleitung vorgestellten Zielsetzungen bezogen, konnte folgendes erreicht werden:

- *Diskussion der bestehenden Ansätze, um räumliche Daten in den Linked Data Kontext zu integrieren.*

Es wurden verschiedene bestehende Ansätze zur Umformung von räumlichen Daten in RDF vorgestellt und diskutiert. Auf Seite der Vektordaten wurde auf bestehende Vorschläge zurückgegriffen, namentlich auf den aufkommenden GeoSPARQL-Standard. Dadurch wird beabsichtigt die Daten bereits im zukünftigen Format zu halten. Dementsprechend hohe Erwartungen werden an GeoSPARQL gestellt, dass der Standard eine wichtige Lücke im Umgang mit räumlichen Daten schließen kann. Hierzu zählen besonders die Durchführung von geometrischen Abfragen und die globale Akzeptanz des Standards, damit die angebotenen Daten direkt abgefragt und verwendet werden können. Für einen Erfolg des Standards müssen auch die entsprechenden Applikationen (z. B. Triple Stores) die angekündigten Funktionen von GeoSPARQL implementieren.

Ein wesentlicher Vorteil von GeoSPARQL ist die Verwendung von WKT bzw. GML, um die Geometrien zu repräsentieren. Dadurch kann der String, welcher die Beschreibung der Geometrie in einem der beiden Formate enthält, ausgelesen und beispielsweise direkt in einem GIS verwendet werden.

Für die Repräsentation von Rasterdaten kann aufgrund dieser Arbeit keine Methode besonders hervorgehoben werden. Entweder werden die Raster in irgendeiner Form vektorisiert oder sie werden beispielsweise gemäss der GML Coverage Syntax in einem String abgelegt (vgl. Usery u. Varanka, 2011). Die Ausarbeitung weiterer Methoden wäre wünschenswert, um zwischen diesen eine Evaluation durchzuführen und daraus schlussendlich eine Einigung für eine gemeinsame Methode hervorgehen würde (z. B. Erweiterung von GeoSPARQL mit Rasterfunktionen).

Neben der Frage wie die Geodaten als RDF formatiert werden können, spielen auch deren räumlichen Beziehungen eine wichtige Rolle. Um diese abzubilden kann zwischen zwei entgegengesetzten Möglichkeiten unterschieden werden:

Die eine Möglichkeit basiert auf der expliziten Angabe der topologischen Beziehungen. Dazu müssen diese Topologien durch Vorprozessierungen ermittelt werden. Somit können räumliche Abfragen formuliert werden, aber nur wie die Ontologie dies vorgesehen hat. Dazu werden keine Geometrien benötigt. So wäre es einen Versuch wert die topologischen Beziehungen intensiver zu untersuchen, damit für die Lösung der Anwendungsfälle weniger räumliche Operationen benötigt werden. Dies bedingt aber andererseits eine verstärkte Vorprozessierung mit räumlichen Operationen.

Die andere Möglichkeit räumliche Beziehungen einzubinden bezieht sich auf geometrische Abfragen, wie dies verstärkt beim Ansatz ARO eingesetzt wird. Das heisst, die räumlichen Beziehungen werden nicht vorprozessiert, sondern können während der Abfrage bzw. des Lösungsprozesses ermittelt werden. Mit der Umformung der Geometrien in RDF folgt konsequenterweise auch die Forderung nach geometrischen Abfragen bzw. geometrischen Prozessierungen. Denn RDF ist nicht primär ein Austauschformat, sondern ein "Beschreibungsformat", worauf unterschiedliche Abfragen durchgeführt werden sollen. Wie bereits bekannt befinden sich die geometrischen Abfragen über RDF-Daten noch in Entwicklung, weshalb beispielsweise bei ARO auf externe räumliche Operationen zurückgegriffen wird (Heath u. Goodwin, 2011). Der vorgeschlagene GeoSPARQL-Standard bezieht die beiden beschriebenen Möglichkeiten zur Abbildung räumlicher Beziehungen mit ein.

- *Entwicklung möglicher Ansätze zum Umgang mit Rasterdaten im Kontext von Linked Data.*

Es wurden zwei verschiedene Ansätze entwickelt, die Raster- wie auch Vektordaten miteinbeziehen. Der Ansatz VKE hat sich intensiver mit der Rasterproblematik auseinandergesetzt und eine Möglichkeit aufgezeigt, wie Raster als RDF gespeichert und abgefragt werden könnten. Der Ansatz ARO hat nicht die Repräsentation der Raster in RDF zum Ziel, sondern speichert lediglich die Metadaten als Tripel ab. Die generell mit räumlichen Daten verbundenen Schwierigkeiten werden an externe Technologien ausgelagert. Daher nimmt ARO die Möglichkeit zur räumlichen Prozessierung, in Form einer aus verschiedenen Technologien kombinierten Übergangslösung, vorweg, bis räumliche Daten adäquat innerhalb des Linked Data Kontext verwendet werden können.

Durch die Strukturierung eines Rasters werden viele Informationen in RDF umgeformt, was ein wesentliches Problem darstellt. Denn durch die Strukturierung werden die im Raster implizit vorhandenen Informationen explizit als Tripel abgelegt. Damit verbunden ist die Entscheidung, welche Informationen und wie detailliert diese aus einem Raster extrahiert werden sollen. Diese Tatsache ist mitunter ein Grund, weshalb die bestehenden Ansätze für Vektoren nicht einfach auf Raster übertragen werden können.

- *Prüfen der verschiedenen Ansätze durch Implementierung praktischer Anwendungsfälle (proof of concept).*

Mit dem entwickelten Ansatz ARO konnten alle Anwendungsfälle gelöst werden. Jedoch handelt es sich beim Lösungsansatz ARO um einen Hybriden, indem einerseits auf Linked Data zurückgegriffen wird, aber andererseits auch externe Technologien hinzugenommen werden müssen. Hierbei sei angemerkt, dass die Anwendungsfälle auch mit einem GIS lösbar sind. Der grosse Nutzen von Linked Data entsteht vor allem, wenn die eigenen Datenbestände über Vokabulare/Ontologien strukturiert und fremde Datenquellen eingebunden werden. Das Letztere konnte in dieser Arbeit nur beschränkt umgesetzt werden, da heute noch zu wenig Datenbestände in RDF veröffentlicht sind, um eine grössere Anzahl an sinnvollen Verknüpfungen zu fremden Daten bereitzustellen.

In der Arbeit spielen die Verknüpfungen vor allem im zweiten komplexen Use Case “Betroffene Gemeinden” eine zentrale Rolle. Dabei hat sich aber auch gezeigt, dass die Bereitstellung von Verknüpfungen teilweise mit einigen Schwierigkeiten verbunden ist. So muss ermittelt werden, welche Ressourcen aus den unterschiedlichen Datenquellen zueinander passen, um eine korrekte Verknüpfung vornehmen zu können. So kann beispielsweise ein GeoNames-Objekt mit dem Toponym “Bern” die gleichnamige Stadt, den Kanton in der Schweiz oder ein Dorf im US-Bundesstaat Kansas repräsentieren. Zudem kommen bei GeoNames die mit nutzergenerierten Daten verbundenen Schwierigkeiten hinzu. So können unterschiedliche Schreibweisen oder Fehler die Suche nach den passenden Ressourcen erschweren. Beispielsweise wird eine Gemeinde bei der swisstopo als “Schwanden (GL)” bezeichnet, während sie bei GeoNames nur als “Schwanden” erfasst ist. Deshalb mussten beim Anwendungsfall “Betroffene Gemeinden” Methoden zugezogen werden, um trotz verschiedener Schreibweisen die entsprechenden GeoNames-Objekte zu finden (*name disambiguation*). Eine weitere Problematik könnte das Auffinden von Datenquellen sein, mit denen potentielle Verknüpfungen möglich wären. Denn einem Datenanbieter können nicht alle im Datenraum angebotenen Datensätze geläufig sein, d. h., es muss bekannt sein, wo nach solchen Verknüpfungen gesucht werden soll. Eine zusätzliche Herausforderung ergibt sich, wenn diese Verknüpfungen automatisiert generiert werden sollen. Die Bereitstellung von Verknüpfungen gestaltet sich als nicht einfach und bildet einen wichtigen Forschungsbestandteil im Linked Data Bereich (Heath u. Goodwin, 2011; Heath u. Bizer, 2011).

11.3 Fazit

Die wichtigsten drei Punkte noch einmal zusammengefasst:

- Es bestehen vielversprechende Ansätze zur Verwendung von räumlichen Daten mit Linked Data, die aber noch einiger Entwicklungsarbeit bedürfen. Dazu gehört besonders die Integration von Rasterdaten.
- Es besteht das Bedürfnis mit Linked Data Technologien räumliche Prozessierungen durchführen zu können, wodurch die bestehenden Abfragemöglichkeiten erweitert würden. Entsprechende Beiträge befinden sich in der Ausarbeitung.
- Linked Data dient der Veröffentlichung, Strukturierung, Verknüpfung sowie Abfrage von Daten und weniger der Datenprozessierung. Dementsprechend sollte auch die Abbildung der topologischen Beziehungen durch Ontologien berücksichtigt werden, da somit auf einfache Weise räumliche Abfragen (attributiv) möglich sind.

11.4 Ausblick

Die Entwicklung von Lösungen zum Umgang mit Geodaten im Kontext von Linked Data steht noch am Anfang (Heath u. Goodwin, 2011). Es bestehen noch einige Unklarheiten bezüglich der Repräsentation, sowie der möglichen Prozessierungen von räumlichen Daten. Zur Repräsentation von Vektordaten bestehen einige Ansätze, aber kein gemeinsamer Standard. Daher wäre es wünschenswert sich erst in diesem Bereich auf einen Standard zu einigen, damit unter anderem darauf basierend Lösungen für Rasterdaten entwickelt werden können. Dementsprechend sollten die Erwartungen an Linked Data noch nicht zu hoch gesteckt werden. Zudem muss man sich immer bewusst sein, dass Linked Data nicht den gleichen Zweck erfüllen soll, wie beispielsweise ein GIS. Trotzdem kann die Implementierung von räumlichen Operationen einen grossen Nutzen bringen, da somit die attributiven Abfragemöglichkeiten durch geometrische ergänzt werden. Dies würde auch die Entdeckung von impliziten Informationen unterstützen, die über geometrische Abfragen zu Tage gefördert werden könnten.

Der Ansatz ARO könnte um einige interessante Komponenten weiterentwickelt werden: Bisher wurde auf die räumlichen Operationen eines GIS zurückgegriffen. Diese könnten durch einen Webservices ersetzt werden, die über die entsprechenden Bibliotheken für solche Operationen verfügen. Die Operationen würden zusätzlich im Triple Store durch Metadaten beschrieben werden, wie z. B. eine Beschreibung der Operation und deren Identifikation über einen URI, welche über das Web verfügbar ist. Dadurch kann sich ein Benutzer nicht nur über Datensätze informieren, sondern auch über verschiedene Prozessierungsmöglichkeiten. Durch Angabe, wie auf beide zugegriffen werden kann, ist der Benutzer in der Lage selber Daten zu prozessieren. Zusätzlich könnte darüber nachgedacht werden, ob und wie die so neu generierten Daten wieder in das Web der Daten integriert werden sollen. Weiter wäre die Entwicklung einer benutzerfreundlichen Applikation notwendig, die alle beschriebenen Funktionen zur Verfügung stellt, damit auch Benutzer ohne Fachkenntnisse eingeschlossen werden können. Aufgrund der Use Cases können zwei wichtige räumliche Operationen ausgeschieden werden: Das sind die Verschneidungen (*intersect*) und die Überprüfung, ob eine Geometrie in einer anderen liegt (*contains*). Würden bereits diese zwei räumlichen Operationen in Linked Data integriert, könnten in Kombination mit der expliziten Angabe der Topologie viele räumliche Fragestellungen gelöst werden.

Eine wesentliche Limite stellt möglicherweise die heute wenig entwickelte Benutzerfreundlichkeit von Linked Data dar, was sicherlich auch daran liegt, dass die Technologie noch in der Entwicklungsphase ist. Um ein breites Publikum zu erreichen, wie dies das Web der Dokumente heute tut, müssen Applikationen entworfen werden, die es den Benutzern einfach machen das Web der Daten zu entdecken und Abfragen zu formulieren. So könnte beispielsweise die Entwicklung eines Browsers im Zentrum stehen, der die jeweiligen Vokabulare/Ontologien der angezeigten Ressourcen darstellen kann und die Möglichkeit bietet mit diesen Vokabularen/Ontologien Abfragen zu formulieren. Zudem sollten die in den Tripeln enthaltenen Informationen visualisiert werden können, wie beispielsweise das Abspielen von Filmen, Anzeigen von Bildern und in Bezug auf diese Arbeit, die Visualisierung der Geodaten. Durch Erweiterungen können zusätzliche Funktionen einbezogen werden, wie etwa ein räumliches Paket, das Geoprozessierungen ermöglicht.

Eine weitere Problematik könnte von den Vokabularen/Ontologien ausgehen. Durch die Strukturierung der Daten können viele unterschiedliche Vokabulare entstehen, da es sich teilweise um spezialisierte Bereiche handelt. Deshalb muss ein möglichst umfassendes und global verwendbares “Grundvokabular” etabliert werden, mit definierten Regeln für Übersetzungen. Eventuell sollte gar die Entwicklung einer “Übersetzungssprache” in Betracht gezogen werden (Bizer u. a., 2009a). Dieses “Grundvokabular” muss wiederum den Benutzern auf einfache Weise vermittelt werden. Denn eine Suche nach bestimmten Ressourcen gestaltet sich schwierig, wenn das Vokabular unbekannt ist. Gleich verhält es sich in dieser Arbeit: Nur wer mit dem Datenmodell vertraut ist, kann auch gezielte Abfragen durchführen.

Die Entwicklung im Linked Data Bereich wird sicherlich weiter anhalten, besonders auch im Hinblick der Realisierung des Semantic Web. Die Möglichkeit die heute anfallenden grossen Datenmengen zu bewirtschaften, ohne den Überblick zu verlieren, ist sicherlich von grossem Interesse. Daher werden räumliche Daten ein wichtiger Bestandteil bleiben und dementsprechend müssen auch offene Fragen in diesem Bereich beantwortet werden. Die Rasterdaten werden früher oder später mehr in das Zentrum des Interesses rücken. Besonders auch mit der erwarteten Veröffentlichung des GeoSPARQL-Standards könnte einiges ins Rollen gebracht werden. Denn einheitliche und anerkannte Standards sind unverzichtbar für einen globalen Erfolg.

Mit der wachsenden Verbreitung von Linked Data, werden Themen wie die Qualität der Daten, der Unterhalt der Verknüpfungen zwischen Datenquellen, der Datenschutz und die Privatsphäre, die Benutzerfreundlichkeit usw. immer wichtiger, ähnlich wie wir es heute bereits im Web der Dokumente vorfinden (Heath u. Goodwin, 2011).² Der Erfolg von Linked Data wird von der Akzeptanz der Communities abhängen. So könnte es zu einer kleinen revolutionären Weiterentwicklung im Web führen oder aber nur eine Parallelexistenz fristen, beschränkt auf fachliche Bereiche.

² vgl. Layered Semantic Web approach (Antoniou u. van Harmelen, 2008, S. 19/20)

- [Antoniou u. van Harmelen 2008] ANTONIOU, Grigoris ; VAN HARMELEN, Frank: *A semantic Web primer*. Cambridge, Massachusetts : MIT Press, 2008
- [Auer u. a. 2009] AUER, Sören ; LEHMANN, Jens ; HELLMANN, Sebastian: LinkedGeoData: Adding a Spatial Dimension to the Web of Data. In: *International Semantic Web Conference (ISWC)*, 2009
- [BAFU 2010] BAFU, Bundesamt für Umwelt: *Bundesinventar der eidgenössischen Jagdbanngebiete*. 2010
- [Battle u. Kolas 2011] BATTLE, Robert ; KOLAS, Dave: GeoSPARQL: Enabling a Geospatial Semantic Web. In: *Semantic Web – Interoperability, Usability, Applicability* (2011). <http://www.semantic-web-journal.net/sites/default/files/swj176.pdf>, 18.07.2011. – submitted (Mai 2011)
- [Berners-Lee 2009] BERNERS-LEE, Tim: *Linked Data – Design Issues*. <http://www.w3.org/DesignIssues/LinkedData.html>, 16.08.2011. Version: 06 2009
- [Bill u. Zehner 2001] BILL, Ralf ; ZEHNER, Marco L.: *Lexikon der Geoinformatik*. Heidelberg : Herbert Wichmann Verlag, 2001
- [Bizer u. a. 2009a] BIZER, Christian ; HEATH, Tom ; BERNERS-LEE, Tim: Linked Data - The Story So Far. In: *International Journal on Semantic Web and Information Systems* 5 (2009), Nr. 3, S. 1 – 22
- [Bizer u. a. 2009b] BIZER, Christian ; LEHMANN, Jens ; KOBILAROV, Georgi ; AUER, Sören ; BECKER, Christian ; CYGANIAK, Richard ; HELLMANN, Sebastian: DBpedia – A crystallization point for the Web of Data. In: *Journal of Web Semantics* 7 (2009), Nr. 3, S. 154 – 165
- [Bizer u. Schultz 2009] BIZER, Christian ; SCHULTZ, Andreas: The Berlin SPARQL Benchmark. In: *International Journal on Semantic Web and Information Systems* 5 (2009), S. 1 – 24
- [Brickley 2006] BRICKLEY, Dan: *W3C Semantic Web Interest Group. Basic Geo (WGS84 lat/long) Vocabulary*. <http://www.w3.org/2003/01/geo/>, 30.08.2011. Version: 02 2006
- [Bulen u. a. 2011] BULEN, Andrew ; CARTER, Jonathan J. ; VARANKA, Dalia E.: *A Program for the Conversion of The National Map Data from Proprietary Format to Resource Description Framework (RDF)*. Reston, Virginia : U.S. Geological Survey, 2011
- [Bundi 2007] BUNDI, Martin: Wild und Jagd im 19. Jahrhundert. In: LÜÖND, Karl (Hrsg.): *Den Jägern auf der Spur: 125 Jahre und eine Zukunft für JagdSchweiz*. Wohlen, Bern : Salm, 2007, S. 37 – 69
- [Burrough u. McDonnell 2006] BURROUGH, Peter A. ; MCDONNELL, Rachael A.: *Principles of Geographical Information Systems*. Oxford : Oxford University Press, 2006

- [Cohn u. a. 1997] COHN, Anthony G. ; BENNETT, Brandon ; GOODAY, John ; GOTTS, Nicholas M.: Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. In: *GeoInformatica* 1 (1997), S. 275 – 316
- [Diestel 2006] DIESTEL, Reinhard: *Graphentheorie*. Berlin, Heidelberg : Springer, 2006
- [Ellenberg 1996] ELLENBERG, Heinz: *Vegetation Mitteleuropas mit den Alpen in ökologischer, dynamischer und historischer Sicht*. Stuttgart : Ulmer, 1996
- [Gonseth u. a. 2001] GONSETH, Yves ; WOHLGEMUTH, Thomas ; SANSONNENS, Bertrand ; BUTTLER, Alexandre: *Die biogeographischen Regionen der Schweiz. Erläuterungen und Einteilungsstandard*. Bern : Bundesamt für Umwelt, Wald und Landschaft, 2001 (Umwelt Materialien)
- [Goodwin u. a. 2008] GOODWIN, John ; DOLBEAR, Catherine ; HART, Glen: Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web. In: *Transactions in GIS* 12 (2008), S. 19 – 30
- [Hart u. Dolbear 2007] HART, Glen ; DOLBEAR, Catherine: What's So Special about Spatial? In: SCHARL, Arno (Hrsg.) ; TOCHTERMANN, Klaus (Hrsg.): *The Geospatial Web. How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*. London : Springer, 2007, S. 39 – 44
- [Hausenblas 2009] HAUSENBLAS, Michael: Anreicherung von Webinhalten mit Semantik – Microformats und RDFa. In: BLUMAUER, Andreas (Hrsg.) ; PELLEGRINI, Tassilo (Hrsg.): *Social Semantic Web*. Berlin, Heidelberg : Springer, 2009, S. 147 – 158
- [Heath u. Bizer 2011] HEATH, Tom ; BIZER, Christian: *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011 (Synthesis Lectures on the Semantic Web: Theory and Technology)
- [Heath u. Goodwin 2011] HEATH, Tom ; GOODWIN, John: Chapter 4. Linking Geographical Data for Government and Consumer Applications. In: WOOD, David (Hrsg.): *Linking Government Data*. New York, Dordrecht, Heidelberg, London : Springer, 2011, S. 73 – 92
- [Hitzler u. a. 2009] HITZLER, Pascal ; KRÖTZSCH, Markus ; BIJAN, Parsia ; PATEL-SCHNEIDER, Peter F. ; RUDOLPH, Sebastian: *OWL 2 Web Ontology Language Primer*. <http://www.w3.org/TR/owl-primer/>, 25.01.2012. Version: 10 2009
- [Hitzler u. a. 2007] HITZLER, Pascal ; KRÖTZSCH, Markus ; RUDOLPH, Sebastian ; SURE, York: *Semantic Web. Grundlagen*. Berlin, Heidelberg : Springer, 2007
- [Landolt 2003] LANDOLT, Elias: *Unsere Alpenflora*. Bern : SAC-Verlag, 2003
- [Liebermann u. a. 2007] LIEBERMANN, Joshua ; SINGH, Raj ; GOAD, Chris: *W3C Geospatial Vocabulary*. <http://www.w3.org/2005/Incubator/geo/XGR-geo/>, 25.01.2012. Version: 10 2007
- [Longley u. a. 2011] LONGLEY, Paul A. ; GOODCHILD, Michael F. ; MAGUIRE, David J. ; W., Rhind D.: *Geographic Information Systems and Science*. Hoboken, N.J. : John Wiley & Sons, 2011
- [Manola u. a. 2004] MANOLA, Frank ; MILLER, Eric ; MCBRIDE, Brian: *RDF Primer*. <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>, 25.01.2012. Version: 02 2004

- [Perry u. Herring 2011] PERRY, Matthew ; HERRING, John: *GeoSPARQL - A geographic query language for RDF data. A proposal for an OGC Draft Candidate Standard*. Open Geospatial Consortium, 2011 <http://www.w3.org/2011/02/GeoSPARQL.pdf>, 11.09.2011
- [Portele 2007] PORTELE, Clemens: *OpenGIS® Geography Markup Language (GML) Encoding Standard*. Open Geospatial Consortium, 2007 http://portal.opengeospatial.org/files/?artifact_id=20509, 14.11.2011
- [Prud'hommeaux u. Seaborne 2008] PRUD'HOMMEAUX, Eric ; SEABORNE, Andy: *SPARQL Query Language for RDF*. <http://www.w3.org/TR/rdf-sparql-query/>, 03.10.2011. Version: 01 2008
- [Reed 2006] REED, Carl: *OGC White Paper. An Introduction to GeoRSS: A Standards Based Approach for Geo-enabling RSS feeds*. Open Geospatial Consortium, 2006 http://portal.opengeospatial.org/files/?artifact_id=15755, 09.11.2011
- [Rohloff u. a. 2007] ROHLOFF, Kurt ; DEAN, Mike ; EMMONS, Ian ; RYDER, Dorene ; SUMNER, John: An Evaluation of Triple-Store Technologies for Large Data Stores. In: MEERSMAN, Robert (Hrsg.) ; TARI, Zahir (Hrsg.) ; HERRERO, Pilar (Hrsg.): *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops* Bd. 4806. Berlin, Heidelberg : Springer, 2007, S. 1105 – 1114
- [Salas u. a. 2011] SALAS, Juan M. ; HARTH, Andreas ; NORTON, Barry ; VILCHES, Luis M. ; DE LEÓN, Alexander ; GOODWIN, John ; STADLER, Claus ; ANAND, Suchith ; HARRIES, Dominic: *NeoGeo Vocabulary: Defining a shared RDF representation for GeoData*. <http://geovocab.org/doc/neogeo.html>, 11.09.2011. Version: 05 2011
- [Shadbolt u. a. 2006] SHADBOLT, Nigel ; BERNERS-LEE, Tim ; HALL, Wendy: The Semantic Web Revisited. In: *IEEE Intelligent Systems* 21 (2006), Nr. 3, S. 96 – 101
- [Smith u. a. 2004] SMITH, Michael K. ; WELTY, Chris ; MCGUINNESS, Deborah L.: *OWL Web Ontology Language Guide*. <http://www.w3.org/TR/owl-guide/>, 25.01.2012. Version: 02 2004
- [Spanos u. a. 2010] SPANOS, Dimitrios-Emmanuel ; STAVOU, Periklis ; MITROU, Nikolas: Bringing Relational Databases into the Semantic Web: A Survey. In: *Semantic Web – Interoperability, Usability, Applicability* (2010). <http://www.semantic-web-journal.net/sites/default/files/swj121.pdf>. – submitted (Juni 2010), Accepted
- [Stadler u. a. 2011] STADLER, Claus ; LEHMANN, Jens ; HÖFFNER, Konrad ; AUER, Sören: LinkedGeoData: A Core for a Web of Spatial Open Data. In: *Semantic Web – Interoperability, Usability, Applicability* (2011). http://www.semantic-web-journal.net/sites/default/files/swj173_2.pdf, 21.11.2011. – submitted (Mai 2011), Accepted
- [swisstopo 2005] SWISSTOPO, Bundesamt für Landestopografie: *DHM25. Das digitale Höhenmodell der Schweiz*. Version: 2005. <http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/height/dhm25.parsysrelated1.41325.downloadList.16713.DownloadFile.tmp/dhm25infode.pdf>, 12.08.2011
- [swisstopo 2007] SWISSTOPO, Bundesamt für Landestopografie: *VECTOR25. Das digitale Landschaftsmodell der Schweiz*. Version: 2007. <http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/landscape/vector25.parsysrelated1.47641.downloadList.50692.DownloadFile.tmp/vector25infode.pdf>, 12.08.2011

- [swisstopo 2011] SWISSTOPO, Bundesamt für Landestopografie: *VECTOR200. Das kleinmassstäbliche digitale Landschaftsmodell der Schweiz*. Version: 2011. <http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/landscape/vector200.parsysrelated1.47641.downloadList.23908.DownloadFile.tmp/vecto200infode.pdf>, 12.08.2011
- [Usery u. Varanka 2011] USERY, E. L. ; VARANKA, Dalia: Design and Development of Linked Data from The National Map. In: *Semantic Web – Interoperability, Usability, Applicability* (2011). http://www.semantic-web-journal.net/sites/default/files/swj180_1.pdf, 21.11.2011. – submitted (Juni 2011), Accepted

Persönliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und die den verwendeten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Zürich, 29.01.2012

Simon Fleischli