

Master Thesis

im Rahmen des
Universitätslehrganges „Geographical Information Science & Systems“
(UNIGIS MSc) am Zentrum für GeoInformatik (Z_GIS)
der Paris Lodron-Universität Salzburg

zum Thema

„Semantische Annotation in OGC Sensor Web Enablement“

vorgelegt von

Dipl. Ing. (FH) Christian Ansorge
U1372, UNIGIS MSc Jahrgang

Zur Erlangung des Grades
„Master of Science (Geographical Information Science & Systems) – MSc(GIS)“

Gutachter:
Ao. Univ. Prof. Dr. Josef Strobl

Wien, 22. April 2010

Danksagung

Die vorliegende Arbeit wäre ohne verschiedene Formen der Unterstützung und Beratung nicht möglich gewesen. Ich möchte diese Gelegenheit nutzen, um mich bei Personen und Institutionen zu bedanken, welche mich im Prozess der Fertigstellung dieser Arbeit unterstützt haben.

An erster Stelle richte ich meinen Dank an das Umweltbundesamt GmbH und meine dortigen Kollegen, welche mich besonders im Gebiet der Semantik unterstützt und inspiriert haben. Ohne Ihre Hilfe wäre mir der Einstieg die vorliegende Thematik nicht möglich gewesen. Persönlich bedanken möchte ich daher besonders bei Katharina Schleidt und Barbara Magagna.

Desweiteren möchte ich mich bei Dr. Desiree Hilbring vom Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB und Dr. Sven Schade vom European Commission Joint Research Center für ihre Unterstützung und ihren Rat danken.

Zu Schluss gilt mein besonderer Dank meiner Lebensgefährtin Anja Masur, welche mich persönlich und fachlich stark unterstützt hat.

Eigenständigkeitserklärung

Ich versichere, diese Master Thesis ohne fremde Hilfe und ohne Verwendung anderer als den angeführten Quellen angefertigt zu haben, und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat. Alle Ausführungen der Arbeit die wörtlich oder sinngemäß übernommen wurden sind entsprechend gekennzeichnet.

Wien, April 2010

A handwritten signature in black ink, appearing to read 'C. Ansorge', written in a cursive style.

Christian Ansorge

Kurzfassung/Abstract

Deutsche Version

Kurzfassung. Während die syntaktische Interoperabilität innerhalb von Komponenten der Geodateninfrastruktur durch die bisher veröffentlichten Standards der OGC weitgehend gewährleistet ist, wird die Rolle der Semantik und der mit ihr zusammenhängenden Konflikte in Geodateninfrastrukturen erst durch das Zusammenwachsen großer, weitgefächerter und internationaler Informationsnetzwerke deutlich. Semantische Annotationen können in Geodateninfrastrukturen auf externe Quellen von Informationen und Wissen verweisen. Durch ihren Einsatz ist es möglich, die semantische Interoperabilität deutlich zu steigern. Allerdings sind die bisherigen Implementierungsmöglichkeiten von semantischen Annotationen heterogen und nicht standardisiert. In dieser wissenschaftlichen Arbeit werden verschiedene Ansätze von semantischen Annotationen in unterschiedlichen Ebenen einer (SWE basierten) Geodateninfrastruktur demonstriert, diskutiert und bewertet. Dabei wurde das Ziel verfolgt, einen Überblick über derzeit verwendete und in der Literatur vorgeschlagene Implementierungsmöglichkeiten von semantischen Annotationen zu bieten.

Schlüsselwörter:

Semantische Annotation, Geosemantik, Semantic Web, Ontologie, Sensor Web, SWE

English Version

Abstract. At present, OGC Standards and specifications enable a certain degree of syntactic interoperability for geo-data infrastructures. In comparison, the importance of the semantic aspect of interoperability has been largely ignored, as it only becomes clearly visible with the advent of widespread international data infrastructures. Semantic Annotations enable references from the data to external knowledge resources like

ontologies, and can greatly improve usability of the data by explicitly specifying the meaning of individual data concepts. This is necessary to enable semantic interoperability of the data. However, the current implementations of semantic annotations are heterogeneous and not standardized. In this paper, we have collated, described and analyzed different approaches to semantic annotations currently in use, based on a hypothetical SWE based information infrastructure. This has been done with the goal of giving the reader an overview of existing solutions of semantic annotation of geospatial data, both in use and described in the literature

Keywords:

Semantic Annotation, Spatial Semantic, Semantic Web, Ontology, Sensor Web, SWE

Inhaltsverzeichnis

1	Einführung	VIII
1.1	Einführung und Problemstellung.....	1
1.2	Zielsetzungen.....	3
1.3	Thesen	4
1.4	Struktur der Thesis	5
2	Grundlagen und Literatur	7
2.1	Wissen	7
2.1.1	Implizites und explizites Wissen.....	9
2.2	Wissensmanagement	10
2.3	Semantik.....	11
2.4	Semantische Interoperabilität und semantische Lücke	13
2.5	Semantic Web.....	18
2.6	Semantische Technologien	20
2.6.1	Semantische Annotation.....	22
2.6.2	Wissensmodellierung und Wissensrepräsentation	22
2.7	Geosemantik und Interoperabilität in der vernetzten Geoinformatik.....	24
3	Ontologie	28
3.1	Systeme der Wissensrepräsentation.....	28
3.2	Thesaurus.....	31
3.3	Ontologie	34
3.3.1	Einsatzmöglichkeiten von Ontologien	35
3.3.2	Aufbau von Ontologien.....	37
3.3.3	Ontologiesprachen	41
3.3.4	Architektur von Ontologien	43
3.4	Beispielontologie.....	46
4	Semantische Annotation	49
4.1	Sensor Web Enablement.....	49

4.1.1	Observation and Measurement.....	52
4.1.2	Sensor Model Language.....	54
4.2	Ebenen und Anwendungen der semantischen Annotation	55
4.2.1	Suche nach geographischen Informationen und Diensten	56
4.2.2	Validierung und Konsistenzprüfung von geographischen Informationen.	57
4.2.3	Validierung von Prozessketten	58
4.2.4	Integration von geographischen Daten	61
4.2.5	Zusammenfassung der möglichen Ebenen der semantischen Annotation	62
4.3	Umsetzung von semantischen Annotationen.....	62
4.3.1	Semantische Annotation von Metadaten.....	63
4.3.2	Semantische Annotation von OGC Diensten	67
4.3.3	Semantische Annotation von Datenschemen	70
4.3.4	Semantische Annotation von Dateninstanzen	73
4.4	Bewertung der Annotationsmöglichkeiten und Thesen.....	79
5	Zusammenfassung, Kritik und Ausblick.....	82
6	Literatur	84
7	Anhang.....	87
7.1	Beispielontologie.....	87
7.2	Observation and Measurement.....	91
7.3	SensorML	94

Abbildungsverzeichnis

Abbildung 1: Wissenspyramide	8
Abbildung 2: Implizites und explizites Wissen.....	10
Abbildung 3: Semantic Web	19
Abbildung 4: Semantische Annotation und Wissensrepräsentation.....	21
Abbildung 5: Semantische Technologien nach Orbst	29
Abbildung 6: GEMET Beispiel für den Ausdruck „Water“	32
Abbildung 7: Auszug einer INSPIRE konformen Metadatenbeschreibung.....	33
Abbildung 8: Beispielontologie (Konzepte und Relationen).....	40
Abbildung 9: Beispielontologie (Instanzen und Relationen).....	40
Abbildung 10: Semantic Web Stack.....	41
Abbildung 11: Ontologie-Architektur	44
Abbildung 12: Zusammenhang Sensornetzwerk - Direktive	47
Abbildung 13: Beispielontologie – CAFE	48
Abbildung 14: Semantische Annotation in <i>OGC Observation and Measurement</i>	51
Abbildung 15: „Temperature“ innerhalb der NASA SWEET Ontologie.....	52
Abbildung 16: Observation-Objekt in Observation and Measurement	53
Abbildung 17: XML-Beispiel - die Semantik auf Datenebene.....	60
Abbildung 18: XML-Schemata-Beispiel – die Semantik auf Schemata-Ebene	61
Abbildung 19: Semantische Annotation von Metadaten – Schema	64
Abbildung 20: Semantische Annotation von Metadaten – INSPIRE Metadatenbeschreibung (GEMET)	65
Abbildung 21: MD_Keyword-Sektions der ISO19115.....	66
Abbildung 22: Semantische Annotation von OGC Diensten – Schema	68
Abbildung 23: <i>Identification Section als UML</i> (Whiteside, 2009).....	69
Abbildung 24: Semantische Annotation von OGC Diensten – Service Identification Section (XML)	70
Abbildung 25: Semantische Annotation von Datenschemen - Schema.....	71
Abbildung 26: Semantische Annotation von Datenschemen – Semantik auf Schemen- Ebene.....	72

Abbildung 27: Semantische Annotation von Datenschemen – Semantik auf Schemenebene.....	72
Abbildung 28: Semantische Annotation von Dateninstanzen – Schema	74
Abbildung 29: Semantische Annotation von Dateninstanzen – URN	75
Abbildung 30: Semantische Annotation von Dateninstanzen – XSD.....	76
Abbildung 31: Semantische Annotation von Dateninstanzen - XML.....	76
Abbildung 32: Semantische Annotation von Dateninstanzen – XLink in SensorML	78
Abbildung 33: Semantische Annotation von Dateninstanzen – O&M	78

1 Einführung

1.1 Einführung und Problemstellung

Seit der Etablierung des Internets sehen wir uns als Gesellschaft in den letzten Jahrzehnten mit einer immer stärkeren Vernetzung von Informationen konfrontiert. Diese technische Entwicklung hat unser Leben und unsere Kommunikation grundlegend verändert. Das Internet ist heute als Medium für den Austausch von Informationen kaum mehr wegzudenken und hat andere Formen der Kommunikation teilweise verdrängt und ersetzt. Dies gilt besonders für den Austausch von Daten. Denn diese werden heutzutage fast nur noch digital gespeichert und innerhalb von Informationsnetzwerken transportiert, was eine erhebliche Effizienzsteigerung bedeutet, da diese Informationen kostengünstig einer hohen Zahl an Nutzern zur Verfügung gestellt werden können. Daher stehen uns heute eine nicht überschaubare Anzahl an Quellen im Internet zur Verfügung, welche textuelle, bildliche und andere Informationen für uns bereitstellen. Das Internet hat sich während der letzten Jahre gewandelt und mit ihnen die Ansprüche ihrer Nutzer. Ging es zu den Anfängen des Webs 2.0 noch darum, Nutzer und ihre persönlichen Inhalte in das Internet einzubinden, so stehen wir heute vor der Herausforderung, diese uns zur Verfügung stehende Menge an Informationen effizient zu nutzen. Der ständig wachsenden Quantität der Daten steht mangelndes Wissen über die Strukturen und die Inhalte der Daten gegenüber. Eine der Gefahren in der Quantität der verfügbaren Daten und Informationen liegt darin, dass vieles an implizitem als auch explizitem Wissen verloren gehen kann, da Informationen nur noch über Schnittstellen und ohne menschlichen Kontakt ausgetauscht werden. Hochkomplexes Wissen und Modelle können bisher nur beschränkt automatisiert und digital ausgetauscht werden. Der quantitative Anstieg der Informationsquellen hat daher nicht zwangsläufig eine qualitative Verbesserung zur Folge. Unterschiedliche Ebenen der Abstraktion im Verständnis zwischen Nutzer und Daten erschweren die Suche und den Umgang mit der notwendigen Information. Komplexe Daten benötigen auch umfassende Beschreibungen, welche ihnen und den von Nutzern an sie gestellten

Ansprüchen gerecht werden können. Nur dadurch kann auch weiterhin die effiziente Nutzung und Kommunikation der abgebildeten Informationen gewährleistet werden.

Neben der reinen Qualität der Daten wird auch der Austausch zwischen den beteiligten Systemen immer wichtiger. Die Fähigkeit von Systemen, Daten – innerhalb von Informationsnetzwerken – untereinander verlustfrei auszutauschen und miteinander zu operieren, – die so genannte Interoperabilität – rückt immer stärker in den Fokus des Interesses.

Diese Entwicklung ist unter Anderem auch im Bereich der Geoinformation zu beobachten. Standen sich in den Anfängen der Geoinformatik noch monolithische Einzelsysteme gegenüber, bei denen der Datenaustausch in der Praxis schier unmöglich erschien, ist die Interoperabilität heutzutage ein Anspruch geworden, an dem sich Geoinformationssysteme messen lassen müssen. Durch die quantitative Verfügbarkeit von Informationen steigt auch der Wunsch diese zu nutzen, wodurch es zwangsläufig zu einem erhöhten Transfer von Daten kommt. Die angesprochene Nutzung setzt allerdings neben den notwendigen technischen Möglichkeiten auch Wissen und Informationen über die Daten selbst voraus. Mit der zunehmenden Verfügbarkeit großer Mengen an geografischen Daten steigt zudem auch die Frage nach der eigentlichen Bedeutung des Inhaltes besagter Informationen. Wie bei Daten mit nicht geographischem Bezug gilt auch hier, dass die angebotene Abstraktionsstufe nicht in jedem Fall ausreichend ist, um die Informationen sinnvoll verwerten und sie für die angestrebten Aufgaben einsetzen zu können. Die Bedeutung der Dateninhalte stellt im Kontext dieser Arbeit wiederum einen Teilaspekt des Datenaustausches dar.

Um diesen Austausch an Daten zu ermöglichen und einen kulturellen als auch finanziellen Mehrwert aus der Generierung neuer Daten zu erzielen, können wir eine immer stärker werdende Harmonisierung der am Informationsnetzwerk beteiligten Komponenten beobachten. Als eine der größten Verbesserungen der letzten Jahre auf dem Gebiet der Interoperabilität ist die Entwicklung verschiedener OGC-Standards¹ zu

1 Webservices wie z.B. WMS, WFS und GML wurden von der OGC (Open Geospatial Consortium, <http://www.opengeospatial.org>, 2009) spezifiziert.

nennen. Sie sind heute weit verbreitet und bilden die operative Basis vieler Geodateninfrastrukturen².

Die Bestrebungen, sowohl die angebotenen Datenstrukturen als auch die eingesetzten Technologien zu standardisieren, können weltweit beobachtet werden. Beispielhaft für den Bereich der Geoinformatik sollen hier INSPIRE³ und SEIS⁴ genannt werden, welche in der weiteren Arbeit eine Rolle spielen werden und von der Europäischen Kommission forciert werden. Ihre Entwicklung folgt der Vision von möglichst grenzenlosem, interoperablem Austausch von Informationen, welche nicht zwangsläufig einen eindeutig geographischen Bezug haben müssen. Als Medium der Datenübertragung dienen dafür unterschiedliche XML-Spezifikationen der bereits genannten OGC wie *GML* oder *Observation and Measurement*. Obwohl die jeweiligen verwendeten XML-Spezifikationen bereits ein gewisses Maß an informeller Ausdruckskraft aufweisen, welche Strukturen und Dateninhalte beschreibt, ergibt sich aus der ständig steigenden Zahl an Quellen für raumbezogene Informationen ein stärkerer Bedarf an Semantik als bisher in OGC-Standards berücksichtigt wurde. Semantische Technologie, wie etwa die Ontologie oder semantische Annotation, kann dabei in verschiedenen Ebenen der Geodateninfrastruktur eingesetzt werden und stellt eine Erweiterung der bisherigen Informationsinstanz dar. Mit dem verstärkten Einsatz von semantischer Technologie kann dem Verlust an Information und Wissen vorgebeugt werden und neben der syntaktischen auch die semantische Interoperabilität sichergestellt werden.

1.2 Zielsetzungen

² Geodateninfrastrukturen (engl. Spatial Data Infrastructure, SDI)

³ INSPIRE, Infrastructure for Spatial Information in the European Community (<http://inspire.jrc.ec.europa.eu/>), Koordination durch das Jointed Research Center (JRC)

⁴ SEIS, Shared Environmental Information System (<http://ec.europa.eu/environment/seis/>), Koordination durch die Europäische Umweltagentur (EEA)

Die Zielsetzung dieser wissenschaftlichen Arbeit ist es, Möglichkeiten der Implementierung von semantischen Verweisen und deren generelle Funktionsweise in bestehenden OGC XML-Spezifikationen aufzuzeigen. Semantische Verweise (semantische Annotationen) zeigen in der Regel auf externe Wissensrepräsentationen, welche zusätzliches Wissen über Ausdrücke und Methoden bieten, und damit eine Erweiterung der bestehenden Information darstellen. Derartige Konzepte bzw. konzeptuelle Modelle sind notwendig, um Daten und Informationen in einer vernetzten Infrastruktur vollständig interoperabel nutzen zu können. Die Umsetzung von semantischen Annotationen soll sich auf die *OGC Sensor Web Enablement* - Spezifikationen beziehen. Die Implementierung von semantischen Verweisen in sonstigen XML-Spezifikationen aus dem Bereich der Geoinformatik soll nur am Rande diskutiert werden.

Neben der rein technischen Umsetzung der semantischen Verweise sollen semantische Technologien wie etwa die Ontologie eingeführt und diskutiert werden, da diese eine zentrale Schlüsselrolle darstellt. Zudem soll eine eigene Ontologie zu Demonstrationszwecken entwickelt werden, an welcher die Funktionsweise von semantischen Annotationen demonstriert werden soll. Das Zielpublikum dieser Arbeit sind Experten, welche mit Informationsnetzwerken – insbesondere im Bereich der Geoinformatik – bereits vertraut sind, aber über geringe Kenntnis im Bereich der Semantik verfügen.

1.3 Thesen

Im Abschnitt Einführung und Problemstellung werden Probleme angerissen und beschrieben, welche ihren Ursprung in fehlender bzw. mangelnder semantischer Beschreibung von Informationen haben. Folgende Thesen stehen im Mittelpunkt dieser wissenschaftlichen Arbeit und im Zusammenhang mit den beschriebenen Zielen.

- Eine semantische Erweiterung und das Injizieren von XML mit Verweisen auf externe Ressourcen ist in OGC XML-Spezifikationen möglich

- Wissen, auf welchem Informationen (im XML-Format) implizit beruhen, kann mittels semantischen Technologien abgebildet werden
- Im Bereich der Geodateninfrastruktur sind weitere Ansätze zur semantischen Erweiterung bestehender Informationsstrukturen und Konzepte notwendig
- Mit den Verweisen auf externe semantische Ressourcen können Konzepte für verschiedene Fach- und Wissenschaftsbereiche angesprochen werden

1.4 Struktur der Thesis

In diesem Abschnitt soll auf die Struktur der Thesis eingegangen werden, welche auf den Zielen der Arbeit und der methodischen Herangehensweise aufbaut.

Diese wissenschaftliche Arbeit ist in drei Abschnitte gegliedert. Diese drei Bereiche stehen miteinander im Zusammenhang und bauen aufeinander auf.

Der erste Bereich der Thesis umfasst die Grundlagen und Auseinandersetzung mit der zur Verfügung stehenden Literatur. In diesem Abschnitt soll eine Einführung in die verwendeten Begrifflichkeiten und Technologien gegeben werden. Zudem befasst sich dieser Bereich mit dem der Arbeit zugrunde liegenden Problem der semantischen Interoperabilität.

Der zweite Bereich widmet sich der Ontologie als eine Möglichkeit der Wissensrepräsentation und semantischen Technologie. Hierbei soll neben der technischen Umsetzung auch auf die Rolle von Ontologien in Informationsnetzwerken eingegangen werden. Dieses Kapitel soll zudem dem besseren Verständnis der Anforderungen und Möglichkeiten im Umgang mit semantischen Technologien dienen. Abschließend soll eine Applikationsontologie zu Demonstrationszwecken vorgestellt werden.

Der dritte Bereich befasst sich mit der konkreten Umsetzung der Verweise auf externe semantische Quellen in XML – der semantischen Annotation. Hierbei sollen verschiedene Methoden, solche Verweise in XML zu implementieren, demonstriert und diskutiert werden. Ein weiterer wichtiger Punkt dieses Abschnittes bildet die Analyse, in

welcher Rolle und in welchen Bereichen semantische Verweise in Geodateninfrastrukturen sinnvoll eingesetzt werden können. Zudem werden die Spezifikationen des OGC *Sensor Web Enablement* diskutiert, auf denen die Mehrzahl der verwendeten Beispiele beruht. Darauf aufbauend sollen abschließend Schlussfolgerungen bezüglich weiterer Schritte zur Einführung semantischer Technologien in operativen Geodateninfrastrukturen gezogen werden.

2 Grundlagen und Literatur

In diesem Kapitel soll eine Einführung in die Begriffe und Grundlagen erfolgen, welche für die Beantwortung der Fragestellung dieser Thesis von Relevanz sind. Da die angeschnittenen Themenfelder weit gefächert und inhaltlich jeweils komplex sind, kann diese Thesis nur einen, in Umfang und inhaltlicher Tiefe, beschränkten Einblick bieten. Neben der inhaltlichen Beschreibung der Begriffe und Technologien wird innerhalb der einzelnen Kapitel versucht, sich mit der zur Verfügung stehenden Literatur auseinander zu setzen.

2.1 Wissen

Es ist notwendig, an dieser Stelle den Begriff des Wissens genauer zu spezifizieren und zu erläutern.

Der Begriff des „Wissens“ wird von zahlreichen klassischen Disziplinen der Geisteswissenschaft wie der Philosophie, Linguistik als auch von Naturwissenschaften sowie der Psychologie behandelt. Es existieren hierbei allerdings deutliche Unterschiede in Deutung und Definition, welche sich im Wesentlichen auf unterschiedliche Sichtweisen und Standpunkte zurückführen lassen. Daher existiert bis heute keine allgemein gültige und anerkannte Definition des Begriffs „Wissen“. (U. Roumois, 2007)

Im Kontext dieser wissenschaftlichen Arbeit und des Wissensbegriffes stehen vor allem der Ansatz sowie die Methoden des Wissensmanagement und der Semantik. Daher soll hierbei ein gängiger Ansatz aus dem Bereich des Wissensmanagement wiedergegeben werden, welcher versucht den Begriff des „Wissens“ zu definieren. Dieser Ansatz stellt den Begriff und die Bedeutung von Wissen in den Kontext zu seiner Nutzung. Wissen dient dabei in erster Linie als Voraussetzung für richtiges Handeln. In diesem Kontext wird zwischen Wissen, Informationen und Daten unterschieden. Diese Unterscheidung wurde im Zuge der Betrachtung von Wissen durch die Informationswissenschaften notwendig, welche versucht Wissen zu strukturieren, zu externalisieren und zu

transferieren. Basierend auf der Unterscheidung von Wissen, Information und Daten entsteht Wissen daher als Produkt eines Lernprozesses, indem er Informationen mit bereits bestehendem Wissen verknüpft. (U. Roumois, 2007)

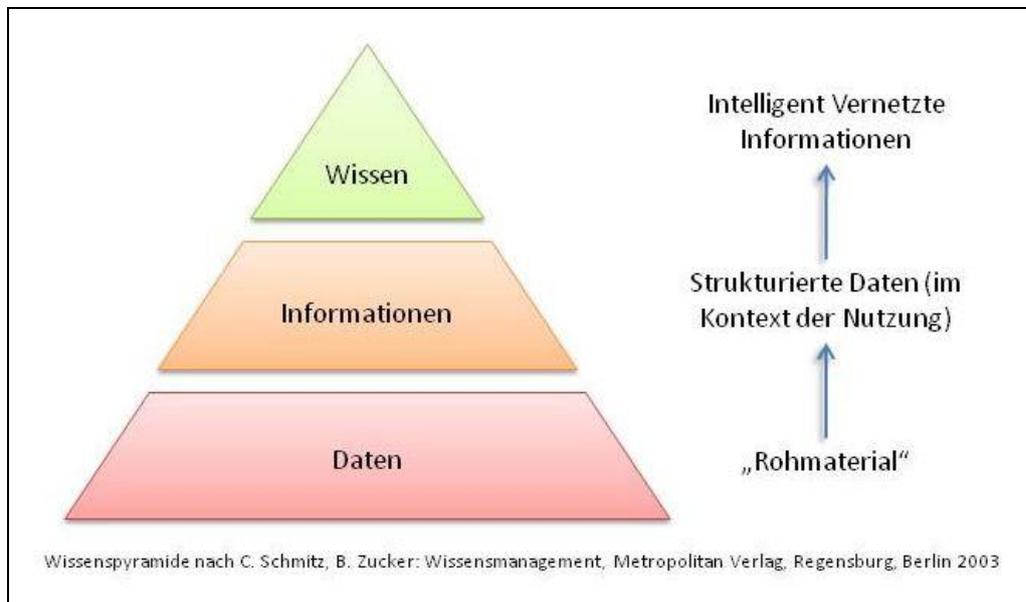


Abbildung 1: Wissenspyramide

Daten können daher materiell (zum Beispiel als Zeichenfolge) wahrgenommen und gespeichert werden. In einem bestimmten (semantischen) Kontext können Daten eine Bedeutung erhalten, was sie zu Informationen aufwertet. Dies setzt eine Relevanz und Nachfrage seitens des Nutzers voraus.

Informationen wiederum stellen eine weitere dynamische und immaterielle Qualitätsstufe von Daten dar. Ob Daten wiederum als Informationen wahrgenommen werden, hängt vom Empfänger ab. Im Prozess des Wahrnehmens von Daten können diese interpretiert sowie strukturiert und als Informationen weiter verarbeitet werden.

Ist eine wahrgenommene Information für einen Empfänger relevant, kann dieser sie mit bereits vorhandenem Wissen verknüpfen und daraus neues Wissen generieren. Diese Verknüpfung und Vernetzung von Informationen und bestehendem Wissen kann als Prozess verstanden werden.

Auf dieser Gliederung aufbauend kann Wissen als immateriell, nicht greifbar sowie subjektiv bezeichnet werden. Darüber hinaus existiert Wissen nur im Kopf des

Menschen. Wissen kann allerdings wiederum kommuniziert werden, wodurch es erneut den Prozess der Daten, Informationen und schlussendlich des Wissens durchläuft. (U. Roumois, 2007)

2.1.1 Implizites und explizites Wissen

Für die weiteren Ausführungen und Erklärungen ist es vor allem wichtig, den Unterschied zwischen explizitem und implizitem Wissen zu schärfen. Basierend auf der Trennung von Daten, Informationen und Wissen kann letzteres auf verschiedene Weisen gegliedert und erneut unterteilt werden. Diese Unterteilungen sind oft fachspezifisch und nur im Kontext ihres Betrachtungsstandpunktes anwendbar.

Wichtig ist in diesem Zusammenhang vor allem die Unterscheidung der zwei Arten von Wissen, dem **expliziten** und **impliziten Wissen**, welche besonders in dem Bereich der Informationswissenschaften eine große Rolle spielen. Diese Unterscheidung wurde von Michael Polanyi 1966⁵ erstmals getroffen. Diese beiden Arten des Wissens lassen sich auch als transferierbares und nicht transferierbares Wissen bezeichnen. Diese Trennung ist im Zusammenhang der Thesis wichtig, da die Existenz dieser beiden Arten des vorhandenen Wissens und ihrer unterschiedlichen Anforderungen an die Kommunikation zur semantischen Lücke und zum Verlust an Semantik in Informationsstrukturen führen kann. (U. Roumois 2007)

Explizites Wissen lässt sich strukturieren, transferieren und artikulieren. Der Inhaber dieses Wissens ist sich dessen bewusst und kann es externalisieren. Ein Beispiel hierfür ist eine wissenschaftliche Arbeit, welche einen Sachverhalt sehr genau und ausführlich beschreibt. Explizites Wissen kann zudem identisch mit implizitem Wissen sein. Die Unterscheidung findet hierbei durch die Repräsentation und den Ort der Speicherung statt. (U. Roumois, 2007)

Implizites Wissen hingegen ist nur schwer zu artikulieren oder zu transferieren, ist in der Regel hoch komplex und von einer geringen zeitlichen Dauer. Als Beispiel für

⁵ Vgl. dazu Michael Polanyi, 1966

implizites Wissen kann man das Wissen eines Experten über einen hochkomplexen und individuellen Prozess (wie zum Beispiel die Herstellung von Wein) nennen, welcher nur schwer beschreibbar ist. (Mertens, Bodendorf et al., 2005) Implizites Wissen bildet sich im Kopf eines Menschen. Erst in dem Augenblick, in welchem das implizite Wissen formalisiert wird und dadurch transferierbar sowie kommunizierbar wird, findet die Umwandlung des externalisierten Wissens in explizites Wissen statt. (U. Roumois, 2007) Die Handhabung und der Umgang mit implizitem Wissen gewinnt zunehmend an Bedeutung, da sich zeigt, dass ein erheblicher Anteil des vorhandenen Wissen noch nicht explizit vorliegt.

Wissen	
Implizites Wissen	Explizites Wissen
Das persönliche Wissen eines Mitarbeiters über einen komplexen Aggregationsprozess von Geodaten	Die formale Spezifikation (der Mitarbeiter ist einer der Autoren), welche den komplexen Aggregationsprozess beschreibt und ihn inhaltlich für andere nachvollziehbar macht

Abbildung 2: Implizites und explizites Wissen

Die Unterscheidung dieser beiden Wissensarten ist im Kontext dieser Arbeit wichtig, da es eine der Bestrebungen dieser Thesis ist, implizites Wissen zu externalisieren und Daten durch Anreicherung von Kontext zu Informationen zu veredeln.

2.2 Wissensmanagement

Das Wissensmanagement in seiner vollen Komplexität soll hierbei bewusst nicht näher erläutert werden, da es nur als Einführung und Rahmen für weitere verwandte Begriffe und Technologien dient.

Das „Wissensmanagement“ dient heute als Sammelbegriff für Methoden und Prozesse, welche sich mit dem Umgang mit Wissen befassen. In diesem Zusammenhang wird Wissen vor allem als eine Ressource verstanden, welche betriebswirtschaftlich sinnvoll genutzt werden muss. Nach Roumois (2007) leitet sich der Bedarf von Wissensmanagement aus folgenden Tatsachen ab:

- Zunehmende Digitalisierung und rasantes Wachstum des globalen Wissens bzw. der global verfügbaren Informationen
- Entkopplung von Wissen und Produktion
- Entmaterialisierung der Wertschöpfung (Wert ist nicht länger auf Materielles beschränkt)
- Überangebot an Information führt zum Mangel an Wissen
- Der Wert des Wissens wird auf seine marktwirtschaftliche Bedeutung reduziert

Bezüglich der Definition des Begriffes „Wissensmanagement“ existieren innerhalb der Literatur verschiedene Ansichten⁶. Allgemein lässt sich allerdings aussagen, dass Wissensmanagement bedeutet, sich bewusst mit Wissen sowie den damit in Zusammenhang stehenden Anforderungen, Methoden und Technologien auseinanderzusetzen und ihnen einen entsprechenden Stellenwert einzuräumen. (Franken, Gadatsch, 2002)

Da Wissen und Informationen immer stärker an Bedeutung gewinnen, wird auch das Wissensmanagement heutzutage in vielen wissenschaftlichen Disziplinen, wie den Wirtschaftswissenschaften, der Informatik oder den Sozialwissenschaften bewusst eingesetzt. Neben dem Einsatz in verschiedenen wissenschaftlichen Fachbereichen spielt Wissensmanagement vor allem im Bereich der privaten Wirtschaft eine zunehmend wichtigere Rolle, um Informationen und Wissen in den Unternehmen sinnvoll und effizient nutzen zu können. Dies resultiert daraus, dass Wissen zunehmend als wichtiger Faktor und Machtpotential der Unternehmen angesehen wird, welches generiert und genutzt werden muss.

2.3 Semantik

Semantik ist ein Teilgebiet der Linguistik, der Sprachwissenschaft, und befasst sich hauptsächlich mit dem Sinn und der eigentlichen Bedeutung der Sprache. Ein Synonym

⁶ R. Franken, A. Gadatsch (2002), S. 4 f.

für die Semantik ist das Wort „Bedeutungslehre“. Ähnlich wie die zuvor kurz umrissene Fachrichtung des Wissensmanagement findet die Semantik in vielen wissenschaftlichen Disziplinen Anwendung. Dabei sind vor allem die Informationstheorie, soziologische Systemtheorie, Sprachphilosophie und visuelle Kommunikation zu nennen.

Im thematischen Zusammenhang ist besonders die Bedeutung der Semantik in der Informationswissenschaft wichtig. Dabei ist deutlich zwischen der formalen Semantik und der linguistischen Semantik zu unterscheiden. Die formale Semantik befasst sich mit der Komposition der Satzbedeutung und bedient sich dabei mathematischer und logischer Darstellungsmittel. Abhängig vom beleuchteten Teilaspekt wird dieser Teil der Semantik auch *modelltheoretische Semantik*, *wahrheitskonditionale Semantik*, *referenzielle Semantik*, *logische Semantik* oder *Mögliche-Welten-Semantik* bezeichnet. (Löbner, 2003)

Im Fokus der semantischen Interoperabilität und der semantischen Lücke stehend, bezieht sich diese wissenschaftliche Arbeit jedoch stärker auf die linguistische Semantik, daher auf die Bedeutung von eigentliche Ausdrücken sowie deren Abbildung in Modellen und Kommunikation. Im Kontext der linguistischen Semantik ist unter Anderem die Unterscheidung zwischen **Syntax** und **Semantik** in Bezug auf Daten wichtig. Ebenso soll auf die Unterscheidung zwischen dem **Sinn** und der **Bedeutung** eines Ausdrucks eingegangen werden, was gleichzeitig die Kernaufgabe der Semantik ist. Diese Begriffe spielen eine zentrale Bedeutung bei der Interoperabilität und dem Semantic Web, welche darauf aufbauend beschrieben wird.

Syntax bezieht sich auf die formale normative Struktur, in der Daten abgelegt werden. Die Syntax innerhalb der Linguistik beschreibt Regeln, durch welche Wörter in Wortgruppen, Sätzen und Phrasen zusammengesetzt werden. Man kann sie in diesem Kontext auch als Grammatik verstehen. In der Informatik bezeichnet die Syntax eine Menge an Regeln, welche den Aufbau und die Struktur von Dokumenten (wie XML-Dokumenten) oder Programmen (etwa durch Programmiersprachen) reglementieren. (P. Hitzler, 2008) In diesem Zusammenhang steht auch der Sinn eines Wortes bzw. eines Ausdrucks. Dieser gibt den Inhalt, welcher sich aus der internen Relation der Wörter und Zeichen innerhalb der Sprache ergibt, an. Daher ist der Sinn eines Ausdrucks sowohl mit seiner Syntax als auch mit der Semantik verbunden.

Semantik steht wie bereits beschrieben für die Bedeutung und die dadurch abgebildete Information durch Sprachen, Ausdrücke, Wörter oder Phrasen. Die Bedeutung bezieht sich auf einen Zusammenhang zwischen den Ausdrücken und referenziert auf ein externes reales oder fiktives Konzept.

Bezogen auf das Beispiel „*Ich mag Spaghetti!*“ bedeutet die Syntax dieses Ausdrucks den internen Zusammenhang zwischen Subjekt, Verb und Objekt. Die Semantik dieses Ausdrucks drückt hingegen aus, dass der Autor eine Vorliebe für eine spezielle Form italienischer Teigwaren hat.

Semantik ist für alle weiteren Begrifflichkeiten von enormer Bedeutung, da Semantic Web und die damit verbundene semantische Interoperabilität Probleme adressieren, welche im Bereich der Geoinformatik - gemessen an ihren Auswirkungen - bisher im zu geringen Ausmaß diskutiert wurden. An der Stelle, an welcher wir mit immer mehr Informationen konfrontiert sind, wird die Bedeutung der Information und ihre effiziente Verarbeitung und Nutzung zunehmend wichtiger.

2.4 Semantische Interoperabilität und semantische Lücke

Im folgenden Abschnitt wird die semantische Interoperabilität erläutert, wie auch einige Probleme, welche sich durch ihr Fehlen ergeben können.

Grundsätzlich muss hierbei festgehalten werden, dass Interoperabilität in der Kommunikation zwischen zwei Agenten⁷ auftritt. Dabei ist wie bereits erwähnt zwischen der syntaktischen und der semantischen Interoperabilität zu unterscheiden. Interoperabilität in der Informationstechnologie bezeichnet die Fähigkeit von Softwareelementen, Daten und Informationen untereinander auszutauschen und diese zu

⁷ Agenten stehen hierbei für (engl.) Agent, welche im Zusammenhang mit dem Semantic Web verwendet wird und sowohl für Systeme, Maschinen als auch Menschen stehen kann. Allerdings wird, besonders im Zusammenhang mit dem Semantic Web, „Agent“ als selbstständig operierende (IT-) Komponente angesehen.

interpretieren. Dies bezieht sich sowohl auf die syntaktische als auch die semantische Ebene der Kommunikation zwischen den Komponenten (R. Heutschi, 2007). Syntaktische Interoperabilität bezeichnet die Kenntnis des gegenseitigen Schnittstellenformates und gilt als Voraussetzung für Interoperabilität im Allgemeinen. Semantische Interoperabilität jedoch bezeichnet die Kenntnis der Bedeutung der Schnittstellen und Datenformate, um diese im entsprechenden Kontext sinnvoll interpretieren und nutzen zu können. Das Wissen der Bedeutung der Dateninhalte und Strukturen ist besonders wichtig, da wir uns heute mit rasant wachsenden Informationsnetzwerken und ständig steigendem Datentransfer konfrontiert sehen. Zu den hauptsächlichen Problemen von verteilten heterogenen Informationsnetzwerken mit hohem Anteil an Datenaustausch gehört neben der redundanten Datenhaltung auch der Semantikverlust (L. Behrend & H. Pundt, 1998).

Wie bereits erwähnt, ist die Quantität und Heterogenität der Information eines der gegenwärtig größten Probleme des Internets. Interoperabilität ist bereits seit langem als wichtiges Thema erkannt worden und in einigen technischen Bereichen existieren bereits viel versprechende Ansätze zur Realisierung. XML basierte Standards zum Datenaustausch haben bereits zu einem hohen Grad an syntaktischer Interoperabilität geführt. Die semantische Interoperabilität, welche auf der syntaktischen aufbaut, wurde bisher jedoch weitgehend vernachlässigt. Es fehlen etablierte Technologien und Standards für die Suche, Verknüpfung und Bewertung von Informationen in Bezug auf ihre Bedeutung.

Ein einfaches Beispiel soll hierbei die Bedeutung der Semantik für vernetzte Informationsnetzwerke verdeutlichen. Auf eine Suchanfrage bezüglich des Terms „*Mine*“ müsste man im Sinne der Semantik zuerst spezifizieren, ob ich mich für Kriegswaffen, Bergbau oder Kugelschreiber interessiere, da das Wort „*Mine*“ jeweils verschiedene Bedeutungen in den unterschiedlichen Bereichen besitzt. Der Mensch kann anhand des Inhaltes der Antworten erkennen, auf welche Bedeutung des Terms „*Mine*“ sich die Antwort bezieht. Die Maschine kann den Kontext und Zusammenhang der Daten, welche als Suchantwort zurückgegeben werden, jedoch nicht interpretieren und daher keinen Mehrwert aus der Datenmenge ableiten. Der Computer kann die gefundene Information zwar wiedergeben, allerdings nicht verstehen bzw. auf ihre Aussagekraft hin bewerten. Daher wird die eigentliche Interpretation der Daten bzw. Informationen nur vom menschlichen Nutzer durchgeführt. Folgend der Idee und den

Zielen des Semantic Web wäre der Computer durch Einsatz von semantischen Technologien in der Lage, die Konzepte, auf welche sich die möglichen Antworten beziehen, zu erkennen und zu unterscheiden. Das Fehlen dieses zweiten (semantischen) Aspektes der Interoperabilität wird auch als **semantische Lücke** bezeichnet. Daher kommt es zu einem Verlust an Semantik und einer Diskrepanz zwischen der Bedeutung eines Sachverhaltes in einer natürlichen Sprache und deren Abbildung innerhalb von Informationsnetzwerken. Die semantische Lücke ist ein Begriff, welcher hauptsächlich durch die Informationswissenschaften geprägt ist und in welcher er mit zunehmendem Zusammenwachsen der nationalen und internationalen Informationsnetzwerke an Bedeutung gewinnt.

Im Zusammenhang mit der Informationswissenschaft und Semantik können wir drei grundlegende Konflikte, welche auf der fehlenden semantischen Interoperabilität beruhen, unterscheiden, welche durch Patrick Maué (2009) wie folgt beschrieben werden:⁸

- Applikationsspezifisches Wissen
- Hierarchische Probleme
- Multilinguale Probleme

Applikationsspezifisches Wissen kann Konflikte erzeugen, die durch fehlendes spezifisches Wissen über die Software entstehen. Konflikte dieser Art entstehen oft in der Phase der Entwicklung und späteren Anpassung der Applikation und sind oft nur durch persönliches (implizites) Wissen zu lösen. Ein passendes Beispiel hierfür ist z.B. eine ungeschickte Abkürzung („Opreis“ als Abkürzung für „Originalpreis“) eines XML-Elementes, welche innerhalb des Entwicklungsteams zwar bekannt, aber aus bestimmten Gründen nie bereinigt, sondern übernommen wurde. Damit erschließt sich die Bedeutung des Elements nicht mehr automatisch (weder dem Menschen und noch weniger der Maschine) und die Interpretation der Daten bleibt einer kleinen Gruppe an Nutzern (welche über genau dieses Wissen verfügen) vorbehalten.

Hierarchische Probleme sind einige der am häufigsten auftretenden Probleme semantischen Ursprungs. Hierbei handelt es sich um die individuellen fachlichen

⁸ P. Maué, 2009

Unterschiede zwischen Datenanbietern und Datennutzern. Es kommt häufig zu hierarchischen Problemen, da davon ausgegangen werden kann, dass Datenanbieter ein entsprechend höheres fachliches Niveau (und Wortschatz) besitzen als Datensuchende. Während der normale Nutzer bspw. den Term „Steinzeit“ benutzt sind die für ihn relevanten Daten möglicherweise unter den Begriffen „Neolithikum“ oder „Jungsteinzeit“ („Neolithikum“ und „Jungsteinzeit“ stellen wiederum Synonyme dar) zu finden.

Multilinguale Probleme stehen besonders im vielsprachigen Europa im Vordergrund, während sie in den homogen englischen Sprachräumen eher in den Hintergrund treten. Multilinguale Probleme treten auf, wenn verschiedene Systeme und Datenformate auf unterschiedlichen Sprachen beruhen. Dies betrifft sowohl die Struktur bzw. das Schema der Daten als auch die Dateninstanzen selbst. Ein einfaches Beispiel hierfür ist der Term „Wasser“, welcher je nach Sprache „Vatten“, „Water“ oder „Aqua“ entspricht. Auch die Verwendung des Ausdrucks „H₂O“ oder „H₂O“ für Wasser fallen in die Gruppe der Multilingualität, da diese sich nicht nur auf natürliche Sprachen (wie Nationalsprachen) bezieht.

Zusammenfassend lässt sich feststellen, dass semantische Probleme aus der Vielfalt der Sprache resultieren. Neben diesen drei genannten semantischen Problemen treten weitere semantische Sonderfälle auf, welche für die semantische Interoperabilität von Bedeutung sind. Patrick Maué⁹ sieht im Rahmen der Geoinformatik folgende semantische Phänomene als potentielle Konflikte:

- Synonym
- Homonym
- Polysem
- Toponym

Ein **Synonym** bezieht sich auf ein anderes Wort, welches dieselbe Bedeutung hat wie das Synonym. Synonymie wird auch als Bedeutungsgleichheit bezeichnet. Als Beispiel wäre hier der schon genannte Ausdruck „Wasser“ zu nennen, dessen Synonym unter Anderem „H₂O“ oder „H₂O“ wäre. Alle Ausdrücke unterscheiden sich in Inhalt und

⁹ P. Maué, 2009

Umfang, haben aber dieselbe Bedeutung. Ein weiteres Beispiel ist das Wort „Jungsteinzeit“ welches als Synonym unter Anderem durch das Wort „Neolithikum“ bezeichnet werden kann. Durch die Verwendung einer semantisch unterstützten Suche würde der Nutzer nach dem eigentlichen Wort, aber auch nach seinen Synonymen suchen können und dadurch aussagekräftigere Ergebnisse erhalten.

Ein **Homonym** bezeichnet ein Wort, welches bei gleicher Schreibweise verschiedene Bedeutungen haben kann. Diese Bedeutungen beziehen sich im Falle des Homonyms auf unterschiedliche und nicht von einander abgeleitete Konzepte. Ein typischer Vertreter eines Homonyms ist das bereits genannte Wort „Mine“, welches Kriegswaffen oder Bergbaustollen bezeichnen kann. Bei diesem Beispiel würde eine herkömmliche, schlüsselwortbasierte Suche wesentlich mehr Ergebnisse aus gänzlich verschiedenen Bereichen zurückgeben als den Suchenden tatsächlich interessieren.

Neben dem Homonym ist auch das Phänomen der **Polysemie** für den Bereich der Geoinformatik relevant. Ein Polysem ist eng mit dem Homonym verwandt. Im Gegensatz zum Homonym leiten sich hier die unterschiedlichen Bedeutungen des identischen Wortes von einander ab. Dabei entstammen die unterschiedlichen Bedeutungen demselben Konzept. Ein Beispiel für Polysemie ist das „Schloss“, welches sowohl das Bauwerk als auch die Schließvorrichtung (an der Tür) bezeichnet. Beide unterschiedlichen Bedeutungen entstammen dem Konzept „schließen“ bzw. „schützen“ und sind von diesem abgeleitet.

Ein **Toponym**¹⁰ ist in diesem Kontext kein direkter semantischer Konflikt, da es den Namen einer geographischen Lokalität ausdrückt (Städte, Länder, Gebirge, etc.). Ein Toponym kann aber in vielen Fällen ein Polysem sein, da Ortsnamen oft nicht eindeutig sind. So ist die Bezeichnung „Altstadt“ als Suchbegriff nicht eindeutig, da dieses Wort als Toponym mehrfach auftreten wird und sich jeweils auf eine andere Stadt beziehen wird. Straßennamen, Stadtnamen und Flussnamen zählen ebenso zu den Polysemen. Der Ausdruck „Reichenbach“ bezeichnet beispielsweise sowohl verschiedene Städte

¹⁰ Toponym, zu Deutsch “Ortsname”

und Gemeinden als auch Ortsteile, Bauwerke und Gewässer.¹¹ Weiterhin kann ein Toponym auch ein Synonym sein.

Das Verständnis für diese Probleme, die in der Kommunikation und beim Transfer von Wissen und Informationen auftreten, ist wichtig. Basierend auf diesen, als semantische Lücke bezeichneten Diskrepanzen wurde die Vision des semantischen Netzwerks¹² entwickelt, welche im folgenden Kapitel näher erläutert wird.

2.5 Semantic Web

Das Semantic Web ist für diese Arbeit von besonderer Bedeutung, da es zum einen die mangelnde semantische Interoperabilität zwischen Agenten thematisiert und eine Vision von grenzenloser Kommunikation zwischen Agenten propagiert. Die semantische Annotation stellt zudem eine Schlüsseltechnologie des Semantic Webs dar, weswegen selbiges im Rahmen dieser Arbeit diskutiert wird.

Die Idee des Semantic Web beruht auf einem Artikel von Tim Berners-Lee, dem Begründer des World Wide Web, aus dem Jahre 2001.¹³ In diesem Artikel beschreibt Berners-Lee seine Vision der interoperablen Kommunikation zwischen Agenten. Diese Agenten, welche miteinander vernetzt sind, tauschen untereinander Informationen aus, bewerten und interpretieren diese und treffen Entscheidungen. All dies erfordert eine neue Art der Interoperabilität, wie sie im propagierten Web 2.0 nicht zu finden ist.¹⁴ Die Intention und die zugrunde liegende Idee ist es, Inhalte und Bedeutungen von Wörtern und Sätzen auch Maschinen und Computern zugänglich zu machen. Eine wesentliche Schwäche des damaligen Internets wurde dahingehend identifiziert, dass die Bedeutung

¹¹ Wikipedia, Suchbegriff "Reichenbach"

¹² Semantisches Netz, (engl. Semantic Web)

¹³ Tim Berners-Lee, James Hendler, Ora Lassila: The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In: Scientific American, 284 (5), S. 34–43, May 2001

¹⁴ Diese Aussage bezieht sich auf den Stand des Erscheinens des Artikels „The Semantic Web“ (2001)

der Inhalte ausschließlich auf menschliche Interpretation und menschliches Verständnis ausgelegt war. Während sich das World Wide Web vornehmlich mit der Verbreitung von Informationen befasst, werden durch das Semantic Web Möglichkeiten und Technologien aufgezeigt, welche die Verknüpfung der Bedeutung der Ausdrücke und Informationen ermöglichen. Das Semantic Web ist daher ein Konzept beziehungsweise eine Idee, welche auf Technologien der Semantik und des Wissensmanagement beruht. Diese Technologien und Konzepte werden jedoch bis jetzt nur zögerlich in Projekten eingesetzt.¹⁵

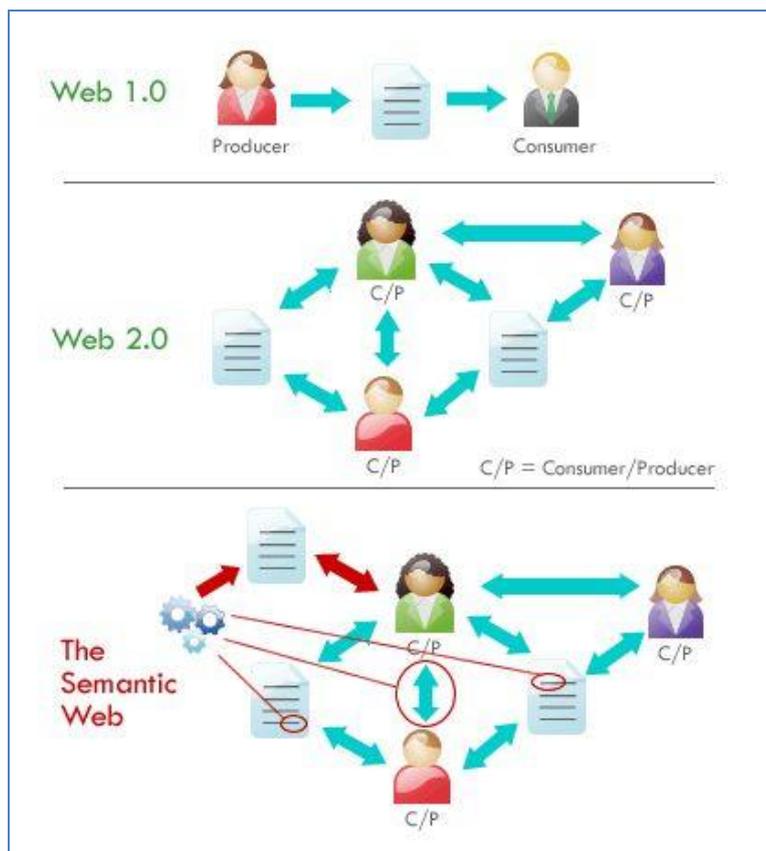


Abbildung 3: Semantic Web (<http://blogs.nesta.org.uk/innovation/2007/07/the-future-is-s.html>, 2009)

¹⁵ Zum Beispiel:

Theseus, Projekt zum Semantic Web (<http://www.theseus-programm.de/home/default.aspx>, 2009)

Swoogle, Semantische Suchmaschine (<http://swoogle.umbc.edu>, 2009)

Semantic Gov, Semantic Web im Bereich EGovernment (<http://www.semantic-gov.org>, 2009)

Das Semantic Web wird auch als Web 3.0 bezeichnet, da es die nächste technische Evolutionsstufe und eine Erweiterung für das Internet bedeutet. Das bisherige Web 2.0 ist an einem Punkt angekommen, an welchem einer quantitativ wachsenden Informationsfülle eine ineffektive Nutzung der Inhalte, welche aufgrund der Quantität zunehmend automatisiert betrieben werden, gegenübersteht. Tim Berners-Lee erkannte bereits 2001 in seinem zuvor genannten Beitrag die Grenzen des Web 2.0 sowie die Möglichkeiten und das Potential, welches in der sinnvollen Nutzung der Daten – nicht nur ihrer Verbreitung - liegen. Unter dem allgemein üblichen Begriff *Web 2.0* versteht man unter Anderem eine stärkere Beteiligung der Nutzer an der Bereitstellung von Daten und Informationen auf verschiedenen Plattformen des Internets.¹⁶ Der Begriff Web 2.0 steht daher weniger für Ziele und Technologien als für das Verständnis und die breite Akzeptanz des Internets in unserer Gesellschaft.¹⁷ Das Semantic Web versteht sich als Erweiterung der bereits vorhandenen Inhalte. (J. Behrendt, K. Zeppenfeld, 2008)

Das Semantic Web baut auf den Errungenschaften des Internets innerhalb der letzten 10 Jahre auf, welches die stärkere semantische Verknüpfung der verbreiteten Informationen überhaupt erst notwendig gemacht hat. Es geht dem Semantic Web im Gegensatz zum Web 2.0 nicht um die weitere Einspeisung und Verbreitung von Daten und Informationen in das Internet, sondern vielmehr um deren effiziente Nutzung. Das Ziel des Semantic Web ist es daher, Daten und Informationen direkt und indirekt der maschinellen Auswertung zugänglich zu machen. Maschinen sollen mit angebotenen Informationen unabhängig operieren können.

2.6 Semantische Technologien

¹⁶ Zum Beispiel:

Flickr, (<http://www.flickr.com>, 2009)

Wikipedia (<http://de.wikipedia.org>, 2009)

¹⁷ Mit dem Begriff Web 2.0 eng verbunden sind Begriffe wie „Social Tagging“, „User Generated Content“ oder „Crowdsourcing“, welche prägend für diese Evolutionsstufe des Internets sind

Die im Rahmen des Semantic Web vorgestellten und angewandten Technologien sollen hier nur erwähnt werden, da sie im Abschnitt Das Ziel dieses Kapitels ist es, zu erläutern, welche Relevanz das Thema der Semantik und des Semantic Webs für die Geoinformatik hat.

Bereits ein Jahr nach dem Erscheinen von Semantic Web spricht Egenhofer in „*Toward the semantic geospatial web*“ die Probleme im Bereich der Geographie an. Egenhofer zieht Parallelen zwischen der Situation des Webs 2.0 und der immer stärker vernetzten Geoinformationslandschaft. So sind fehlende Interoperabilität und ein hoher Verlust an Semantik auch in der Geoinformatik zu finden. Durch die Spezifikationen und Standards des OGC konnte bereits ein gewisses Niveau der syntaktischen Interoperabilität erreicht werden, die semantische Interoperabilität wurde allerdings weitgehend vernachlässigt.

Egenhofer sieht den Bedarf nach mehr semantischer Unterstützung vor allem in zwei Bereichen. Zum einen im Bereich der Suche nach Daten (Discovery) und zum anderen in der interoperablen Zusammenarbeit von Operanten (Computing). Desweiteren bemerkt Egenhofer in seinem Artikel, dass das Semantic Web zwar einen generischen Rahmen für die Arbeit mit Ontologien und Verweisen bietet, dieser aber der Vielfalt und den Ansprüchen der Geoinformatik nicht gerecht wird.

Derzeitige Suchmechanismen basieren hauptsächlich auf Schlüsselwörtern oder Textvergleichen, was sich seit 2002 bis heute nicht wesentlich verändert hat. Im Bereich der Geoinformatik werden zudem räumliche Filter eingesetzt und die Suchfunktion um die räumliche Komponente erweitert. Diese Methoden sind nicht ausreichend, um das Potential auszuschöpfen und dem Nutzer sinnvolle und effektive Antworten auf seine Suchanfragen zu liefern. Hierbei besteht nach wie vor Bedarf an semantischer Unterstützung beim Vorgang der Suche nach Informationen und Informationsquellen. Wir verfügen heute mit verschiedenen Standards für die Beschreibung von Metadaten zwar über syntaktische Interoperabilität und können Metadaten sammeln und verarbeiten, jedoch unterscheiden sich die Metadaten durch unterschiedliche Terminologie und Sprachen. Wir können hierbei die bereits angesprochenen hierarchischen und multilingualen Probleme erkennen. (Klien, 2009 & Egenhofer, 2002)

Ein weiterer Punkt, an welchem Egenhofer semantische Unterstützung für notwendig hält, ist die interoperable Zusammenarbeit zwischen verschiedenen Operanten. Hierbei ist es das Ziel, Informationen effizient und nutzbringend weiterzugeben. Als Adressaten dieser Informationen sieht Egenhofer sowohl die Maschine, welche die Information verarbeitet, als auch den Menschen, welcher die Informationen verstehen muss. Diese Art der semantischen Interoperabilität ist das Hauptziel des Semantic Webs. Hier haben wir durch die bereits angesprochenen Spezifikationen und Standards des OGC ein gewisses Maß an syntaktischer Interoperabilität erreicht. Eine wirkliche Interoperabilität schließt hierbei jedoch die Kenntnis der Bedeutung der Daten ein. Obwohl Interoperabilität in der Geoinformatik als wichtiges Thema diskutiert wird und OGC Web Services allgemein als Lösung des Problems begriffen werden, bedeuten diese jedoch nur eine syntaktische Interoperabilität. Für eine Zusammenarbeit von Web Services oder Maschinen im Allgemeinen muss die Kenntnis der Bedeutung der jeweilig transferierten Daten vorhanden sein. Im Bereich der Geoinformatik und der damit zusammenhängenden Web Services finden häufig flache Datenstrukturen Einsatz. Dies bedeutet, dass man sich in der Übermittlung von Daten (Metadaten) einschränkt und wichtige Informationen nicht transferiert werden können, da diese zu komplex sind. In der Praxis beschränken wir uns daher häufig auf Abkürzungen (wie z.B. „bgld“ für Burgenland oder „uba“ für Umweltbundesamt), welche die Bedeutung der Information reduzieren und oft nur schwer erkennen lassen. Ist dies der Fall, sind wir von echter Interoperabilität noch weit entfernt (Klien, 2009 & Egenhofer, 2002).

Auch W. Kuhn sieht die Interoperabilität in der vernetzten Geoinformatik durch die Teilung in Daten und Methoden erschwert. Zudem betont Kuhn in seinem Artikel zum Thema „Geospatial Semantics“, dass Semantik an sich nichts Neues auf dem Gebiet der Geoinformatik darstellt. Vielmehr werden semantische Technologien bereits seit längerer Zeit benutzt, um Interoperabilität zu unterstützen bzw. diese zu gewährleisten. Kuhn sieht daher die Notwendigkeit eines „Semantic Reference Systems“ für geographische Daten, um eine grundlegende Interoperabilität zu gewährleisten. (Kuhn, 2005)

Dies stellt nur eine Auswahl an Problemen und Ideen dar, welche mit mangelnder Interoperabilität und dem Verlust an Information zusammenhängen. Dadurch wird vor

allem deutlich, dass auch der Sektor der Geoinformatik durch die zunehmende Vernetzung von Daten und Services mit den Problemen konfrontiert ist, welche zu den Visionen des Semantic Web führten. Geodaten rücken zunehmend in den Fokus der breiten Öffentlichkeit und können daher nicht getrennt vom Internet betrachtet werden. Die mangelnde Interoperabilität wird zudem durch den hohen Grad an Heterogenität (Daten, Methoden als auch Services) erschwert. Diese Probleme und der Bedarf nach mehr Homogenität und Interoperabilität wurden erkannt, auch wenn die zahlreichen Bestrebungen und Ansätze, im Gegensatz zum Semantic Web, bisher keine weite Verbreitung gefunden haben. Die semantische Erweiterung unserer bisherigen Informationsnetzwerke ist zwar keine alleinige Lösung, aber eine der Schlüsselkomponenten, um deren Potential und Möglichkeiten auch für die Geoinformatik nutzen zu können. Durch das OGC wurde im Rahmen des „*Geospatial Semantic Web Interoperability Experiment Report*“ untersucht, welche Schritte im Zusammenhang mit der Schaffung eines Geospatial Semantic Web (GSW) und damit der Verbesserung der semantischen Interoperabilität unternommen werden müssen. Zu den weiteren Aktivitäten des OGC, welchem in der Standardisierung der vernetzten Geoinformatik eine Schlüsselrolle zukommt, bezüglich der Semantik gehören u. A.:

- „*Geospatial Semantic Web Interoperability Experiment Report*“, 2005
- Untersuchung der notwendigen Schritte zur Schaffung eines Geospatial Semantic Web aufbauend auf den Konzepten und Methoden des Semantic Web
- „*Semantic Annotations in OGC Standards*“, 2009
- Untersuchung wie semantische Annotation in OGC Standards (z. B. WFS, GML, Metadaten) realisiert werden kann
- *SensorML* und *Observation & Measurement*
- Standards aus dem Bereich Sensor Web (*OGC Sensor Web Enablement*), welche ein hohes Maß an semantischer Ausdrucksstärke besitzen und erstmals Techniken wie Annotation und Ontologien seitens der OGC vorschlagen

Neben den Aktivitäten der OGC widmen sich weltweit zahlreiche Forschungsprojekte dem Thema der semantischen Interoperabilität auch auf dem Bereich der Geoinformatik. Geosemantik gewinnt proportional zur wachsenden Vernetzung von Geodaten durch das Internet an Bedeutung.

genauer beschrieben werden. Im Fokus der Ansätze und Technologien steht die Erweiterung des bestehenden Internets sowohl um eine Ebene der semantischen Beschreibung von Daten und Diensten als auch der Abbildung von Wissen und Konzepten, auf welche sich diese semantischen Metadaten beziehen. Damit erhebt diese kurze Einführung keinen Anspruch auf Vollständigkeit, sondern versucht vielmehr die eingesetzten Technologien in zwei grundlegende Bereiche einzuteilen und eine ungefähre Funktion zu vermitteln.

Die semantischen Technologien lassen sich in zwei grobe Bereiche unterteilen: die **Repräsentation von Wissen** in Modellen und das **Referenzieren** auf genannte Modelle. Die Abbildung 4: Semantische Annotation und Wissensrepräsentation verdeutlicht den Zusammenhang dieser beiden technischen Komponenten.

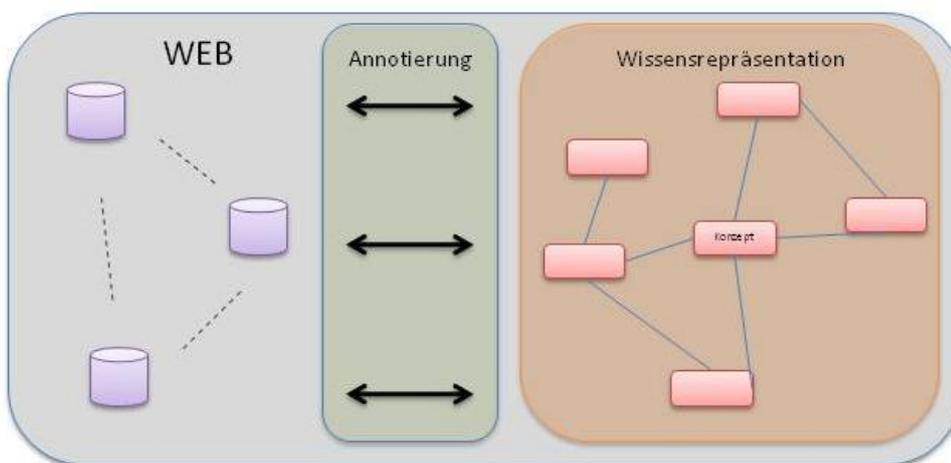


Abbildung 4: Semantische Annotation und Wissensrepräsentation

Um externes Wissen in Form von Modellen nutzen zu können, müssen Bezüge und Verweise auf diese Wissensressourcen hinweisen. Dadurch besteht zwischen diesen beiden technischen Aspekten eine Abhängigkeit. Allerdings müssen semantische Annotationen nicht zwangsläufig auf externe Ressourcen verweisen, sondern die Wissensmodellierung kann zum Beispiel bereits auf der jeweiligen Ebene der Annotation (z.B. durch URNs) vorgenommen werden. Die Etablierung von Modellen als externe Ressource stellt jedoch den klassischen Fall im Sinne des Semantic Webs dar.

2.6.1 Semantische Annotation

Grundlegend ist hierzu anzumerken, dass in der Englisch dominierten Fachliteratur der Begriff „**Annotation**“ verwendet wird. Dieser Begriff bezeichnet - im Kontext der semantischen Annotation - unter Anderem das Zufügen von Referenzen auf externe Quellen wie Ontologien oder Thesauri. Es wird in der weiteren Ausführung an dem Begriff „Annotation“ festgehalten.

Besonders die semantische Annotation steht im Fokus dieser wissenschaftlichen Arbeit und wird später im Kapitel

Semantische Annotation behandelt. Daher geht es in diesem kurzen Abschnitt nur um das Grundverständnis dieser Verweise.

Im Kontext des Semantic Webs und der Informatik allgemein dienen Annotationen der Einbindung von Quelltexten oder Metadaten und haben keine direkte Auswirkung auf die Syntax der Dateninstanz. Das Semantic Web sieht Annotationen innerhalb von Webseiten als Art semantische Ebene der Beschreibung vor.¹⁸ Diese Verweise enthalten daher Informationen, die der Maschine nicht zur Verfügung stehen, oder Verweise auf externe Wissensquellen¹⁹.

2.6.2 Wissensmodellierung und Wissensrepräsentation

Die Modellierung und Darstellung von Wissen in Form von Ontologien oder anderen Systemen sowie das Referenzieren von Inhalten des Internets zu diesen Modellen sind die wesentlichen Methoden des Semantic Web, welche auch im Zusammenhang dieser Arbeit eine wichtige Rolle spielen. Das Ziel ist es, eine Ebene mit komplexen, individuellen Metainformationen zu schaffen, welche von Maschinen interpretiert und verarbeitet werden kann. Diese Metainformationen gehen weit über das hinaus, was üblicherweise in der Geoinformatik unter dem Begriff Metadaten verstanden wird und ermöglicht eine interoperable (syntaktische als auch semantische) Verwendung der Informationen. Die Problemstellung der fehlenden Interoperabilität sowie die Ansätze der Lösung lassen sich vom allgemeinen Internet auch auf derzeitige Entwicklungen und Tendenzen in der Geoinformatik übertragen.

Die Felder der Wissensmodellierung und Wissensrepräsentation sind logisch voneinander abhängig. Die Wissensrepräsentation beschäftigt sich mit der Abbildung der Bedeutung von Ausdrücken durch eine formale Sprache oder Symbolik. Das grundsätzliche Ziel der Abbildung von Wissen ist die Vermittlung und der Transfer des Wissens zwischen zwei oder mehreren getrennten Domänen. Zudem ist die Abbildung

¹⁸ Annotation (<http://de.wikipedia.org/wiki/Annotation>, 2009)

¹⁹ Externe Wissensquellen im konkreten Kontext sind z.B. Ontologien, Thesauri, ect.

von Wissen bzw. Propositionen die Grundvoraussetzung für Reasoning²⁰, welches basierend auf dem bestehenden Wissen operiert und daraus einen Mehrwert ableiten soll. (R. Brachman, H. Levesque, 2004)

Zu den verschiedenen Systemen der Wissensrepräsentation gehören unter Anderem²¹:

- Taxonomy (Taxonomie)
- Thesaurus
- Conceptual Model (Konzeptuelle Modelle)
- Logical Theory (Formale Logik)

Bei der obigen Auflistung steigt der Grad der semantischen Interoperabilität von der Taxonomie zur logischen Theorie an. Damit verbunden ist nach Orbst allerdings eine steigende Generalisierung der Aussage, sodass die Wahl der entsprechenden Werkzeuge anhand des zu erfüllenden Zweckes entschieden werden muss. (Orbst, 2006). Das in dieser Arbeit favorisierte Werkzeug der Wissensrepräsentation – die Ontologie - ist den letzten beiden Punkten und somit den ausdrucksstärksten semantischen Werkzeugen zuzuordnen.

Im Rahmen des Semantic Web ermöglicht die Wissensmodellierung und Wissensrepräsentation, explizites Wissen formal und für Maschinen interpretierbar abzubilden, wie in Abbildung 2 und Abbildung 3 verdeutlicht wird.

Der Aspekt der Wissensmodellierung wird ebenso wie die Annotation in einem gesonderten Kapitel intensiver behandelt (Kapitel Ontologie und Kapitel Semantische Annotation).

2.7 Geosemantik und Interoperabilität in der vernetzten Geoinformatik

²⁰ „Reasoning“, englisch für Schlussfolgerung, ist ein in der Wissensrepräsentation und künstlichen Intelligenz genutzter Fachbegriff, welcher für die Ableitung von Kenntnissen aus bestehenden Wissensrepräsentationen steht

²¹ Da einige der Ausdrücke englischer Literatur entstammen und auch im deutschen Sprachraum in ihrer englischen Form benutzt werden, wird die folgende Aufzählung homogen in Englisch wiedergegeben

Das Ziel dieses Kapitels ist es, zu erläutern, welche Relevanz das Thema der Semantik und des Semantic Webs für die Geoinformatik hat.

Bereits ein Jahr nach dem Erscheinen von Semantic Web²² spricht Egenhofer²³ in „*Toward the semantic geospatial web*“ die Probleme im Bereich der Geographie an. Egenhofer zieht Parallelen zwischen der Situation des Webs 2.0 und der immer stärker vernetzten Geoinformationslandschaft. So sind fehlende Interoperabilität und ein hoher Verlust an Semantik auch in der Geoinformatik zu finden. Durch die Spezifikationen und Standards des OGC²⁴ konnte bereits ein gewisses Niveau der syntaktischen Interoperabilität erreicht werden, die semantische Interoperabilität wurde allerdings weitgehend vernachlässigt.

Egenhofer sieht den Bedarf nach mehr semantischer Unterstützung vor allem in zwei Bereichen. Zum einen im Bereich der Suche nach Daten (Discovery) und zum anderen in der interoperablen Zusammenarbeit von Operanten (Computing). Desweiteren bemerkt Egenhofer in seinem Artikel, dass das Semantic Web zwar einen generischen Rahmen für die Arbeit mit Ontologien und Verweisen bietet, dieser aber der Vielfalt und den Ansprüchen der Geoinformatik nicht gerecht wird.

Derzeitige Suchmechanismen basieren hauptsächlich auf Schlüsselwörtern oder Textvergleichen, was sich seit 2002 bis heute nicht wesentlich verändert hat. Im Bereich der Geoinformatik werden zudem räumliche Filter eingesetzt und die Suchfunktion um die räumliche Komponente erweitert. Diese Methoden sind nicht ausreichend, um das Potential auszuschöpfen und dem Nutzer sinnvolle und effektive Antworten auf seine Suchanfragen zu liefern. Hierbei besteht nach wie vor Bedarf an semantischer Unterstützung beim Vorgang der Suche nach Informationen und Informationsquellen. Wir verfügen heute mit verschiedenen Standards für die Beschreibung von Metadaten zwar über syntaktische Interoperabilität und können Metadaten sammeln und

²² T. Berners-Lee, 2001

²³ M. Egenhofer, 2002

²⁴ OGC, Open Geospatial Consortiums

verarbeiten, jedoch unterscheiden sich die Metadaten durch unterschiedliche Terminologie und Sprachen. Wir können hierbei die bereits angesprochenen hierarchischen und multilingualen Probleme erkennen. (Klien, 2009 & Egenhofer, 2002)

Ein weiterer Punkt, an welchem Egenhofer semantische Unterstützung für notwendig hält, ist die interoperable Zusammenarbeit zwischen verschiedenen Operanten. Hierbei ist es das Ziel, Informationen effizient und nutzbringend weiterzugeben. Als Adressaten dieser Informationen sieht Egenhofer sowohl die Maschine, welche die Information verarbeitet, als auch den Menschen, welcher die Informationen verstehen muss. Diese Art der semantischen Interoperabilität ist das Hauptziel des Semantic Webs. Hier haben wir durch die bereits angesprochenen Spezifikationen und Standards des OGC ein gewisses Maß an syntaktischer Interoperabilität erreicht. Eine wirkliche Interoperabilität schließt hierbei jedoch die Kenntnis der Bedeutung der Daten ein. Obwohl Interoperabilität in der Geoinformatik als wichtiges Thema diskutiert wird und OGC Web Services allgemein als Lösung des Problems begriffen werden, bedeuten diese jedoch nur eine syntaktische Interoperabilität. Für eine Zusammenarbeit von Web Services oder Maschinen im Allgemeinen muss die Kenntnis der Bedeutung der jeweilig transferierten Daten vorhanden sein. Im Bereich der Geoinformatik und der damit zusammenhängenden Web Services finden häufig flache Datenstrukturen Einsatz. Dies bedeutet, dass man sich in der Übermittlung von Daten (Metadaten) einschränkt und wichtige Informationen nicht transferiert werden können, da diese zu komplex sind. In der Praxis beschränken wir uns daher häufig auf Abkürzungen (wie z.B. „bgld“ für Burgenland oder „uba“ für Umweltbundesamt), welche die Bedeutung der Information reduzieren und oft nur schwer erkennen lassen. Ist dies der Fall, sind wir von echter Interoperabilität noch weit entfernt (Klien, 2009 & Egenhofer, 2002).

Auch W. Kuhn sieht die Interoperabilität in der vernetzten Geoinformatik durch die Teilung in Daten und Methoden erschwert. Zudem betont Kuhn in seinem Artikel zum Thema „Geospatial Semantics“²⁵, dass Semantik an sich nichts Neues auf dem Gebiet der Geoinformatik darstellt. Vielmehr werden semantische Technologien bereits seit längerer Zeit benutzt, um Interoperabilität zu unterstützen bzw. diese zu gewährleisten.

²⁵ W. Kuhn, 2005

Kuhn sieht daher die Notwendigkeit eines „Semantic Reference Systems“ für geographische Daten, um eine grundlegende Interoperabilität zu gewährleisten. (Kuhn, 2005)

Dies stellt nur eine Auswahl an Problemen und Ideen dar, welche mit mangelnder Interoperabilität und dem Verlust an Information zusammenhängen. Dadurch wird vor allem deutlich, dass auch der Sektor der Geoinformatik durch die zunehmende Vernetzung von Daten und Services mit den Problemen konfrontiert ist, welche zu den Visionen des Semantic Web führten. Geodaten rücken zunehmend in den Fokus der breiten Öffentlichkeit²⁶ und können daher nicht getrennt vom Internet betrachtet werden. Die mangelnde Interoperabilität wird zudem durch den hohen Grad an Heterogenität (Daten, Methoden als auch Services) erschwert. Diese Probleme und der Bedarf nach mehr Homogenität und Interoperabilität wurden erkannt, auch wenn die zahlreichen Bestrebungen und Ansätze, im Gegensatz zum Semantic Web, bisher keine weite Verbreitung gefunden haben. Die semantische Erweiterung unserer bisherigen Informationsnetzwerke ist zwar keine alleinige Lösung, aber eine der Schlüsselkomponenten, um deren Potential und Möglichkeiten auch für die Geoinformatik nutzen zu können. Durch das OGC wurde im Rahmen des „*Geospatial Semantic Web Interoperability Experiment Report*“²⁷ untersucht, welche Schritte im Zusammenhang mit der Schaffung eines Geospatial Semantic Web (GSW) und damit der Verbesserung der semantischen Interoperabilität unternommen werden müssen. Zu den weiteren Aktivitäten des OGC, welchem in der Standardisierung der vernetzten Geoinformatik eine Schlüsselrolle zukommt, bezüglich der Semantik gehören u. A.:

- „*Geospatial Semantic Web Interoperability Experiment Report*“, 2005
- Untersuchung der notwendigen Schritte zur Schaffung eines Geospatial Semantic Web aufbauend auf den Konzepten und Methoden des Semantic Web
- „*Semantic Annotations in OGC Standards*“, 2009

²⁶ z.B. Applikationen wie Google Earth oder Google Maps

²⁷ J. Liebermann, 2006

- Untersuchung wie semantische Annotation in OGC Standards (z. B. WFS, GML, Metadaten) realisiert werden kann
- *SensorML*²⁸ und *Observation & Measurement*²⁹
- Standards aus dem Bereich Sensor Web (*OGC Sensor Web Enablement*), welche ein hohes Maß an semantischer Ausdrucksstärke besitzen und erstmals Techniken wie Annotation und Ontologien seitens der OGC vorschlagen

Neben den Aktivitäten der OGC widmen sich weltweit zahlreiche Forschungsprojekte³⁰ dem Thema der semantischen Interoperabilität auch auf dem Bereich der Geoinformatik. Geosemantik gewinnt proportional zur wachsenden Vernetzung von Geodaten durch das Internet an Bedeutung.

²⁸ SensorML, OGC Standard zur Beschreibung von Sensor Metadaten und Prozessen (<http://www.opengeospatial.org/standards/sensorml>, 2009)

²⁹ Observation & Measurement, OGC Standard zur Beschreibung von Messungen

³⁰ z.B.

SWING, Semantic Web Services Interoperability for Geospatial Decision Making (<http://www.swing-project.org/>, 2009)

GDI-GRID, Geodateninfrastruktur-Grid, <http://www.gdi-grid.de/>

3 Ontologie

Das Ziel dieses Kapitels ist es, das Thema der Wissensrepräsentation weiter einzuführen.

Die Wissensrepräsentation in ihren verschiedenen Formen ist ein wichtiger Bestandteil des Semantic Web. Zudem ist sie als Referenzobjekt wichtiger Bestandteil der vorgestellten Methode der semantischen Annotation, welche in dieser Arbeit gesondert behandelt wird. Das folgende Kapitel soll daher einen Überblick über die verschiedenen Systeme der Wissensrepräsentation darstellen. Darauf aufbauend soll besonders die Ontologie, welche einen sehr hohen Grad an semantischer Ausdruckskraft besitzt, genauer betrachtet werden.

Im Abschluss des Kapitels soll zudem die Beispielontologie vorgestellt werden, welche im daran anschließenden Kapitel über die semantische Annotation für Demonstrationszwecke verwendet wird.

3.1 Systeme der Wissensrepräsentation

Die grundsätzlichen Ziele der Wissensrepräsentation sind es, bestehendes Wissen anhand von Symbolen und Sprachen explizit und formal abzubilden. Dazu gehören vor allem komplexe Zusammenhänge. Historisch gehen die ersten Ansätze der Darstellung und Strukturierung von Wissen in die Antike zurück, welche allerdings für die vorliegende Arbeit nicht relevant sind.³¹ Für weitere Ausführungen, welche den Umfang dieser Arbeit übersteigen, sei auf die angeführte Fachliteratur verwiesen.

Wie bereits im Kapitel 2.6.2 beschrieben, werden auf dem Gebiet der Wissensrepräsentation verschiedene Systeme und Methoden eingesetzt, um Wissen

³¹ Aristoteles (384 v. Chr. – 322 v. Chr.) stellt Überlegungen zur Klassifikation und Unterteilung von Lebewesen an; vgl. dazu M. Stuck (2008)

abzubilden und zu strukturieren. Diese Systeme entstammen verschiedenen Ansätzen der Nutzung als auch verschiedenen Gebieten der Wissensrepräsentation. Dabei ist zu beachten, dass die Wissensrepräsentation als Feld sehr weit gefächert ist und sich die vorliegende Arbeit daher ausschließlich auf die Bereiche der Wissensordnung und Ontologie beschränkt.

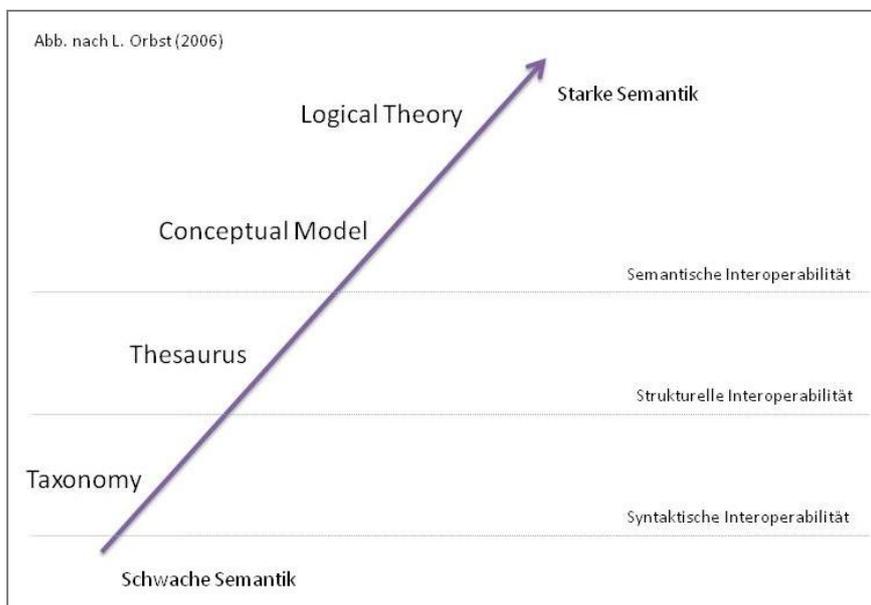


Abbildung 5: Semantische Technologien nach Orbst

Heute können wir verschiedene wissenschaftliche Ansätze der Gliederung und Strukturierung der Methoden des Wissensmanagement beobachten, welche sich vom jeweiligen Standpunkt der Autoren ableiten.³² Da sich diese Arbeit dem Thema der Wissensrepräsentation vom Feld der Informationstechnologie hin nähert, wird der Ansatz von L. Orbst³³ verwendet, welcher zwischen Taxonomie, Thesauri, konzeptuellen Modellen sowie der logischen Theorie unterscheidet. Dieser groben Gliederung der Methoden der Wissensrepräsentation können verschiedene konkrete Technologien zugeordnet werden:

³² Vgl. dazu W. Stock und M. Stock, 2008

³³ L. Orbst, 2006

- Taxonomie
 - Katalog
 - Glossar
 - Taxonomie
- Thesaurus
 - Klassifikation
 - Thesaurus
- Konzeptuelles Modell
 - Ontologien
 - Semantische Netze
 - Frames
- Formale Logik (Logical Theory)
 - Prädikatenlogik
 - Axiomensysteme

Die hierbei genannten Technologien und Methoden der Wissensrepräsentation erheben keinerlei Anspruch auf Vollständigkeit. Vielmehr soll damit ein Überblick über bestehende Methoden geboten werden. Zudem sei bereits an dieser Stelle auf einen semantischen Konflikt in der Definition von Ontologien hingewiesen. In der Fachliteratur wird der Ausdruck Ontologie in folgenden beiden Varianten gebraucht:

- Als Oberbegriff für semantische Ontologien, welcher unter Anderem auch Thesauri, Taxonomien sowie Ontologien (entsprechend ihrer zweiten Bedeutung) umfasst (L. Orbst, 2006, T. Gruber 1998)
- Als Ausdruck für eine spezielle Form der Wissensrepräsentation, welche sich durch Konzepte, Instanzen und Relationen abhebt sowie nach dem Thesaurus die nächste Stufe der semantischen Ausdruckskraft darstellt und sich von diesem abgrenzt

In weiterer Folge sollen sowohl der Thesaurus als auch die Ontologie wegen ihrer besonderen Bedeutung für die semantische Interoperabilität und ihres Einsatzes in der Geodateninfrastruktur genauer beschrieben werden.

3.2 Thesaurus

Der Thesaurus bezeichnet ein kontrolliertes Vokabular, dessen Inhalte durch Relationen verknüpft sind. Diese Relationen bezeichnen in der Regel Synonyme, können aber genauso auf Ober- bzw. Unterklassen des Begriffes verweisen. Der Thesaurus stellt ein Modell, bestehend aus einer Sammlung von Begrifflichkeiten und Relationen, dar, welches ein oder mehrere Themengebiete beschreiben kann. Thesauri stellen in ihren Möglichkeiten und ihrem Umfang die semantische Vorstufe zur eigentlichen Ontologie dar und finden weite Verbreitung und vielfältigen Einsatz.

Heutzutage finden wir zahlreiche Beispiele für den Einsatz von verschiedenen Thesauri. Ein Beispiel für einen Thesauri aus dem europäischen Raum ist der GEMET³⁴, welcher im Auftrag der Europäischen Umweltagentur³⁵ entwickelt wurde und durch *European Environment Information and Observation Network*³⁶ publiziert wird. Dieser Thesaurus ist derzeit in 28 Sprachen verfügbar und daher multilingual. Mit ihm wird versucht, das Themengebiet der Umwelt und Umweltbeobachtung mit speziellem Fokus auf Europa abzudecken. Der GEMET spielt eine wichtige Rolle bei europäischen Initiativen zur Geoinformation und Umweltinformation wie INSPIRE oder SEIS.

³⁴ GEMET, GEneral Multilingual Environmental Thesaurus (<http://www.eionet.europa.eu/gemet>, 2009)

³⁵ European Environment Agency (EEA, <http://www.eea.europa.eu/>, 2009)

³⁶ European Environment Information and Observation Network (EIONET, <http://www.eionet.europa.eu/>, 2009)

Thematic Listings | [INSPIRE Spatial Data Themes](#) | [Alphabetic Listings](#) | [Hierarchical Listings](#) | [Search Thesaurus](#)

Select language: [ar](#) [bg](#) [cs](#) [da](#) [de](#) [el](#) [en](#) [en-US](#) [es](#) [et](#) [eu](#) [fi](#) [fr](#) [ga](#) [hu](#) [it](#) [lv](#) [mt](#) [nl](#) [no](#) [pl](#) [pt](#) [ro](#) [ru](#) [sk](#) [sl](#) [sv](#)

water (substance)

Definition:
Common liquid (H2O) which forms rain, rivers, the sea, etc., and which makes up a large part of the bodies of organisms. (Source: PHC)

broader terms

- natural material

narrower terms

- drinking water
- hot water
- mineral water
- raw water
- spring water
- water for agricultural use
- water for consumption
- water for industrial use

related terms

- bathing water
- water (geographic)

Scope note:
scope note is not available

Groups:
PRODUCTS, MATERIALS

Themes:
chemistry
water

عربي:	ماء (مادة)
Български:	Вода (вещество)
Čeština:	voda [hmota]
Dansk:	vand
Deutsch:	Wasser
Ελληνικά:	νερό
English (US):	water (substance)
Español:	agua
Eesti keel:	vesi
Euskara:	ur (substantzia)
Suomi:	vesi
Français:	eau (substance)
Ghaeilge:	uisce (substaint)
Magyar:	víz (anyag)
Italiano:	acqua (sostanza)
Lietuvių kalba:	vanduo (medžiaga)
Latviešu:	ūdens (viela)
Malti:	ilma (sostanza)
Nederlands:	water
Norsk:	vann (stoff)
Polski:	woda (substancja)
Português:	água (substância)
Română:	apă (substanță)
Русский:	вода (вещество)
Slovenčina:	voda
Slovensčina:	voda (snov)
Svenska:	vatten

[PDF Files](#) | [Administration](#) | [Alphabets](#) | [About GEMET](#) | [Web services](#) | [Definition sources](#)

GEMET - Concepts, version 2.3, 2009-07-13

Abbildung 6: GEMET Beispiel für den Ausdruck „Water“

Durch Abbildung 6 ist die Weboberfläche des GEMET, welche über <http://www.eionet.europa.eu/gemet> (2009) erreichbar ist, dargestellt, um einen Eindruck über den Umfang und die Möglichkeiten dieses Thesaurus vermitteln.

Der Thesaurus spielt durch seine Funktion als kontrolliertes Vokabular bereits heute eine wichtige Rolle in Teilen der Geoinformatik. Ein Thesaurus ermöglicht es als externe Referenzliste, eine festgelegte Menge an Begrifflichkeiten zur Verfügung zu stellen. Dies wird beispielsweise durch den Metadatenstandard ISO 19115³⁷ genutzt, welcher weltweit (neben anderen Standards für Metadaten) eingesetzt wird. Innerhalb dieses Standards ist die Verwendung von Schlüsselwörtern (engl. Keywords) vorgesehen, welche optional aus einem kontrollierten Vokabular entstammen können,

³⁷ ISO 19115, <http://www.iso.org>, ISO 19115 ist Teil der Serie zur Normierung von Geodaten und Geoinformation

welches wiederum angegeben werden kann und muss. Hier kann eine erste Form der semantischen Annotation festgestellt werden, welche der Gegenstand dieser Arbeit ist. ISO 19115 bildet auch die Basis, auf welcher in adaptierter Form die Metadaten der INSPIRE-Direktive beruhen. Für die INSPIRE konforme Metadatenbeschreibung wurden einige Elemente des ursprünglichen ISO 19115 bezüglich ihrer Rolle³⁸ und ihrer Kardinalität hin abgewandelt. So wurde die in ISO 19115 noch optionale Angabe eines kontrollierten Vokabulars dahingehend abgewandelt, dass jeder Metadatensatz mindestens ein Schlüsselwort des bereits erwähnten GEMET enthalten muss, um mit den Metadatenrichtlinien³⁹ von INSPIRE konform zu sein.

```

...
<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gco:CharacterString>Environmental Monitoring Facility</gco:CharacterString>
    </gmd:keyword>
    <gmd:keyword>
      <gco:CharacterString>waste, waste disposal, hazardous waste </gco:CharacterString>
    </gmd:keyword>
    <gmd:thesaurusName>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>GEMET Thesaurus Version 1.0</gco:CharacterString>
        </gmd:title>
        <gmd:date>
          <gmd:CI_Date>
            <gmd:date>
              <gco:Date>2004</gco:Date>
            </gmd:date>
            <gmd:dateType>
              <gmd:CI_DateTypeCode codeListValue =
"http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml#CI_DateTypeCode" codeList =
"publication">publication</gmd:CI_DateTypeCode>
            </gmd:dateType>
          </gmd:CI_Date>
        </gmd:date>
      </gmd:CI_Citation>
    </gmd:thesaurusName>
  </gmd:MD_Keywords>
</gmd:descriptiveKeywords>
...

```

Abbildung 7: Auszug einer INSPIRE konformen Metadatenbeschreibung unter Verwendung von GEMET

³⁸ Rollen wie Mandatory, Optional, Conditional

³⁹ INSPIRE Metadata Regulation vom 03.12.2008, <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/101>

Die Abbildung 7 zeigt einen Auszug aus einer INSPIRE konformen Metadatenbeschreibung⁴⁰. Darin ist das Element <gmd:descriptiveKeywords> gezeigt, welches Schlüsselwörter des aufgeführten GEMET Thesaurus Version 1.0 enthält.

Zusammenfassend lässt sich festhalten, dass Thesauri das Wissen und die Nomenklatur eines Fachgebietes strukturiert repräsentieren können. Zudem bieten sie neben der Struktur einen oft multilingualen Wortschatz (bezogen auf das Gebiet), welcher die semantische Interoperabilität unterstützt und erleichtert (P. Maué, 2009).

3.3 Ontologie

Der Begriff „Ontologie“ ist nicht eindeutig definiert und wird sowohl in der Informationswissenschaft als auch in der Philosophie in verschiedenen Bedeutungen verwendet. Im vorliegenden thematischen Rahmen orientieren wir uns an der Bedeutung des Wortes Ontologie aus den Informationswissenschaften.

Die Unterscheidung zwischen Ontologie und Thesaurus ist oft nicht klar und am besten von den jeweiligen Anwendungsfällen ableitbar. Ontologien dienen in den Informationswissenschaften zur Abbildung und Kommunikation von Wissen und sind damit essentieller Teil der Ideen des Semantic Web. Durch die Fachliteratur wird die Ontologie unter anderem auch als „explizite formale Spezifikation einer Konzeptualisierung“ (T. Gruber, 1998) bezeichnet. Diese Definition umfasst auch semantische Methoden wie den Thesaurus, Klassifikationen und Taxonomie. Daher wird im Zusammenhang der Definition nach Gruber zwischen dem Grad der Konzeptualisierung des Wissens unterschieden.

Wie bereits erwähnt, bestehen in der Fachliteratur unterschiedliche Auffassungen des Begriffes Ontologie und dessen begrifflicher Abgrenzung. Ontologie bezeichnet (hier) eine Technologie, wie sie im Kapitel „Aufbau von Ontologien“ beschrieben ist. Zudem dient der Begriff Ontologie in seiner Definition als formale Spezifikation einer Konzeptualisierung als Überbegriff für semantische Methoden wie Thesauri, Frames,

⁴⁰ Das Schema und die Abbildung der ISO 19115 wird durch die Norm ISO 19139 festgelegt

Prädikatenlogik oder Taxonomie.⁴¹ Hiermit soll festgehalten werden, dass sich die folgenden Erläuterungen auf die Ontologie als spezifische Methode beziehen und nicht auf die Bedeutung als Sammelbegriff für die verschiedenen Formen der Wissensrepräsentation.

Wie bereits angedeutet, bildet die Ontologie nach dem Thesauri die nächste Stufe der semantischen Ausdrucksfähigkeit. Der hauptsächliche Unterschied zwischen Thesauri und Ontologien ist die Fähigkeit der Ontologie, komplexes Wissen (Allgemein- und Spezialwissen) zu repräsentieren. Die Ontologie stellt wesentlich mehr dar als nur ein kontrolliertes Vokabular mit Begrifflichkeiten einer Domäne. Die Ontologie ist vielmehr eine komplexe Repräsentation von Wissen einer Domäne, welches ein abgegrenztes Szenarium beschreibt. (T. Gruber, 1998) Ontologien dienen anders als Thesauri der Zusammenführung, Integration und Verarbeitung von Wissen, wohingegen Thesauri eher im Gebiet der Wissensstrukturierung und als kontrollierte Vokabulare eingesetzt werden.

Besonders im Hinblick auf die Informationswissenschaften müssen wir zudem zwischen der formalen und informalen Ausdrucksweise einer Ontologie unterscheiden. Informale Ontologien bedienen sich größtenteils der natürlichen Sprache, verzichten auf Formalismen und finden vor allem in der Philosophie Verwendung. Formale Ausdrucksweise bezeichnet dagegen die Verwendung von scharf definierten Begriffen und Semantik durch eine formale Sprache. Formale Ontologien werden fachgebietsübergreifend als Methode verwendet, Spezialwissen deklarativ und maschinenlesbar auszudrücken. Die formalen Ausdrucksweisen von Ontologien werden im Unterkapitel Ontologiesprachen behandelt.

3.3.1 Einsatzmöglichkeiten von Ontologien

Im Konzept und in der Architektur des Semantic Web, spielt die Ontologie - als Methode der Wissensrepräsentation – eine Schlüsselrolle, wodurch ebenfalls das Konzept der Ontologie in den Informationswissenschaften beflügelt wurde. Die

⁴¹ Vgl. dazu T. Gruber, 1998 und N. Guarino, 1998 sowie W. Stock, 2008

wesentlichen Gründe für den Einsatz von Ontologien im Zusammenhang mit Informationsnetzwerken sind:

- Ontologie als Mediator in einer inhomogenen Informationsinfrastruktur (**semantische Interoperabilität**) zur Überbrückung der semantischen Inhomogenität und Multilingualität
- Ontologien für die Unterstützung der Wissenssuche, Informationsgewinnung (*Information Retrieval*) und Personalisierung der Informationen
- Ermöglicht und verbessert die Kommunikation zwischen Agenten untereinander als auch zwischen Agent und Mensch
- Ontologien für die Unterstützung der Wissensvisualisierung

Basierend auf diesen Eigenschaften haben sich Ontologien in den vergangenen Jahren vor allem in den Informationswissenschaften zu einem populären Thema entwickelt. Allerdings muss ebenso festgehalten werden, dass ihr Einsatz trotz der theoretischen Möglichkeiten immer noch mit Schwierigkeiten behaftet ist. Dies liegt vor allem an der Komplexität und dem Umfang, den derartige Wissensrepräsentationen entfalten können. Eine Ontologie ist als ein Werkzeug zu verstehen, welches ein auf sie abgestimmtes Umfeld benötigt um einen wirklichen Mehrwert für das Informationsnetzwerk zu gewährleisten. Daher kann man bis heute nicht von einer weiten Verbreitung und einem umfassenden Einsatz von Ontologien sprechen, da sich viele Aspekte des Semantic Web und der Ontologie immer noch im Stadium der Forschung befinden.

Dennoch haben Ontologien durch die Informationswissenschaft bereits in den Bereichen Einzug gehalten, in denen vor allem die schnelle Verfügbarkeit von relevanten Informationen im Vordergrund steht. Dazu zählen heute vor allem die Biomedizin, der Katastrophenschutz und die Kriminalistik. Ein konkretes Beispiel für den Einsatz von semantischen Technologien in der Medizin ist die Forschungsgruppe *Ontologies in Medicine and Life Sciences*⁴², *Onto-Med*, welche sich u. A. auf Top-Level Ontologien fokussiert. Heute können wir weltweit zahlreiche solche Beispiele für den interdisziplinären Einsatz und die Entwicklung von Ontologien sowie Nutzungskonzepten beobachten.

⁴² Onto-Med, Ontologies in Medicine and Life Sciences (<http://www.onto-med.de/>, 2009)

Eine weitere Anwendung von Ontologien ist die maschinelle Generierung von neuem Wissen bzw. die Folgerung aus bestehendem Wissen. Man kann vereinfacht ausdrücken, dass durch *Reasoning*⁴³ aus bestehenden Aussagen und Formulierungen neue Information abgeleitet werden kann. Voraussetzung für das *Reasoning* ist eine formale Spezifikation der Konzeptualisierung und die Anwendung von *Description Logic*⁴⁴. (Maué, 2009) *Reasoning* ist ein wichtiger Aspekt im Einsatz von Ontologien. Da die Funktionalität des *Reasoning* für die vorliegende Arbeit nicht relevant ist, wird an dieser Stelle nur auf die angefügten Referenzen verwiesen. Ebenso soll die im Kapitel Beispielontologie verwendete Ontologie nur Demonstrationszwecken für die semantische Annotation dienen und keineswegs die Möglichkeit zum *Reasoning* bieten.

Die beschriebenen Einsatzmöglichkeiten von Ontologien lassen sich ohne große Adaptionen auch auf das Gebiet der Geoinformatik übertragen. Wie auch bei anderen Fachbereichen, welche Ontologien einsetzen, ist auch hier zu beachten, dass eine Ontologie keinen unabhängigen Operator, sondern vielmehr ein Werkzeug und einen Mediator darstellt. Eine Ontologie muss daher als Teilstück in einer Informationsinfrastruktur verstanden werden und bedarf Applikationen, welche auf sie zugreifen und ihre Informationen verwerten können sowie Annotation auf verschiedenen Ebenen, um auf sie zu referenzieren.

3.3.2 Aufbau von Ontologien

Grundsätzlich ist beim Aufbau der Ontologie zwischen dem Konzept und der Realisierung zu unterscheiden. Mit Konzept (Aufbau) sind in diesem Zusammenhang die Repräsentation des Wissens und die Struktur der Ontologie zu verstehen. Mit Realisierung (Syntax) werden die Methoden und Sprachen verstanden, mit denen die semantischen Strukturen ausgedrückt und ausgetauscht werden.

⁴³ Reasoning, engl. für Schlussfolgerung; da zu diesem englischen Fachausdruck keine gebräuchliche deutsche Übersetzung verwendet wird, wird der englische Term verwendet

⁴⁴ Vgl. dazu R. Brachmann & H. Levesque, 2004 oder P. Hitzler, 2008

Ähnlich dem Thesaurus stellt eine Ontologie ein Wissensnetz und abstraktes Modell dar. Wie bereits erwähnt, sind die Möglichkeiten der Modellierung von Wissen durch die Ontologie wesentlich umfangreicher als durch die eines Thesaurus. Wir sprechen daher von unterschiedlichen Stufen der Konzeptualisierung.⁴⁵

Durch das Ontology Definition Metamodel⁴⁶ werden vier Bestandteile genannt, welche für eine Konzeptualisierung notwendig sind. (Maué, 2009) Die unterschiedlichen Bezeichnungen entstammen unterschiedlicher Terminologien in der Literatur und Anwendungen.

- Konzepte, Klassen oder Begriffe
- Individuen oder Instanzen
- Relationen oder Eigenschaften
- Axiome

Konzepte, auch als **Klassen** oder **Begriffe** bezeichnet, sind das Grundgerüst der Struktur der Ontologie. Ein Konzept steht in der Regel für eine Begrifflichkeit, wie zum Beispiel „Station“ oder „Schadstoffe“. Die Klassen haben dabei entsprechende Ober- und Unterbegriffe, wobei die Klassen der Ontologie keiner strengen hierarchischen Baustruktur unterliegen wie zum Beispiel ein XML-Dokument. Unterbegriffe können dabei von ihren entsprechenden Oberbegriffen Eigenschaften erben. Konzepte können daher auch mit Klassen der objektorientierten Programmierung verglichen werden. Konzepte bzw. Klassen können miteinander in Beziehung stehen, was wiederum als Relation bezeichnet werden kann. Zudem können Konzepte über eine Anzahl an Attributen verfügen und sind in der Regel nach Taxonomien strukturiert.

Individuen, auch als **Instanzen** bezeichnet, bilden die Wissensbasis innerhalb der Ontologie. Instanzen sind in der Regel Objekte bzw. Begriffe und sind Konzepten zugeordnet. So kann beispielsweise „Ozon“ eine Instanz der Klasse „Schadstoffe“ sein. Ein bisher nicht geklärtes Problem ist die genaue Verwendung von Instanzen in Beziehung zu Klassen, da sie sich von Ontologie zu Ontologie unterscheiden kann. Ein

⁴⁵ Nach T. Gruber, 1998

⁴⁶ Erstellt als Whitepaper durch die Arbeitsgruppe der „Object Management Group“ (OMG) (<http://www.omg.org/ontology/>, 2009)

Beispiel für diese Form des Konfliktes ist der Ausdruck „Homo Sapiens“, welcher zum einen ein Konzept für weitere Instanzen wie „Hans Muster“ aber auch selbst eine Instanz (zum Beispiel des Konzeptes „Säugetiere“) darstellen kann. Diese Variationen sind durch den Anwendungsfall der Ontologie und deren spezifische Sichtweise bedingt.

Relationen – beziehungsweise **Eigenschaften** – bezeichnen Beziehungen zwischen Konzepten und Individuen. Ein Beispiel für eine Relation zwischen den Instanzen „Luftmessstation 123“ und „Ozon“ ist „misst“⁴⁷.

Axiome entstammen der Prädikatenlogik und dienen zur Formulierung von Aussagen, welche immer wahr sind. Innerhalb einer Ontologie dienen Axiome zur Konsistenzprüfung und zur Darstellung von Wissen, welches mit den bereits genannten Komponenten nicht dargestellt werden kann. Ein Beispiel für ein Axiom ist die Festlegung, dass „PM10“ nicht gleich „PM25“⁴⁸ ist. Axiome sind besonders im Zusammenhang mit der „Open World Assumption“ sinnvoll, welche alles als grundsätzlich möglich ansieht, was nicht in der Ontologie spezifiziert wurde.

Die Abbildungen 8 und 9 stellen eine Beispielontologie⁴⁹ grafisch dar, in welcher Konzepte, Relationen und Instanzen farblich hervorgehoben wurden. Die Darstellung der Instanzen leitet sich dabei von den obigen Konzepten ab.

⁴⁷ Daher „Luftmessstation 123“ „misst“ „Ozon“

⁴⁸ Perimeter (PM), engl. für Feinstaub

⁴⁹ Die Abbildung der gezeigten Beispielontologie entstammt
[http://de.wikipedia.org/wiki/Ontologie_\(Informatik\)](http://de.wikipedia.org/wiki/Ontologie_(Informatik)), 2009

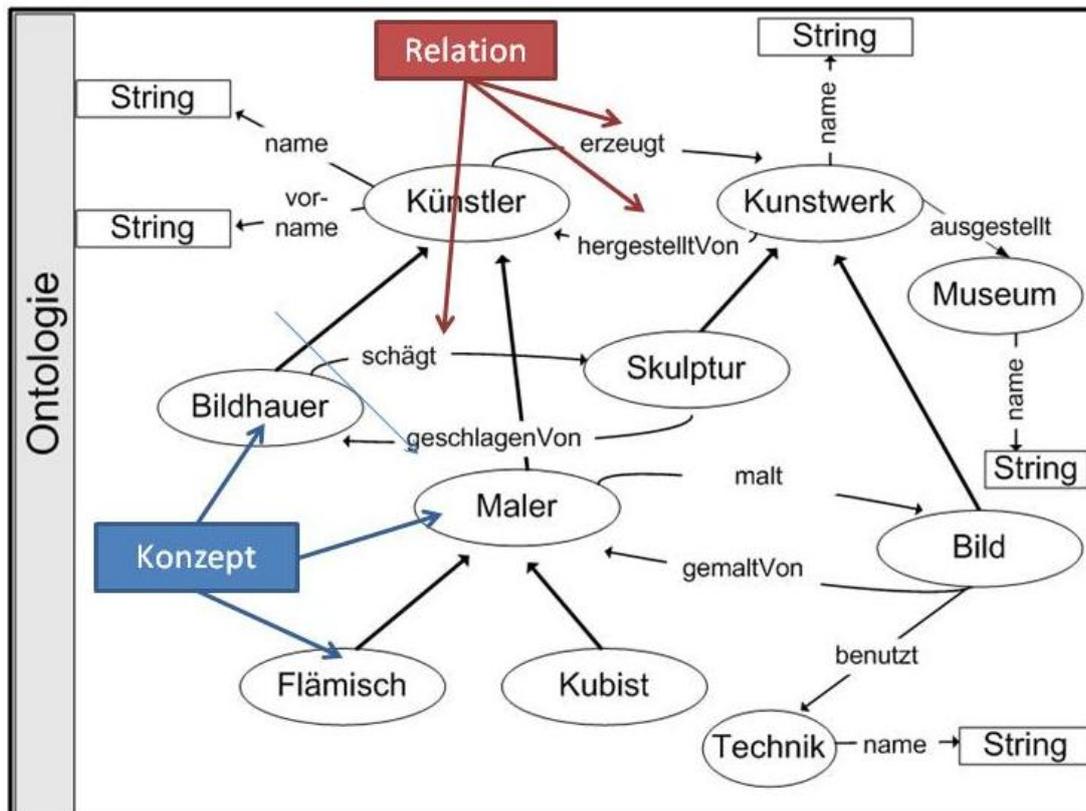


Abbildung 8: Beispielontologie (Konzepte und Relationen)

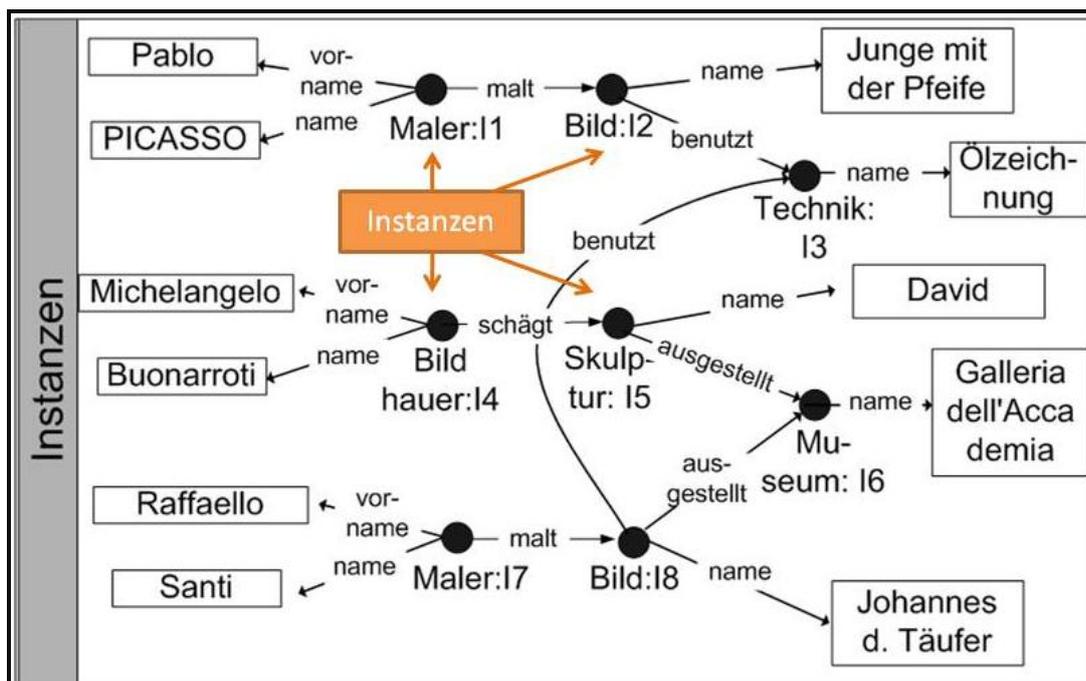


Abbildung 9: Beispielontologie (Instanzen und Relationen)

3.3.3 Ontologiesprachen

Im folgenden Absatz soll nur ein kurzer Überblick über die derzeit verwendeten formalen Sprachen gegeben werden, mit welchen Ontologien in den Informationswissenschaften ausgedrückt werden können.

Die unten dargestellte Abbildung 10, bekannt als „The Semantic Web Stack“⁵⁰, zeigt die Architektur des Semantic Webs. Sie zeigt, welche Techniken aufeinander aufbauen. Ebenso drückt sie aus, dass die Idee des Semantic Webs eine Erweiterung des Webs 2.0 ist und daher auf den bestehenden Strukturen und Inhalten aufbaut.

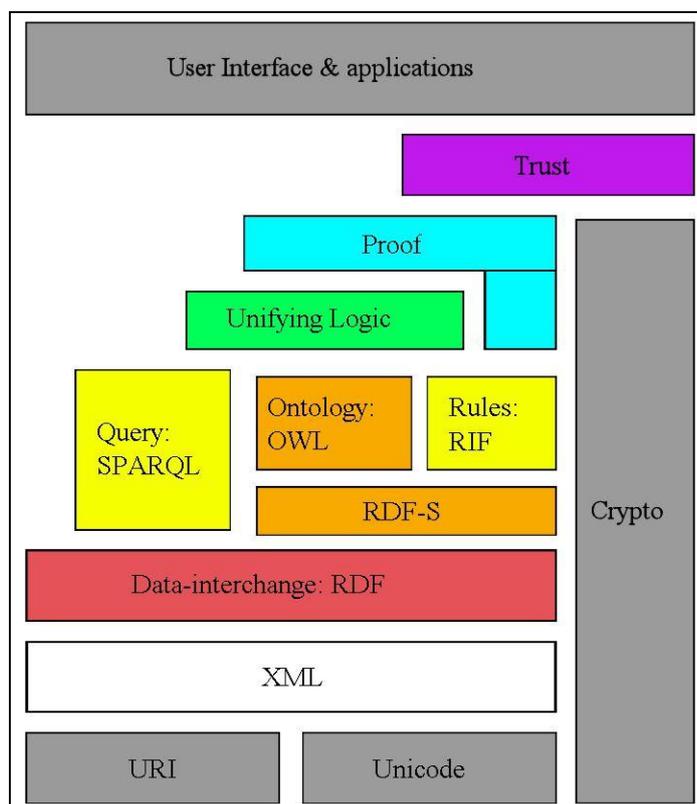


Abbildung 10: Semantic Web Stack

Im gegebenen thematischen Kontext interessieren uns vor allem die formalen Sprachen der semantischen Technologien. Dazu zählen:

⁵⁰ Eingeführt durch Berner-Lee; Abbildungsquelle: http://de.wikipedia.org/wiki/Semantisches_Web, 2009

- Resource Description Framework (RDF)
- RDF Schema (RDFS)
- Web Ontology Language (OWL)
- SPARQL

Resource Description Framework (RDF)⁵¹ ist ein Standard des W3C, welcher auf bereits existierenden Standards wie URI oder XML basiert. Dieser Standard wurde 2004 durch das W3C veröffentlicht. RDF wurde für die Repräsentation von Metainformationen über im Internet verfügbare Quellen entwickelt. Dazu gehören beispielsweise der Autor, Titel oder der Inhalt von Webseiten. Damit soll es Anwendungen ermöglicht werden, den Inhalt und die Metadaten von Quellen korrekt verarbeiten zu können. RDF stellt eine der Grundsäulen des Semantic Webs dar und findet heute als Standard in zahlreichen Applikationen Anwendung.

Inhaltlich basiert RDF auf Ressourcen, welche jeweils in Relation zueinander stehen können. Strukturell geschieht dies durch die Verwendung sogenannter Tripel. Tripel sind im Zusammenhang mit RDF jeweils aus Subjekt, Prädikat und Objekt aufgebaut. RDF ermöglicht dadurch die Modellierung einfacher Ontologien und die Ableitung von implizitem Wissen (P. Hitzler, 2008).

Resource Description Framework Schema (RDFS)⁵² definiert separate Vokabulare, welche wiederum durch RDF und OWL (teilweise) verwendet werden können. Durch RDFS können Eigenschaften und Klassen definiert werden, welchen Ressourcen zugeordnet werden können. Das Verhältnis von RDFS zu RDF ist vergleichbar mit dem Verhältnis von XML-Schemata zu XML.

⁵¹ Resource Description Framework (RDF, <http://www.w3.org/RDF/>, 2009)

⁵² Resource Description Framework Schema (RDFS, <http://www.w3.org/TR/rdf-schema/>, 2009)

Web Ontology Language (OWL)⁵³ bezeichnet einen Standard des W3C zur Repräsentation von komplexen Zusammenhängen. Dieser Standard wurde 2004 durch das W3C veröffentlicht. Da die Darstellung komplexer Zusammenhänge durch RDF nicht möglich ist, wurde mit OWL eine semantisch ausdrucksstarke Sprache zur Repräsentation von Wissen definiert, welche auf formaler Logik beruht. OWL stellt heute das verbreitetste Werkzeug für die Wiedergabe komplexer Zusammenhänge dar. Dabei muss OWL selbst allerdings in drei Teilsprachen untergliedert werden. Diese Trennung wurde bewusst eingerichtet, um den verschiedenen Ansprüchen auf dem Gebiet der Wissensrepräsentation gerecht zu werden (P. Hitzler, 2008).

OWL-LITE ist eine Teilsprache von OWL DL und OWL FULL, welche über eine vergleichsweise geringe semantische Ausdrucksstärke verfügt. Sie dient daher hauptsächlich zum Erschaffen von Modellen wie Thesauri oder Taxonomien.

OWL-DL (OWL Description Logic) ist eine Teilsprache von OWL FULL, welche OWL-LITE enthält. Sie beinhaltet Einschränkungen in Bezug auf RDF, um eine Beschreibungslogik gewährleisten zu können.

OWL-FULL enthält sowohl OWL-LITE als auch OWL-DL vollständig und ist semantisch sehr ausdrucksstark. Allerdings verzichtet es auf die Einschränkungen von OWL-DL weswegen es als einzige OWL-Teilsprache nicht entscheidungsfähig (*reasonable*) ist.

3.3.4 Architektur von Ontologien

Obwohl Ontologien als einer der Schlüsselbausteine hin zu verbesserter Interoperabilität dienen, können wir derzeit auf diesem Bereich einen gewissen Grad an Heterogenität beobachten. Diese Differenzen beruhen vor allem auf den unterschiedlichen Einsatzzwecken (als kontrolliertes Vokabular bis hin zum Reasoning) und der daraus resultierenden unterschiedlichen technischen Umsetzung als auch der unterschiedlichen konzeptuellen Herangehensweise.

⁵³ Web Ontology Language (OWL, <http://www.w3.org/TR/owl-semantics/>, 2009)

Daher unterscheiden wir vier Stufen der Ontologien, welche zueinander in Relation stehen⁵⁴. Diese Typen der Ontologien unterscheiden sich in der Art und Form des Wissens als auch in ihrer konzeptuellen Struktur. Da diese Ausdrücke englischer Literatur entstammen und auch im deutschsprachigen Kontext hauptsächlich im englischen Original wiedergegeben werden, sollen hier ebenfalls die englischen Originalbegriffe verwendet werden:

- Application Ontology
- Domain Ontology
- Task Ontology
- Top-Level Ontology oder Foundational Ontology

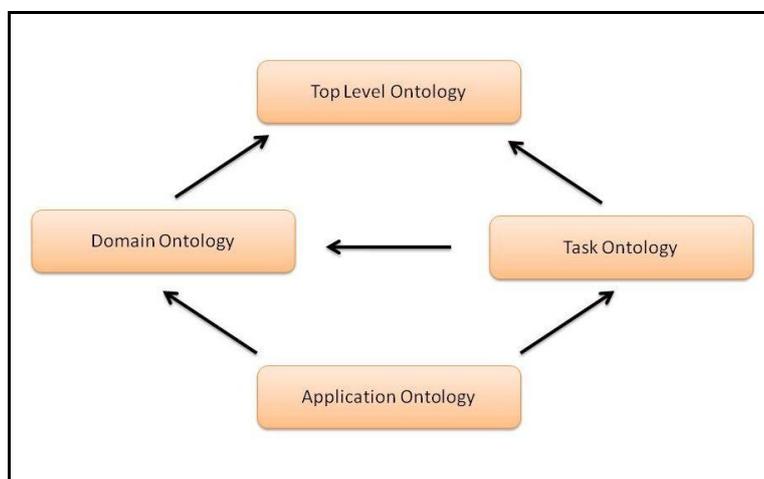


Abbildung 11: Ontologie-Architektur

Der Ausdruck **Application Ontology** (engl. für Anwendungsontologie), auch als „Resource Ontology“ bezeichnet, stellt eine Spezifikation einer Konzeptualisierung eines sehr eingeschränkten Bereiches dar. Diese Form der Ontologien wird in geschlossenen Systemen eingesetzt und ist nicht für den Austausch konzipiert oder frei verfügbar. Sie enthält in der Regel spezifisches Wissen bezüglich anwendungsinterner Vorgänge als formale Repräsentation von Wissen. Anwendungsontologien können auf

⁵⁴ Nach H. Stuckenschmidt, 2009

Task- oder Domainontologies referenzieren und somit deren dargestelltes Wissen nutzen und nach Bedarf verfeinern (Maué, 2009 und Stuckenschmidt, 2009).

Die **Domain Ontology** (engl. für Domänenontologie) bezeichnet die Repräsentation von Wissen bezüglich eines abgegrenzten Bereiches (wie zum Beispiel ein medizinisches Fachgebiet oder Umweltmonitoring). Sie sind in der Regel frei zugänglich und für die weitere Verbreitung des abgebildeten Wissens konzipiert. Domainontologies dienen als höher angesiedelte Spezifikation einer Konzeptualisierung, welche als Referenz für Task- oder Application Ontologies dient. Beispiele für Domain Ontologies sind u. A. GEMET⁵⁵ (technisch gesehen ein Thesaurus) oder SWEET⁵⁶, (Maué, 2009, Stuckenschmidt, 2009) welche beide dem Umweltbereich entstammen sowie CIDOC⁵⁷ aus dem Bereich des Denkmalschutzes und des Bibliothekswesens (Gill, 2004).

Die **Task Ontology** (engl. für Aufgabenontologie) ist in der Literatur nur selten erwähnt und dient als mögliches Bindeglied zwischen der Wissensrepräsentation und Modellen, um die Lücke zwischen der Domain- und der Application Ontology zu überbrücken (Stuckenschmidt, 2009).

Die **Top Level Ontology**, auch Foundational Ontology, Formal Ontology oder Upper Level Ontology genannt, hat die Aufgabe, verschiedene Fachdomänen mit ihren jeweiligen Wissensrepräsentationen zu überbrücken. Diese Form der Ontologie bildet kein Spezial- sondern Allgemeinwissen ab, was sie von den Domain Ontologies unterscheiden. Das Gebiet der Top Level Ontologies befindet sich derzeit noch im Stadium der Forschung und Spezifikation, wodurch keine konkreten erfolgreichen Beispiele genannt werden können (Maué, 2009, Stuckenschmidt, 2009).

⁵⁵ GEMET (GEneral Multilingual Environmental Thesaurus), Thesaurus bezüglich Umweltterminologie der Europäischen Umweltagentur (<http://www.eionet.europa.eu/gemet>, 2009)

⁵⁶ SWEET, Semantic Web for Earth and Environmental Terminology, umfangreiche Ontologie (modularer Aufbau) der NASA bezüglich Umweltterminologie

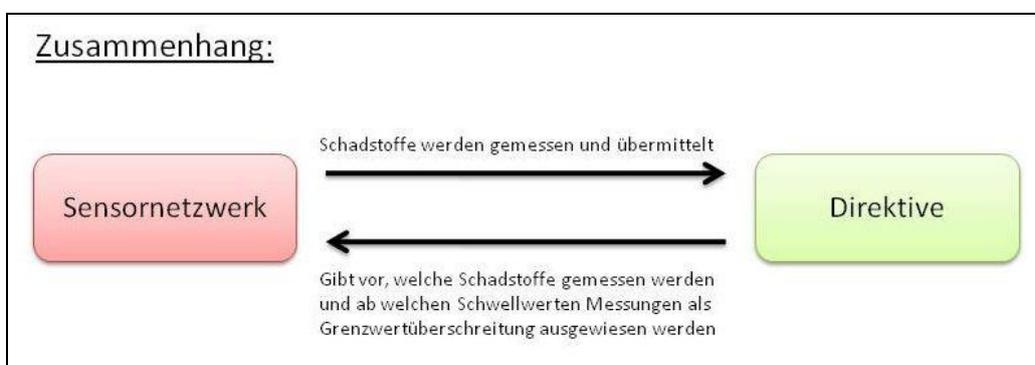
⁵⁷ CIDOC CRM, Domänenontologiekonzept, welches sich dem Bereich des Denkmalschutzes widmet, <http://cidoc.ics.forth.gr/>

3.4 Beispielontologie

Ziel des folgenden Abschnittes ist es, die verwendete Referenzontologie kurz vorzustellen. Dabei soll sowohl auf die Struktur als auch auf die Erstellung der Ontologie kurz eingegangen werden. Anführend muss zudem festgehalten werden, dass diese Applikationsontologie ausschließlich Demonstrationszwecken dient. Sie erhebt keinerlei Anspruch auf thematische Vollständigkeit und ist nicht für den Einsatz von *Reasoning* konzipiert. Sie soll zwei primäre Funktionen erfüllen:

- Im Internet abrufbares Referenzobjekt für die Demonstration von semantischen Annotationen sein
- Den Aufbau und Einsatz von Ontologien im Zusammenhang mit Informationsnetzwerken anhand dieses konkreten einfachen Beispiels verdeutlichen

Die vorliegende Ontologie bildet den stark vereinfachten Prozess der Umweltbeobachtungen ab. Diese Erfassung von Umweltdaten geschieht im Bezug auf eine Richtlinie der Europäischen Kommission zur Lufthygiene⁵⁸ (CAFE). Diese Richtlinie stellt den Kontext und den Zweck dar, weswegen Umweltdaten mittels Sensornetzwerken erfasst werden. In der Ontologie ist somit der einfache Zusammenhang zwischen dem Schadstoff und der CAFE-Richtlinie dargestellt.



⁵⁸ Cleaner Air For Europe (CAFE) Directive

(<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52001DC0245:EN:NOT>, 2009)

Abbildung 12: Zusammenhang Sensornetzwerk - Direktive

Die Kernaussage der Ontologie ist, wie in der Darstellung verdeutlicht, der Zusammenhang zwischen einem Informationsnetzwerk, welches Daten und Informationen transportiert und einer juristischen Verordnung, welche den Anlass gibt, diese Daten überhaupt zu sammeln und zu kommunizieren.

Innerhalb einer Messung werden die gemessenen Eigenschaften (*observed properties*) semantisch annotiert. Damit verweisen sie auf die entsprechenden Konzepte innerhalb der Ontologie. Dies ermöglicht als Folgeschritt weitere Rückschlüsse bezüglich der Bedeutung der gemessenen Werte. Daraus kann im konkreten Beispiel der Rückschluss abgeleitet werden, dass sich die Messung auf die Berichtspflichten durch die CAFE-Richtlinie bezieht. Im Kontext einer Messung, welche mit einem OGC-Standard wie *Observation and Measurement* abgebildet ist, können solche Zusammenhänge kaum oder nur schwer abgebildet werden. Selbst mit einem semantisch ausdrucksstarken Standard wie *Observation and Measurement* können komplexe Zusammenhänge nur sehr begrenzt abgebildet werden. Daher dient die vorliegende Demonstrationsontologie als ergänzendes Wissensmodell, welches bei Bedarf ausgewertet werden kann.

In dem Versuch, den aufgezeigten Zusammenhang auszudrücken, folgt diese Ontologie der *Open World Assumption*. Diese Annahme ist das Gegenteil der *Closed World Assumption* und bezeichnet alles als unbekannt, was nicht innerhalb dieses Modelles abgebildet ist. Daher muss im konkreten Modell die Frage, ob auch NO_x⁵⁹ durch die CAFE-Direktive erforderlich ist, als „unbekannt“ beantwortet werden, da dieses Wissen nicht im Modell abgebildet wurde.

Auf die technischen Aspekte der Beispielontologie soll hierbei nicht weiter eingegangen werden. Sie ist mittels *Protégé*⁶⁰ erstellt worden und lehnt sich in ihrem Grundkonzept an *SWEET2*⁶¹ der NASA an, welches allerdings erheblich umfangreicher ist⁶².

⁵⁹ NO_x, Sammelbegriff für verschiedene Formen von Stickstoffoxiden

⁶⁰ Protégé, Open Source Ontologie Editor (<http://protege.stanford.edu/>, 2009)

⁶¹ SWEET2 Ontologie der NASA liegt derzeit (Februar 2010) als Beta-Version vor (<http://sweet.jpl.nasa.gov/ontology/>, 2009)

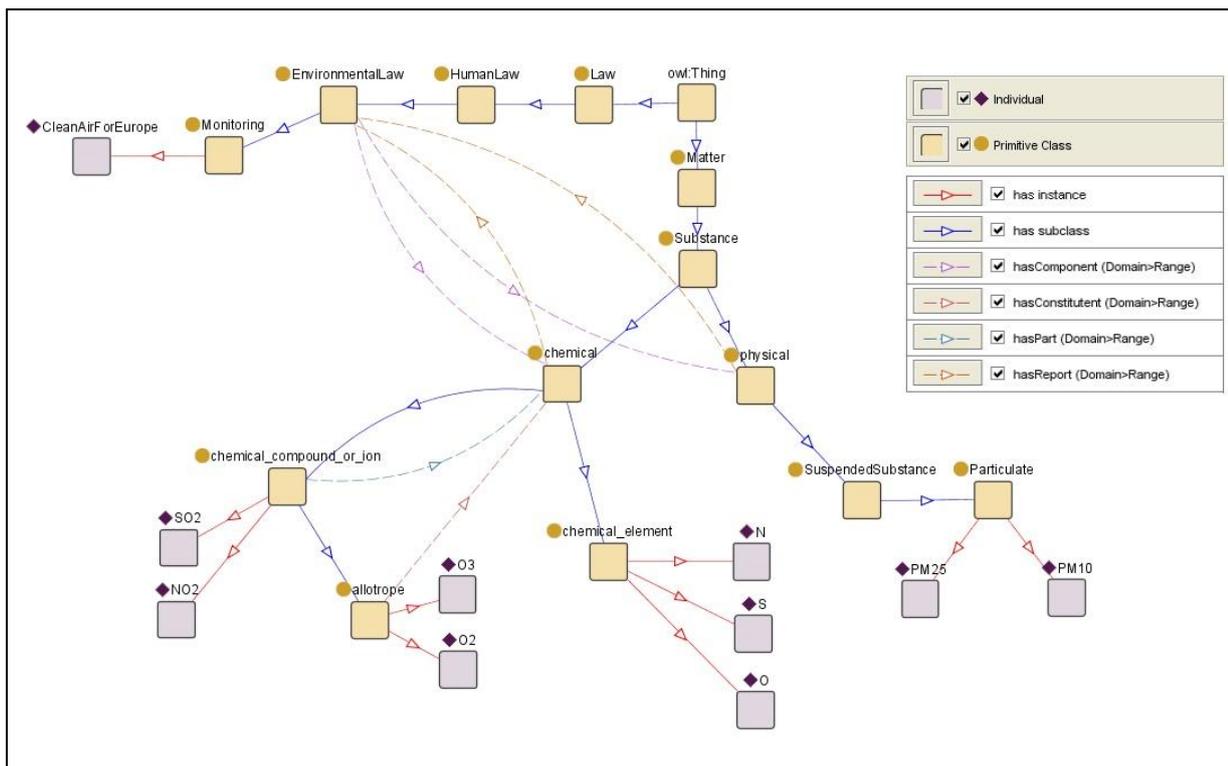


Abbildung 13: Beispielontologie – CAFE

Die folgende Abbildung wurde durch ein Plug-In innerhalb von *Protégé (Jambalaya)* erstellt und stellt grafisch die Zusammenhänge zwischen den enthaltenen Klassen und Instanzen dar. Die eigentliche Ontologie wird jedoch durch OWL repräsentiert und findet sich sowohl im Anhang dieser Arbeit als auch unter http://www.katzengeist.de/attachments/File/CAFE_Phenomena.owl.

⁶² SWEET2 umfasst über 4100 Konzepte (hauptsächlich aus dem Bereichen Umweltphänomene)

4 Semantische Annotation

Das Ziel dieses Kapitels ist es, die Methodik der semantischen Annotation vorzustellen. Dies geschieht an konkreten Beispielen aus dem Bereich des Umweltmonitoring. Neben der eigentlichen Methodik der Annotation sollen innerhalb dieses Kapitels die Sensor Web Enablement (SWE) Spezifikationen der OGC vorgestellt werden. Diese XML-Spezifikationen stellen die syntaktische Grundlage und den Konterpart für die dargestellten Annotationsbeispiele dar.

4.1 Sensor Web Enablement

Hinter dem Namen *Sensor Web Enablement* (SWE)⁶³ verbirgt sich eine Initiative der OGC zur Standardisierung. Wie bereits aus dem Namen hervorgeht, stehen Sensoren und Sensordaten im Fokus der Bestrebungen von SWE. Der Name soll allerdings nicht den Eindruck erwecken, dass die darin enthaltenen Technologien nur auf direkte Sensordaten beschränkt sind. Vielmehr bieten sich die Spezifikationen und Standards allgemein für die Kommunikation von Beobachtungen und Messungen an. Durch diese Initiative wurde ein Rahmen für zahlreiche Standard und Spezifikationen geschaffen, welcher sowohl Datenmodelle als auch entsprechende Webservices enthält. Dazu gehören⁶⁴:

- Observation and Measurement (O&M)
- Sensor Model Language (SensorML)
- Transducer Markup Language (TML)
- Sensor Observation Service (SOS)
- Sensor Planning Service (SPS)
- Sensor Alert Service (SAS)⁶⁴

⁶³ Sensor Web Enablement (SWE, <http://www.opengeospatial.org/ogc/markets-technologies/swe>, 2009)

⁶⁴ Zum Zeitpunkt März 2010

- Web Notification Service (WNS)⁶⁵

Diese Spezifikationen dienen der Interoperabilität von Sensordaten und damit verbundenen Beobachtungen. Allerdings sind weder Daten noch Services an Sensoren gebunden oder auf diese beschränkt, da jegliche mit einer Messung oder Beobachtung verbundenen Informationen über diese Standards kommuniziert werden können. Sensoren⁶⁶, wie im Namen der Spezifikationen erwähnt, spielen dabei keine direkte Rolle. Vielmehr geht es um die Kommunikation von Messungen und Beobachtungen jeglicher Art, welche unter anderem auch mittels Sensoren erfasst werden können. Dabei spielt die räumliche Komponente, welche bei bisherigen Spezifikationen der OGC im Mittelpunkt stand, eine vergleichsweise untergeordnete Rolle.

Es wird an dieser Stelle auf weitere Details und Einzelheiten bezüglich SWE verzichtet, da dies mit der semantischen Annotation nicht direkt in Zusammenhang steht. Jedoch sollen im Folgenden die beiden hauptsächlich in SWE verwendeten Datenmodelle kurz vorgestellt werden, da diese die syntaktische Grundlage für die später vorgestellte semantische Annotation darstellen.

Die *Sensor Web Enablement Suite* der OGC ist zudem für die Arbeit relevant, da sowohl durch *Observation and Measurement* als auch durch die *Sensor Model Language* erstmals Elemente der semantischen Annotation direkt in OGC Standards vorgeschlagen werden. Beschränkten sich bisherige Standards wie die *Geographic Markup Language* (GML) darauf, eine syntaktische Interoperabilität zu gewährleisten, so berücksichtigen die Spezifikationen des SWE-Rahmens die bisher nicht behandelte semantische Interoperabilität. Die Datenmodelle des SWE-Rahmens werden (derzeit) vor allem im Bereich der Umweltinformatik eingesetzt, welche eine klare Terminologie und einen hohen Grad an Metainformationen benötigt. Obwohl weite Teile von O&M wie auch SensorML generisch sind, ist es schwer alle - für wissenschaftliche Prozesse oder Umweltmonitoring relevanten - Informationen darin zu modellieren. Durch die Ansprüche dieser Form der Messungen, dessen Metainformationen sowohl rechtliche

⁶⁵ Sowohl der *Sensor Alert Service* als auch der *Web Notification Service* sind bisher keine offiziellen Standards des OGC, Stand März 2010

⁶⁶ Der Begriff des „Sensor“ ist dabei nicht eng definiert, da auch menschliche Beobachter als Sensoren operieren können

Aspekte als auch naturwissenschaftliche Parameter enthalten können, bietet sich die Nutzung von externen Quellen der Wissensrepräsentation an, auf welche im Bedarf referenziert werden kann.

```

...
<om:parameter>
  <swe:Quantity definition = "http://sweet.jpl.nasa.gov/ontology/property.owl#Temperature">
    <!-- Beispiel für eine semantische Annotation -->
    <swe:uom xlink:href = "urn:ogc:def:uom:UCUM:Cel"/>
    <swe:value>22.3</swe:value>
  </swe:Quantity>
</om:parameter>
...

```

Abbildung 14: Semantische Annotation in *OGC Observation and Measurement* Dokumentation

Die Abbildung 14⁶⁷ zeigt die Verwendung eines Verweises auf die bereits erwähnte SWEET Ontologie der NASA⁶⁸ durch die offizielle OGC Spezifikation bezüglich *Observation and Measurement*. Konkret zeigt das Beispiel das Element *<Parameter>*, welches anstatt der Bezeichnung des einfachen Namens des Parameters einen Zeiger auf das Konzept „*Temperature*“ der besagten SWEET Ontologie enthält.

Wie in der nachfolgenden Abbildung 15 ersichtlich, ist das Konzept „Temperatur“ ein Konzept, welches noch weitere Unterkonzepte besitzt. Die dort angesprochene *property.owl* ist eine Subontologie, welche wiederum u. A. das Konzept *Temperatur* enthält. Diese Grafik soll einen Einblick sowohl in die Komplexität als auch in das Potential der semantischen Annotation bieten. Die graphische Abbildung der Relationen und Zusammenhänge erfolgte unter Verwendung des Ontologie-Editors *Protégé* und dem entsprechenden Plug-In *Jambalaya*.

⁶⁷ XML-Beispiel entstammt S. Cox, 2007

⁶⁸ Semantic Web for Earth and Environmental Terminology (SWEET), <http://sweet.jpl.nasa.gov/ontology/>

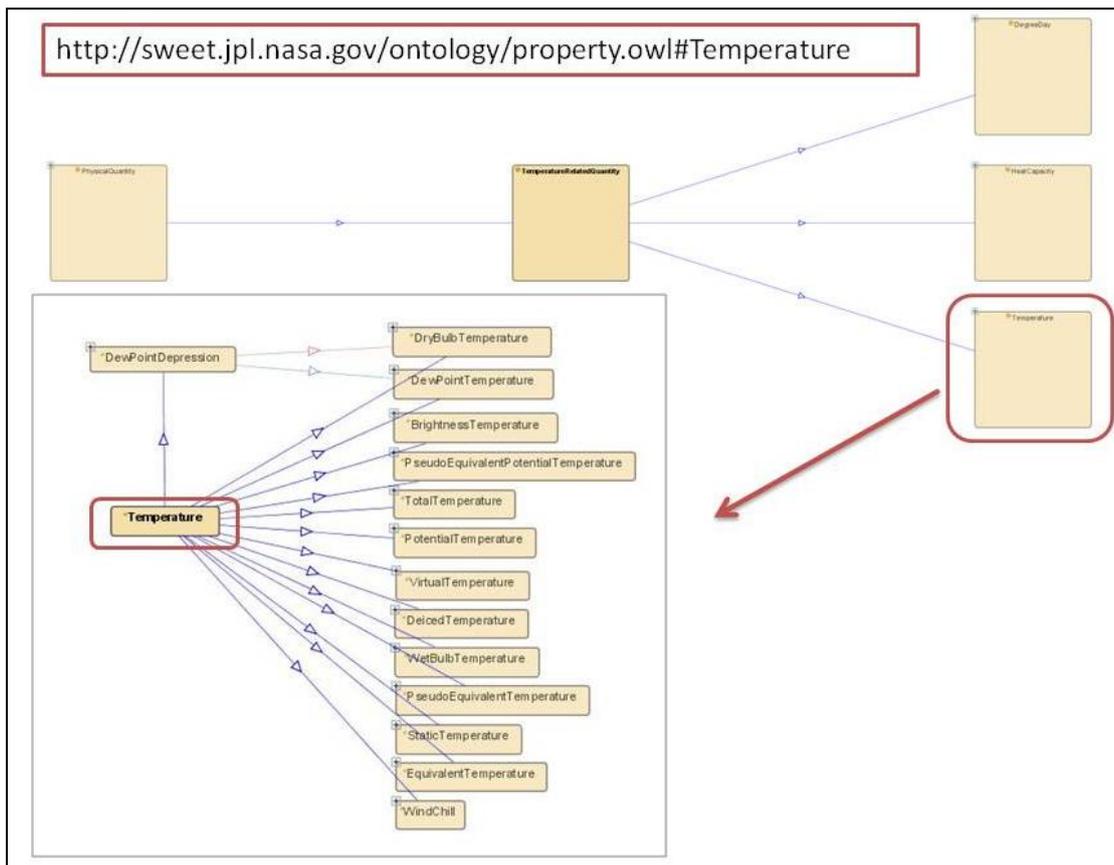


Abbildung 15: „Temperature“ innerhalb der NASA SWEET Ontologie

4.1.1 Observation and Measurement

Die OGC XML-Spezifikation „*Observation and Measurement*“⁶⁹ (O&M) bezeichnet einen OGC Standard, welcher ein Datenmodell zur Kommunikation von Beobachtungen und Messungen darstellt. Obwohl O&M innerhalb des SWE - Rahmens entwickelt wurde, ist dieser Standard keineswegs auf das Feld der Sensordaten beschränkt. Er ist, indem eran einem semantischen Grundmodell orientiert ist, in der Lage, selbst einfachste Beobachtungen⁷⁰ basierend auf einer Fragestellung und einem Resultat auszudrücken. Da diese Spezifikation teilweise sehr generisch und semantisch ausdrucksstark ist, deckt sie ein sehr breites Spektrum an Anwendungsmöglichkeiten

⁶⁹ Observation and Measurement, OGC Standard (<http://www.opengeospatial.org/standards/om>, 2009)

⁷⁰ z.B. „Welche Farbe hat der Apfel?“

ab. Die Dokumentation von O&M ist in zwei getrennte Bereiche unterteilt, die eigentliche Beobachtung⁷¹ und die Verortung⁷², welche getrennt voneinander betrachtet werden können. Während durch den Bereich „Beobachtung“ das semantische Grundkonzept der wissenschaftlichen Messung (u. A. Gemessene Eigenschaft, Zeit, Parameter, Resultat) definiert wird, werden durch den Bereich Verortung die Strukturen der Ortsangabe der Messung definiert.

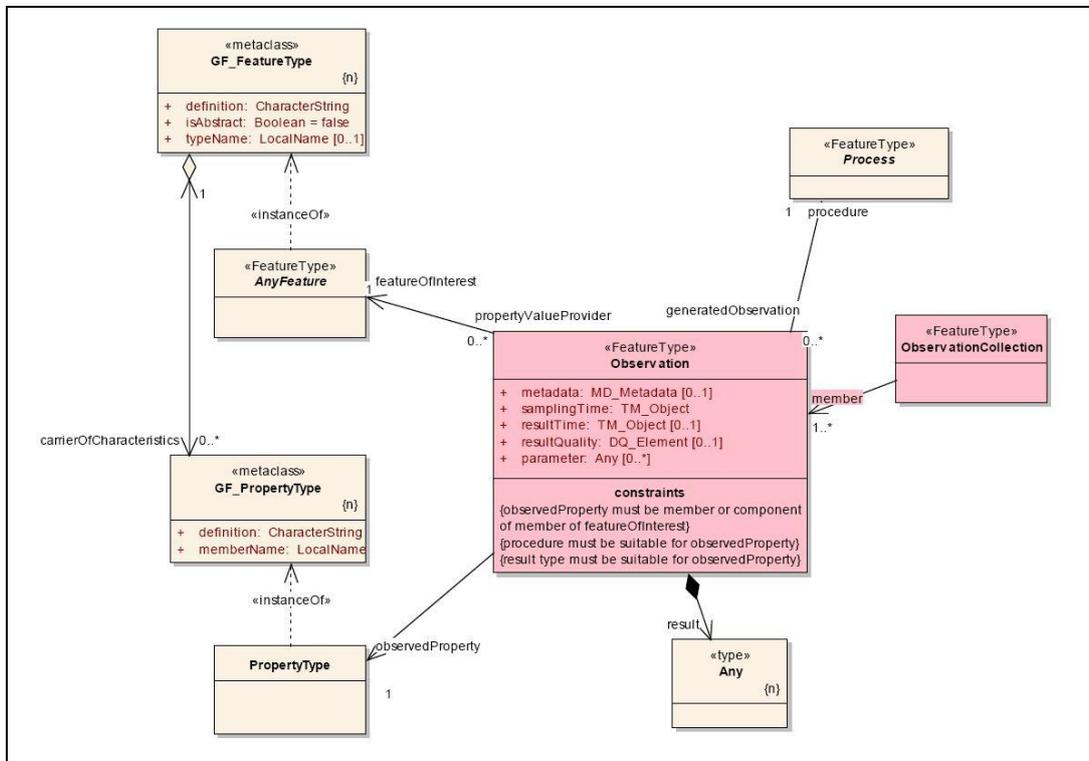


Abbildung 16: Observation-Objekt in Observation and Measurement

Die obige Abbildung 16 gibt die definierte Struktur der Beobachtung (*Observation*) wieder, welche unter Anderem aus folgenden Elementen und Eigenschaften besteht.

- Beobachtung (*Observation*)

⁷¹ Observations and Measurements - Part 1 - Observation schema
(http://portal.opengeospatial.org/files/?artifact_id=22466, 2009)

⁷² Observations and Measurements - Part 2 - Sampling Features
(http://portal.opengeospatial.org/files/?artifact_id=22467, 2009)

- Zeit (*Time*)
- Parameter
- Verortung (Feature of Interest)
- Gemessene Eigenschaft (*Observed Property*)
- Prozessinformation (*Procedure*)
- Resultat (*Result*)

Für das Verständnis der weiteren Zusammenhänge sei hier erwähnt, dass für die Darstellung der Prozessinformationen (*Procedure*) jeder Beobachtung die XML-Spezifikation *SensorML* verwendet wird. Es wird an dieser Stelle darauf verzichtet ausführlicher auf Details bezüglich der *Observation and Measurement XML*-Spezifikation einzugehen. Das XML-Schema wie auch beschreibende Beispiele werden im Anhang dieser Arbeit aufgeführt (S. Cox, 2007).

4.1.2 Sensor Model Language

Die Sensor Model Language⁷³ (SensorML) ist ebenfalls eine XML-Spezifikation von SWE, welche Konzepte und XML-Kodierung für die Abbildung von Sensor- und Messprozessen zur Verfügung stellt. Wie O&M wurde auch SensorML nicht direkt durch die OGC entwickelt, sondern verfügt über eine eigene Entwicklungsgeschichte außerhalb der OGC. SensorML wird besonders im Bereich der Umweltüberwachung eingesetzt, da es hochgradig generisch ist und damit umfangreiche Beschreibungen von komplexen Prozessen ermöglicht.

SensorML wurde 2007 durch die OGC als Standard aufgenommen und ist Bestandteil des SWE-Rahmens. Im Zusammenhang mit *Observation and Measurement* werden

⁷³ Sensor Model Language (SensorML, <http://www.opengeospatial.org/standards/sensorml>, 2009)

durch SensorML Prozesse und Metainformationen wiedergegeben, welche damit in Verbindung stehen.

Wie die bereits erwähnte Spezifikation *Observation and Measurement* schlägt SensorML die Verwendung von semantischen Annotationen innerhalb seiner generischen Strukturen vor (M. Botts, 2007). Das entsprechende Schema von SensorML und Beispiele sind im Anhang angefügt.

4.2 Ebenen und Anwendungen der semantischen Annotation

Für den Einsatz von semantischen Annotationen innerhalb von Geodateninfrastrukturen bieten sich verschiedene Ebenen an, auf welchen derartige Verweise auf externe Wissensmodelle sinnvoll erscheinen. Diese Ebenen sind mit konkreten Anwendungsfällen verbunden, durch welche sie identifiziert werden. Ziel dieses Abschnittes ist es daher, anhand der Anwendungsfälle zu analysieren, in welchen Bereichen der Geodateninfrastruktur Annotationen sinnvoll sein können.

Die Analyse der unterschiedlichen Ebenen der semantischen Annotation soll mit einer Aufstellung der Anwendungsfälle beginnen, in denen semantische Annotation in Zusammenhang mit Geoinformatik relevant erscheint. Die Zusammenstellung der Anwendungsfälle orientiert sich an Bommersbach (2009) und Mauè (2009). Einige dieser Prozesse und Anwendungsfälle wurden zudem bereits im Kapitel Geosemantik und Interoperabilität in der vernetzten Geoinformatik besprochen. Die hier folgende Auflistung von Anwendungsfällen erhebt keinen Anspruch auf Vollständigkeit. Sie ist vielmehr ein Versuch, anhand repräsentativer Anwendungsfälle abzuleiten, auf welchen Ebenen semantische Annotationen sinnvoll und nötig sind (Maué, 2009, Bommersbach, 2009, Klien, 2009):

- Der Prozess der Suche nach geographischen Informationen und Diensten
- Die Validierung und Konsistenzprüfung von geographischen Informationen
- Die Validierung von Prozessketten
- Die Integration von geographischen Daten

4.2.1 Suche nach geographischen Informationen und Diensten

Der Prozess der Suche nach geographischen Informationen und Diensten wurde bereits im genannten Kapitel beschrieben. Die besondere Herausforderung bei der Suche nach geographischen Daten liegt in deren Heterogenität. Die gesuchten Daten können Ausprägungen in den verschiedensten Dimensionen aufweisen, wie etwa räumlich, zeitlich oder thematisch. (Maué, 2009) Durch Standards der OGC sind die Dimensionen des Raumes und der Zeit im Suchprozess handhabbar und eindeutig ansprechbar geworden. Dadurch ist die Ebene der Metadaten ein wichtiger Bestandteil existierender und zukünftiger Geodateninfrastrukturen geworden⁷⁴. Auf dem Bereich der Thematik der Datensätze sehen wir uns allerdings nach wie vor mit dem besprochenen Problem der fehlenden semantischen Interoperabilität konfrontiert. Hierbei ist vor allem die unzureichende thematische Beschreibung existierender Datensätze als Hauptursache zu benennen. Während durch die existierenden Metadatenrichtlinien eine Struktur und damit eine syntaktische Interoperabilität vorgegeben werden, bleiben die thematischen Ansprüche weitgehend unberücksichtigt. Für die Archivierung, Verwaltung und Bereitstellung von Metadaten werden Katalogdienste⁷⁵ verwendet, welche es dem Nutzer erlauben nach Informationen bestimmter Stichworte und Parameter wie Zeit und Raum zu suchen. Dabei beschränken sich die Angaben der thematischen Beschreibung innerhalb der Metadatenprofile⁷⁶ zumeist auf Schlüsselwörter, Kategorisierungen und Freitexte. Berücksichtigt man die Vielfalt der Fachgebiete, in welchen geographische Informationen eingesetzt werden, dann lässt dies erkennen, dass die jeweiligen Spezifikationen der Metadatenprofile einer aussagekräftigen Beschreibung der Dateninhalte nicht gerecht werden können. Hierbei ist hinzuzufügen, dass sich

⁷⁴ z.B. der Discovery-Service und Registry-Service stellt innerhalb von INSPIRE die Metaebene in einer europaweiten Geodateninfrastruktur dar

⁷⁵ z.B. Catalogue Service Web (CSW), Spezifikation der OGC, (<http://www.opengeospatial.org/standards/cat>, 2009)

⁷⁶ Metadatenprofile bzw. Metadatenformate wie ISO19115/19139, Dublin Core, FGDC oder ebRIM

Metadaten im Kontext der Geoinformatik nicht nur auf Daten, sondern auch auf Dienste beziehen können, welche ebenso gesucht und beschrieben werden. (Klien, 2009)

Ausgehend von dem Anwendungsfall – der Suche nach geographischen und thematischen Informationen – bietet sich hier die semantische Annotation an, um die angesprochenen Lücken der thematischen Beschreibung, welche durch die Struktur der Profile nicht ermöglicht wird, zu überbrücken. Die erweiterte Semantik auf der **Ebene der Metadaten** kann aus heutiger Sicht am sinnvollsten durch die Annotation von Schlüsselwörtern (engl. *Keywords*) erreicht werden. Die Annotation hat hierbei die Aufgabe, die Bedeutung semantisch unscharfer Schlüsselwörter durch den Verweis auf Thesauri und Ontologien zu unterstützen und somit dem Nutzer klarere Antworten auf seine Suchanfragen zur Verfügung zu stellen. Durch das Anwenden von *Reasoning* können dem Datensuchenden erweiterte Suchoptionen geboten werden. Zudem würde sich die Anzahl der effektiven Einträge durch *Reasoning* steigern lassen. (Maué, 2009) Zusammenfassend kann hierbei festgehalten werden, dass – im Kontext der Suche nach Information – die semantische Annotation auf der Ebene der Metadaten notwendig und sinnvoll ist. Dabei wird versucht, die semantische Aussagekraft der eingesetzten Schlüsselwörter zu erweitern.

Die semantische Annotation auf der Ebene der Metadaten ist die verbreitetste Form dieser Technologie, da sie sich innerhalb der bereits bestehenden Spezifikationen leicht umsetzen lässt. Heute finden wir zahlreiche Beispiele für die semantisch unterstützte Suche nach geographischer Information. Zu nennen sind unter Anderem das bereits erwähnte Projekt SWING oder das ORCHESTRA⁷⁷-Projekt, welches sich mit Katastrophenmanagement beschäftigt.

4.2.2 Validierung und Konsistenzprüfung von geographischen Informationen

Die Validierung von geographischen Informationen bezieht sich auf die tatsächliche Nutzbarkeit von geographischen Informationen für die Ansprüche des Nutzers und

⁷⁷ ORCHESTRA, EU-gefördertes Projekt (<http://www.eu-orchestra.org/>, 2009)

innerhalb von Systemen. Dies kann sich unter Anderem auf technische oder rechtliche Spezifikationen sowie auf Aspekte der Sicherheit der Daten beziehen. (Maué, 2009)

Die Nutzbarkeit von Daten für bestimmte, definierte Zwecke hat – wie beschrieben - unterschiedlichste Aspekte, welche sich auf verschiedene Ebenen der Geodateninfrastruktur auswirken. Zu berücksichtigen sind neben den Daten, welche (nur teilweise) durch Metadaten beschrieben werden, auch beteiligte Dienste (aus Gründen der syntaktischen Interoperabilität), welche Daten anbieten oder weiterverarbeiten. Daher sind die semantischen Annotationen in diesem Fall sowohl auf der **Ebene der Metadaten** (Metadatenformate) als auch auf der **Ebene der Dienste** selbst (Prozess- und Dienstbeschreibung, z.B. durch *Web Service Description Language*⁷⁸) sinnvoll und notwendig.

Zum Anwendungsfall der Validierung von geographischen Daten gehört neben den Aspekten der Nutzbarkeit auch die Datenkonsistenz. Hierbei spielen die semantische Annotation von geographischen Daten eine große Rolle, da sie auf komplexe externe Modelle und Regeln verweisen können, welche dem Nutzer sonst möglicherweise verschlossen bleiben. Diese Regeln können themenspezifisch oder topologisch sein. Diese teils individuellen Beziehungen können durch Ontologien abgebildet werden, wodurch auch komplexe Konsistenzprüfungen von Daten und Informationen automatisiert erfolgen können. Diese Annotationen können auf der **Ebene der Dienste**, der **Datenschemen** und Formate sowie auf der **Ebene der Dateninstanz** selbst eingesetzt werden.

4.2.3 Validierung von Prozessketten

Neben der Validierung von geographischen Daten stellt die Validierung von Prozessketten einen wichtigen Anwendungsfall der semantischen Annotation dar. Unter Prozessketten sind dabei die Interaktion und Kommunikation zwischen Diensten zu verstehen, welche automatisiert und ohne individuelle Steuerung durch Nutzer

⁷⁸ Web Service Description Language (WSDL), Spezifikation des W3C (<http://www.w3.org/TR/wsd120/>, 2009)

interoperieren sollen. Dieser Anwendungsfall trifft den eigentlichen Kern der Ideen und Visionen des Semantic Webs, da er die Interoperabilität zwischen Maschinen bzw. Agenten zum Ziel hat.

Innerhalb dieses Prozesses lassen sich verschiedene Komponenten identifizieren, in denen eine semantische Annotation möglich und sinnvoll erscheint. Auch hier kommen dafür die Ebene der Prozessbeschreibung, der Datenschemen sowie die Ebene der Dateninstanz in Frage.

Für die Validierung und Interoperabilität zwischen verschiedenen Diensten, wie zum Beispiel dem Web Feature Service⁷⁹ oder dem Web Processing⁸⁰ Service, ist es vor allem notwendig, die syntaktische Interoperabilität zwischen den Diensten zu gewährleisten. Obwohl die Dienste der OGC mit dem Ziel der syntaktischen Interoperabilität entwickelt wurden, unterscheiden sich diese Dienste jedoch in ihren Implementierungen. Daher können auf der Ebene der Dienstbeschreibung semantische Annotationen eingesetzt werden, um die mit den Diensten zusammenhängende und nicht in *Web Service Description Language* (WSDL) abgebildete Information durch externe Quellen bereitzustellen.

Neben der Prozessbeschreibung sind die verschiedenen Datenformate und Dateninstanzen weitere Komponenten, welche validiert werden können. Bezüglich der Datenformate stellt sich vor allem die Frage der strukturellen Konformität aber auch die der Semantik. Der Aspekt der strukturellen Konformität von Daten betrifft sowohl die Kommunikation mit verschiedenen Diensten als auch die Interoperabilität zwischen anderen Datenformaten.

Die Bedeutung der Informationen innerhalb von Prozessketten bezieht sich auf den bereits beschriebenen möglichen Verlust an Semantik durch Verwendung

⁷⁹ Web Feature Service (WFS), Spezifikation des OGC (<http://www.opengeospatial.org/standards/wfs>, 2009)

⁸⁰ Web Processing Service (WPS), Spezifikation des OGC (<http://www.opengeospatial.org/standards/wps>, 2009)

unterschiedlicher Terminologien⁸¹. Die Bedeutung der Dateninstanzen, welche bestimmten Strukturen entsprechen, kann dabei an verschiedenen Stellen angesprochen und definiert werden. Dies ist abhängig vom Konzept des Datenformates. Da wir uns auf die Spezifikation *Observation and Measurement* fokussieren, soll das aufgeführte Beispiel unten zeigen, wie die Bedeutung eines Ausdrucks (Value) auf der Ebene der Dateninstanz gelöst werden kann.

```
<?xml version="1.0" encoding="UTF-8"?>
<om:Observation
  gml:id = "obsTest1"
  xmlns:om = "http://www.opengis.net/om/1.0"
  xmlns:swe = "http://www.opengis.net/swe/1.0.1"
  xmlns:xsi = "http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xlink = "http://www.w3.org/1999/xlink"
  xmlns:gml = "http://www.opengis.net/gml"
  xsi:schemaLocation = "http://www.opengis.net/om/1.0 ../om.xsd">
  <gml:description>Observation test instance: fruit mass</gml:description>
  <gml:name>Observation test 1</gml:name>
  <om:samplingTime>
    <gml:TimeInstant gml:id = "ot1t">
      <gml:timePosition>2005-01-11T16:22:25.00</gml:timePosition>
    </gml:TimeInstant>
  </om:samplingTime>
  <om:procedure xlink:href = "http://www.flakey.org/register/process/scales34.xml"/> <om:observedProperty xlink:href =
"urn:ogc:def:phenomenon:OGC:mass"/> <om:featureOfInterest xlink:href =
"http://wfs.flakey.org?request=getFeature&featureid=fruit37f"/> <om:parameter>
  <swe:Quantity definition = "http://sweet.jpl.nasa.gov/ontology/property.owl#Temperature">
    <swe:uom xlink:href = "urn:ogc:def:uom:UCUM:Cel"/>
    <swe:value>22.3</swe:value>
  </swe:Quantity><!--example of optional soft-typed parameter -->
  </om:parameter>
  <om:result xsi:type = "gml:MeasureType" uom = "urn:ogc:def:uom:OGC:kg">0.28</om:result>
</om:Observation>
```

Abbildung 17: XML-Beispiel - die Semantik der Elemente wird auf Datenebene abgebildet

Wir können im obigen XML-Beispiel⁸² erkennen, dass zum Beispiel die Elemente *<om:observedProperty>* oder *<om:parameter>* als Konzepte für weitere Elemente stehen, welche dann die eigentliche Bedeutung des beobachteten Wertes ausdrücken. Diese Form der Abbildung von Semantik auf Ebene der Dateninstanz ist ein Merkmal der generischen Datenmodelle wie O&M und SensorML.

Als Gegenstück zur Definition der Bedeutung auf der Ebene der Dateninstanz gilt die Umsetzung von Datenformaten durch XML-Schemen, wie am unten aufgeführten Beispiel zu erkennen ist.

⁸¹ Siehe Kapitel Semantische Interoperabilität und semantische Lücke zum semantischen Konflikt des Wortes „Mine“

⁸² Das XML-Beispiel entstammt S. Cox, 2009

```

...
<complexType>
  <sequence minOccurs="1" maxOccurs="unbounded">
    <element name="urn:ogc:def:phenomenon:OGC:mass" type="xs:double" />
  </sequence>
</complexType>
...

```

Abbildung 18: XML-Schemata-Beispiel – die Semantik wird auf Ebene der Schemata abgebildet

In diesem Beispiel ist zu erkennen, dass die Bedeutung eines Elementes innerhalb eines XSD-Dokumentes vorgenommen wird, welches wiederum die Struktur aber auch (im konkreten Beispiel) die Semantik der Dateninstanz definiert.

Die semantische Aussage kann also in Abhängigkeit der Konzeption der Datenschemen sowohl auf der Ebene der Datenformate als auch auf der Ebene der Dateninstanz angesprochen und somit semantisch annotiert werden. Daher sind beider Ebenen potentielle Träger der semantischen Annotation durch diesen konkreten Anwendungsfall.

4.2.4 Integration von geographischen Daten

Dieser letzte beschriebene Anwendungsfall ähnelt dem im Kapitel Validierung von Prozessketten bereits beschriebenen. Unter der Integration von geographischen Daten ist die Zusammenführung und notwendige Transformation von verschiedenen Datenquellen in ein integriertes System zu verstehen. Dies setzt voraus, dass die entsprechenden Datenquellen in Geometrie, Struktur und Semantik interoperabel sind, was bedeutet, dass sie beispielsweise das idente räumliche Gebiet abdecken, innerhalb eines entsprechenden Zeitfensters liegen, strukturell miteinander interoperabel sind oder dass es keine semantischen Lücken gibt (Maué, 2009). Durch die Annotation auf Prozessebene wie auch auf Ebene der Datenschemen kann die strukturelle Interoperabilität gewährleistet werden. Hierzu gehört auch die notwendige Transformation (engl. *Mapping*) der Daten, bei der der Einsatz von Ontologien hilfreich sein kann, um die Umwandlung zwischen den verschiedenen Datenformaten und Schemen zu unterstützen.

Durch die Verwendung von semantischer Annotation und Wissensrepräsentation lassen sich Hürden, wie nicht ausreichende semantische Beschreibung oder differierende Datenformate, bei der Integration von geographischen Informationen bewältigen.

Systeme, welche semantische Annotation auflösen können, wären in der Lage, den Nutzer vor drohenden Konflikten zu warnen oder könnten durch eingebaute Regeln versuchen, Konflikte selbstständig zu lösen (Maué, 2009)

4.2.5 Zusammenfassung der möglichen Ebenen der semantischen Annotation

Durch die aufgeführten Anwendungsbeispiele werden verschiedene Ebenen identifiziert, welche sich für die semantische Annotation anbieten. Diese Ebenen entsprechen jeweils unterschiedlichen Anwendungsfällen.

- **Metadatenbeschreibungen** (u. A. durch Annotation von Schlüsselwörtern)
- **Dienstbeschreibungen** (u. A. durch Verwendung von SAWSDL⁸³)
- **Datenschemen** (u. A. durch Verwendung von SAWSDL)
- **Dateninstanz**

Es ist das Ziel der weiteren Ausführungen, die technische Umsetzung semantischer Annotationen auf der jeweiligen Ebene zu diskutieren und zu bewerten.

4.3 Umsetzung von semantischen Annotationen

Ziel dieses Kapitels ist es, die verschiedenen Möglichkeiten der semantischen Annotation technisch an Beispieldatensätzen zu demonstrieren. Dabei bezieht sich die Diskussion auf Annotationen der in Abschnitt 4.2.5 identifizierten Komponenten der Geodateninfrastruktur. Da hierbei der besondere Fokus auf den Spezifikationen *Sensor Web Enablement* liegt, wurde ein besonderer Schwerpunkt auf die Ebenen der Metadaten und Dateninstanzen gelegt. Im Zusammenhang mit den

⁸³ Semantic Annotation for Web Description Language and XML Schema (SAWSDL), WC3 Specification (<http://www.w3.org/2002/ws/sawSDL/>, 2009)

Annotationsmöglichkeiten in XML-Datenschemata wird zudem der W3C - Standard *Semantic Annotations for WSDL and XML Schema* vorgestellt werden.

Eine weitere wichtige Tatsache ist die, dass nur die semantische Annotation diskutiert und vorgestellt wird. Für eine Verarbeitung der Annotation, das Auflösen des Pfades sowie entsprechende Rückschlüsse aus der referenzierten Ontologie, ist eine entsprechende Applikation notwendig. Solche Applikationen können semantische Kataloge oder Agenten in integrierten Systemen sein, welche den Umfang dieser Arbeit sprengen und nicht im Kontext der Standardisierung der semantischen Annotation stehen würden.

4.3.1 Semantische Annotation von Metadaten

Zur Einführung in diesen Abschnitt muss festgehalten werden, dass der Begriff „Metadaten“ sich hierbei auf den Kontext der Geoinformatik bezieht. In diesem konkreten Zusammenhang sind „Metadaten“ aus historischen Gründen scharf von den eigentlichen „Daten“ abgegrenzt. In anderen Bereichen der Informationswissenschaften ist der Begriff der Metadaten weit weniger etabliert und ist strukturell oft nicht von den eigentlichen Daten zu trennen⁸⁴. Die Beschreibung der Annotation auf der Ebene der Metadaten bezieht sich zudem auf den Suchprozess (durch den Nutzer) nach Daten und Diensten. Die Beschreibung von Prozessen und Diensten im Sinne der interoperablen Zusammenarbeit erfolgt durch gesonderte Kapitel.

Wie bereits beschrieben stellen Metadaten eine wesentliche Komponente der Geodateninfrastruktur dar. In ihnen werden vor allem organisatorische und strukturelle Aspekte der beschriebenen Datensätze festgehalten. Dazu können zum Beispiel Informationen wie Kontaktpersonen, rechtliche Beschränkungen oder das verwendete geographische Bezugssystem gehören. Metadaten stellen daher bereits eine Art der

⁸⁴ Metadaten sind oft Teil der eigentlichen Datenstrukturen und werden daher nicht extern erfasst und verwaltet (wie in der Geoinformatik üblich)

externen Wissensrepräsentation dar, da sie sich auf einen Datensatz beziehen und diesbezüglich Informationen abbilden, welche in dem besagten Datensatz selbst nicht vorhanden sind. Leider müssen wir feststellen, dass diese formalisierte Abbildung in Form von festgelegten (Meta-) Datenformaten⁸⁵ den Ansprüchen vieler Anwendungsfälle nicht gerecht werden kann. In Abhängigkeit vom jeweiligen Datenformat liegt der Fokus dieser Form von Metadaten eher auf organisatorischen, geographischen und strukturellen Aspekten der Datensätze. Die thematische Dimension wird dahingehend nur oberflächlich beschrieben, was häufig nicht ausreicht, um die Bedeutung und Inhalte der Daten anhand ihrer Metadaten zu interpretieren.

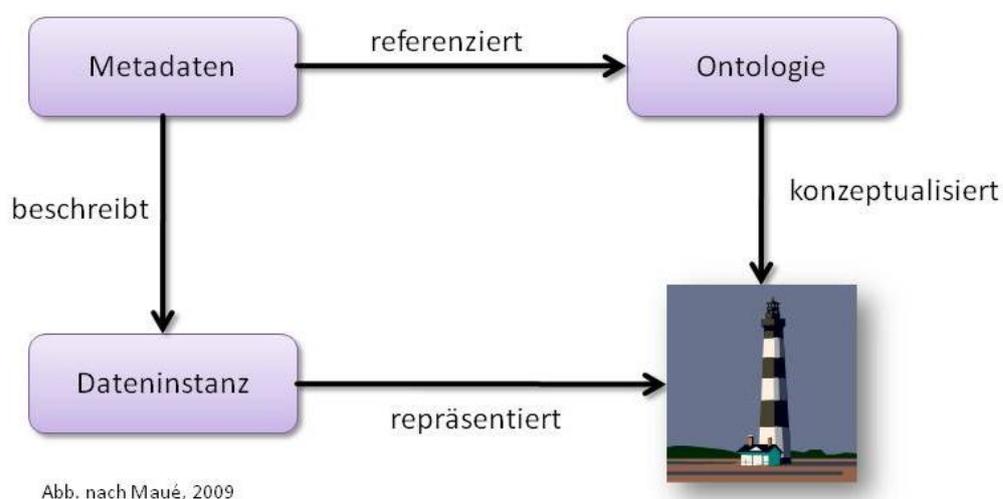


Abbildung 19: Semantische Annotation von Metadaten – Schema

Bezogen auf die in Europa verbreitete Normierung für Metadaten ISO19115⁸⁶ sind die Hauptinformationsquellen für eine thematische Beschreibung der Dateninstanz die Schlüsselwörter (engl. *Keywords*) und der Abstract. Während sich der Abstract als Freitext nicht für Annotation eignet, ist semantische Annotation von Schlüsselwörtern durchaus gebräuchlich. In der oben genannten Norm – der ISO 19115 – ist sogar eine

⁸⁵ Metadatenformate wie z.B. ISO19115/19139, Dublin Core oder FGDC

⁸⁶ Die Norm ISO19115 (http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020, 2009) ist u. A. die Grundlage für die Metadatenbeschreibung in INSPIRE und hat damit für Europa eine herausragende Bedeutung

Form der semantischen Annotation implementiert. Durch die INSPIRE-Richtlinie der Europäischen Kommission wird diese Form der semantischen Annotation bereits als verpflichtend eingeführt. Jeder gültige Metadatensatz muss neben den notwendigen Elementen von ISO19115 und der *INSPIRE Metadata Regulation*⁸⁷ mindestens ein Schlüsselwort enthalten, welches auf den GEMET-Thesaurus verweisen muss. Dieses bereits gezeigte Beispiel wird daher erneut aufgeführt. Dadurch sehen wir uns bereits mit der ersten Form der semantischen Annotation in einer kontinentalen Geodateninfrastruktur konfrontiert. Dies ist erforderlich, um die Verwendung von multilingualen Schlüsselwörtern zu regulieren und eine grundsätzliche Interoperabilität der Metadaten zu gewährleisten. Leider ist der referenzierte GEMET-Thesaurus für einige Fachgebiete und Fragestellungen nicht geeignet oder ausreichend, da er ursprünglich ausschließlich für Umweltthemen konzipiert wurde. Weitere kontrollierte Vokabulare können jedoch optional verwendet werden. Obwohl durch die ISO19115 und INSPIRE der Ausdruck „kontrolliertes Vokabular“ verwendet wird, ist es dennoch möglich auch ausdrucksstärkere semantische Technologien, wie zum Beispiel Ontologien, zu annotieren.

```

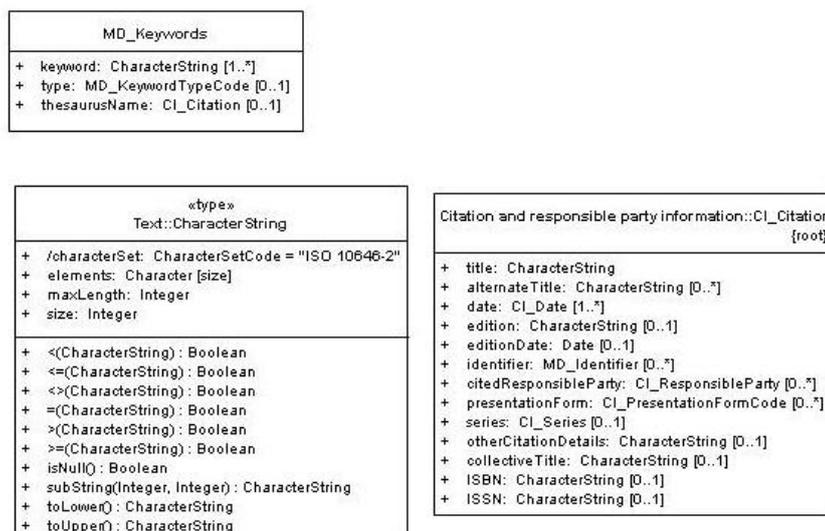
...
<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gco:CharacterString>Environmental Monitoring Facility</gco:CharacterString>
    </gmd:keyword>
    <gmd:keyword>
      <gco:CharacterString>waste, waste disposal, hazardous waste </gco:CharacterString>
    </gmd:keyword>
    <gmd:thesaurusName>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>GEMET Thesaurus Version 1.0</gco:CharacterString>
        </gmd:title>
        <gmd:date>
          <gmd:CI_Date>
            <gmd:date>
              <gco:Date>2004</gco:Date>
            </gmd:date>
            <gmd:dateType>
              <gmd:CI_DateTypeCode codeListValue =
"http://www.isotc211.org/2005/resources/Codelist/gmxCodeLists.xml#CI_DateTypeCode" codeList =
"publication">publication</gmd:CI_DateTypeCode>
            </gmd:dateType>
          </gmd:CI_Date>
        </gmd:date>
      </gmd:CI_Citation>
    </gmd:thesaurusName>
  </gmd:MD_Keywords>
</gmd:descriptiveKeywords>
...

```

Abbildung 20: Semantische Annotation von Metadaten – INSPIRE Metadatenbeschreibung (GEMET)

⁸⁷ Metadata Regulation vom 03.12.2008 (<http://inspire.jrc.ec.europa.eu/index.cfm/pageid/101>, 2009)

In diesem Beispiel wird demonstriert, wie eine Umsetzung der semantischen Annotation nach dem Schema von ISO19115 aussehen kann. Dabei ist anzumerken, dass diese Annotation aus zwei Komponenten besteht. Zum einen der gezeigten Dateninstanz (Metadateninstanz) und zum anderen dem XML-Schema⁸⁸. Durch das Schema werden die Struktur und Datentypen der Annotation geregelt. Im Fall der besagten ISO-Norm ist die semantische Annotation vergleichsweise komplex beschrieben. Neben dem eigentlichen Link ist u. A. die Angabe des Namens, der Version und der Repräsentationsform der externen Referenz möglich. Dies ist in der unten angeführten UML-Darstellung des Elementes *MD_Keywords*⁸⁹ der ISO19115 wiedergegeben.



Quelle: ISO - <http://www.isotc211.org/hmmg/HTML/EARoot/EA12/EA8/EA1985.htm>

Abbildung 21: MD_Keyword-Sektions der ISO19115

Im konkreten Zusammenhang mit der ISO19115 bietet sich daher die Annotation von Schlüsselwörtern besonders an, da die Schlüsselwörter idealerweise die Namen der jeweiligen Konzepte der externen Ontologie darstellen. So kann die generelle Lesbarkeit der Metadaten und annotierten Elemente gewährleistet werden. Durch diese

⁸⁸ XML Schema Definition (XSD), XSD Dokumente definieren die komplexe Struktur und die erlaubten Inhalte eines XML Dokumentes

⁸⁹ Quelle: <http://www.isotc211.org/hmmg/HTML/EARoot/EA12/EA8/EA1985.htm>, 2009

Variante der Annotation ist keine Änderung des bestehenden Schemas nötig und Schlüsselwörter können in der Regel bestehen bleiben. Die Verwendung von bestehenden Standards ohne strukturelle Modifikation sollte, soweit dies möglich ist, beibehalten werden (Maué, 2009).

Neben den hier erwähnten Schlüsselwörtern beinhalten Metadaten natürlich sehr viel mehr Informationen. Diese Elemente sind jedoch zum einen bereits semantisch klar geregelt und beschrieben. Zum Anderen würde eine semantische Annotation dieser Elemente eine Änderung des Standards bedeuten und ist damit nicht zielführend.

4.3.2 Semantische Annotation von OGC Diensten

Metadaten können sich neben den Daten auch auf Dienste (OGC Web Services, OWS) beziehen. Hierbei muss jedoch zwischen den Metadaten der Dienste durch Capabilities - Dokumente (teilweise eine indirekte Beschreibung der Daten) und der Dienstbeschreibung mittels *Web Service Description Language* unterschieden werden. Auf den Einsatz und die Diskussion von SAWSDL im Zusammenhang mit Dienstbeschreibung wird in diesem Kapitel verzichtet. SAWSDL bezieht sich auf *Web Service Description Language*, welche für die technische Beschreibung von Web Diensten im Allgemeinen (u. A. für SOAP-Bindings) verwendet wird. Diese XML - Spezifikation ist in Verbindung mit der Beschreibung von Diensten jedoch nicht GIS - spezifisch, weswegen sie nur im Zusammenhang mit Datenmodellen eingeführt wird.

Im Kontext unseres Verständnisses von Metadaten spielt neben der Beschreibung der Daten selbst auch die Beschreibung der Dienste (welche u. A. die Daten zur Verfügung stellen oder spezifische Prozesse und Transformationen durchführen) eine wichtige Rolle. Zu den angesprochenen Diensten zählen im Kontext der OGC unter Anderem der *Web Mapping Service*, der *Web Feature Service* oder der *Web Processing Service*. Solche Dienste erfüllen verschiedene Aufgaben und die Beschreibung der Dienste kann von der Beschreibung der publizierten Dateninhalte bis zu technischen Aspekten der durchgeführten Prozesse reichen.

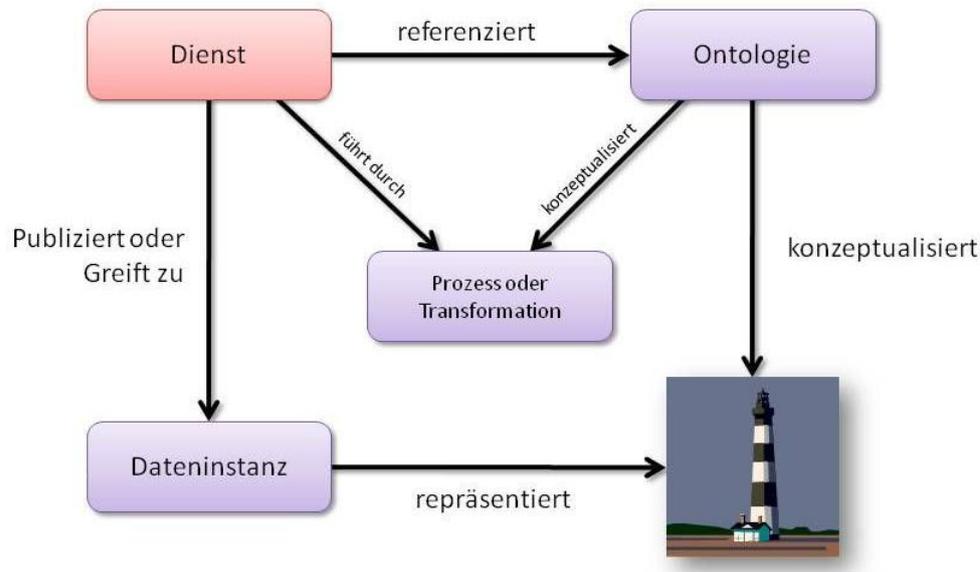


Abbildung 22: Semantische Annotation von OGC Diensten – Schema

Die Beschreibung von Diensten findet in unterschiedlichen Formen statt (wie z.B. durch ISO Norm ISO19119⁹⁰ oder Capabilities-Beschreibungen), welche sich in bestimmten Teilen ergänzen oder überlappen. Teile der inhaltlichen und technischen Beschreibungen von OWS finden sich auch im *Capabilities*⁹¹ - Dokument wieder, welches die Eigenschaften, Operationen und Inhalte von OGC - Diensten wiedergibt. Sowohl der Aufruf (GET oder POST auf HTML-Basis) als auch die Antwort von OWS auf eine entsprechende Anfrage nach den Möglichkeiten des Dienstes, sind durch die OGC standardisiert worden. (Whiteside, 2007) Diese Dienstbeschreibungen durch die *Capabilities* umfassen sowohl strukturelle als auch inhaltliche Aspekte und enthalten unter Anderem die *Service Identification Section*⁹². An dieser Stelle ergibt sich eine weitere Möglichkeit des Einsatzes von semantischen Annotationen. Die *Service Identification Section* als Teil der OWS - *Capabilities*, welche folgend als UML -

⁹⁰ ISO19119, http://www.iso.org/iso/catalogue_detail.htm?csnumber=39890, 2010

⁹¹ Capabilities Dokumente sind die Antwort des OGC Dienstes auf eine GetCapabilities Anfrage, über welche viele OGC Dienste als Basisfunktionalität verfügen

⁹² Vgl. dazu Whiteside, 2007

Modell abgebildet ist, enthält unter Anderem eine Möglichkeit, den OWS mit Schlüsselwörtern zu kategorisieren.

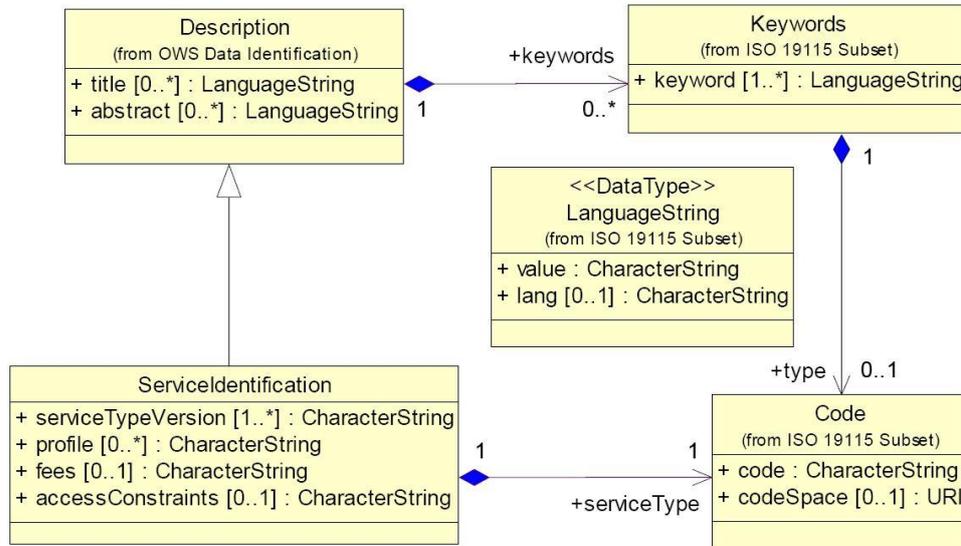


Abbildung 23: Identification Section als UML (Whiteside, 2009)

Die Möglichkeit der Vergabe von Schlüsselwörtern referenziert auf MD - Keywords der bereits erwähnten ISO19115. Wie bereits dargestellt, lassen sich auf diese Weise Schlüsselwörter und entsprechende, kontrollierte Vokabulare standardisiert angeben. Diese Art der semantischen Annotation bezieht sich, wie auch die Annotation der bereits beschriebenen Metadaten (in Relation zu Dateninstanzen), auf die thematische Einordnung und Kategorisierung der bereitgestellten Dienste und Inhalte des OWS. Neben dieser thematischen Beschreibung kann das referenzierte Konzept auch explizites Wissen über die angebotene Transformation oder die Prozesse enthalten. So kann zum Beispiel durch das Schlüsselwort „Intersection“ auf das entsprechende Konzept verwiesen werden und dort die technischen Spezifikationen des Prozesses abbilden. Das folgende Beispiel zeigt die Verwendung von semantischer Annotation innerhalb der *Service Identification Section* eines *Sensor Observation Service*⁹³.

⁹³ Sensor Observation Service (SOS, <http://www.openeospatial.org/standards/sos>, 2009)

```

...
<ows:ServiceIdentification
  xmlns:ogc = "http://www.opengis.net/ogc"
  xmlns:ows = "http://www.opengis.net/ows/1.1"
  xmlns:om = "http://www.opengis.net/om/1.0"
  xmlns:swe = "http://www.opengis.net/swe/1.0">
  <ows:Title>SOSsrv</ows:Title>
  <ows:Abstract>Cascading Sensor Observation Service</ows:Abstract>
  <ows:Keywords>
    <ows:Keyword>CleanAirForEurope</ows:Keyword>
    <ows:Keyword>PM10</ows:Keyword>
    <ows:Keyword>PM25</ows:Keyword>
    <ows:Keyword>O3</ows:Keyword>
    <ows:Type codespace = "http://www.katzengeist.de/attachments/File/CAFE_Phenomena.owl#">ontology</ows:Type>
  </ows:Keywords>
  <ows:ServiceType codeSpace = "http://opengeospatial.net">OGC:SOS</ows:ServiceType>
  <ows:ServiceTypeVersion>1.0.0</ows:ServiceTypeVersion>
  <ows:Fees>NONE</ows:Fees>
  <ows:AccessConstraints>NONE</ows:AccessConstraints>
</ows:ServiceIdentification>
...

```

Abbildung 24: Semantische Annotation von OGC Diensten – Service Identification Section (XML)

Ebenso wie Metadaten, welche sich auf Dateninstanzen beziehen, können auch auf OGC Dienste bezogene Metadaten in Katalogen gespeichert und verwaltet werden.

4.3.3 Semantische Annotation von Datenschemen

Neben Metadaten und Dienstbeschreibungen stellen die Datenschemen die dritte Ebene dar, auf welchen semantische Annotationen eingesetzt werden können. Da diese – thematisch bedingt – im Kontext von OGC - Standards betrachtet werden und es sich bei den verwendeten Datenformaten ausschließlich um XML-Spezifikationen handelt, sprechen wir daher von XML-Schemen oder Datenschemen.

Im Rahmen dieser Arbeit wird die Kenntnis von der Extensible Markup Language (XML) als auch der entsprechenden XML - Schemata (XSD) vorausgesetzt. XML-Schemen haben vorrangig die Aufgabe, die Struktur eines XML-Dokumentes festzulegen, wodurch diese gegen das Schema validiert werden können. Ein XML Schema nutzt dafür selbst XML. Dabei können verschiedene Funktionen des Schemas genutzt werden, um der XML - Instanz selbst noch ein gewisses Maß an Flexibilität zu gewährleisten. So können Gruppierungen und Kardinalität eingesetzt werden. (Hitzler,

2009) Ein Beispiel für ein generisches und flexibles XML - Schemata ist beispielsweise *Observation and Measurement* oder *SensorML*.

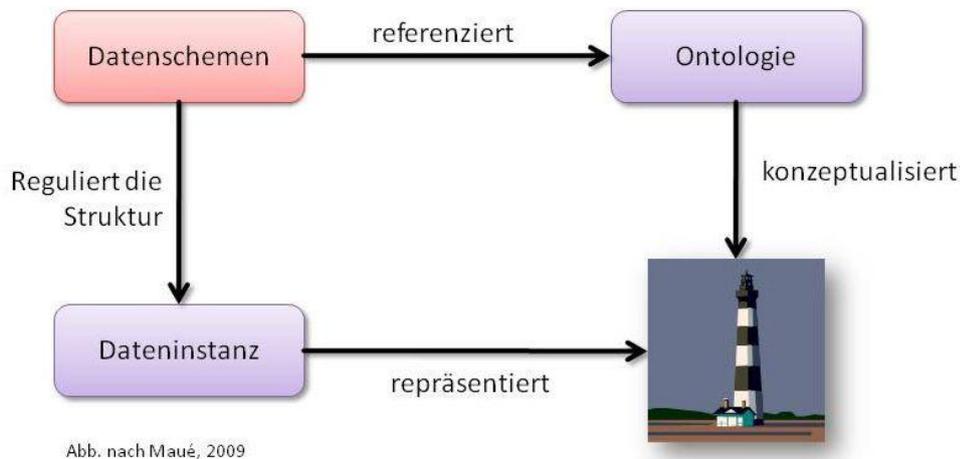


Abbildung 25: Semantische Annotation von Datenschemen - Schema

Eine festgelegte Struktur von XML - Dateninstanzen ermöglicht die Kommunikation dieser Daten und verbessert die syntaktische Interoperabilität. Einer der Nachteile der Datenmodellierung liegt darin, dass es verschiedene Konzepte gibt, auf welcher Ebene die Semantik ausgedrückt wird. Es sind besonders im Hinblick auf XML - Spezifikationen der OGC zwei Tendenzen zu bemerken. Diese beiden Formen der Konzeption wurden bereits im Kapitel Validierung von Prozessketten beschrieben und sollen hier nur kurz angerissen werden.

Zum einen ist es möglich, (z. B. mittels GML) sehr spezifische Datenmodelle zu entwickeln, bei welchen Semantik hauptsächlich auf Seiten der XML - Schemen zum Einsatz kommt. Dies bedeutet, dass die Semantik eines Ausdrucks auf der Ebene der Schemen und nicht auf der Ebene der Dateninstanz erläutert wird. Ein Beispiel dafür ist folgend abgebildet.

```

...
<xs:complexType name = "O3_Messung">
  <xs:complexContent>
    <xs:extension base = "gml:AbstractFeatureType">
      <xs:sequence>
        <xs:element name = "Stationsname" type = "xs:string"/>
        <xs:element name = "O3_Konzentration" type = "xs:double"/>
        <xs:element name = "O3_Tageshöchstwert" type = "xs:double"/>
      </xs:sequence>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>

```

```
...
```

Abbildung 26: Semantische Annotation von Datenschemen – Semantik auf Schemen-Ebene

Die Abbildung 25 zeigt ein XML-Schema, welches die Inhalte und die Semantik streng regelt und eine semantische Annotation nur auf Ebene der Schemata sinnvoll erscheinen lässt.

Die zweite Tendenz der OGC-Datenmodellierung ist, dass XML-Schemata sehr generisch gehalten werden und die genaue Bedeutung einzelner Elemente und Gruppierungen ausschließlich auf der Ebene der XML-Dateninstanz erklärt wird. Daraus ergibt sich, dass eine semantische Annotation vermehrt auf Seiten der Dateninstanz eingesetzt wird. Beispiele für derart generische XML-Schemata sind die schon genannten *Observation and Measurement* und auch *SensorML*, welche beide zu SWE gehören. Beide Standards verfügen über eine Struktur, welche allerdings in sich generisch und bis zu einem gewissen Grad flexibel ist. Dadurch wird beispielsweise die Eigenschaft `<observedProperty>` im Schema festgelegt. Die Bedeutung und Aussage, welche Eigenschaft tatsächlich gemessen wird, findet sich allerdings in der Dateninstanz wieder. Das folgende kurze Beispiel (*Observation and Measurement*) gibt auszugsweise wieder, wie die Bedeutung und Metadaten von Ausdrücken auf Ebene der Dateninstanz ausgedrückt werden.

```
...
<om:observedProperty xlink:href = "urn:ogc:def:phenomenon:OGC:Ozone"/>
...
```

Abbildung 27: Semantische Annotation von Datenschemen – Semantik auf Schemata-Ebene

Die Unterscheidung zwischen diesen beiden Formen der XML-Datenmodellierung ist wichtig, da durch sie entschieden wird, auf welcher Ebene – XML-Schemata oder XML-Dateninstanz – eine semantische Annotation sinnvoll eingesetzt werden kann. Da in dieser Arbeit (unter Anderem) versucht wird, die konkreten Ansatzpunkte für semantische Annotation im Kontext von *Sensor Web Enablement* zu identifizieren, lässt sich die Schlussfolgerung ableiten, dass semantische Annotation auf Ebene der XML-Schemata im Kontext von SWE nicht relevant ist. Dennoch soll hier kurz auf die Möglichkeit der semantischen Annotation von Datenschemen eingegangen werden, da dies für andere OGC Standards (nicht SWE) von Relevanz ist.

Für die semantische Annotation von XML-Schemen bietet sich die W3C-Spezifikation SAWSDL an. Diese Spezifikation bietet Möglichkeiten, semantische Annotation in *Web*

Service Description Language als auch in XML-Schemen zu implementieren. Bezogen auf XML-Schemata bietet diese Spezifikation die Möglichkeit, Elementen das Attribut „*Model Reference*“ hinzuzufügen. Durch dieses Attribut kann auf ein Konzept in einer externen Ontologie verwiesen werden. In dieser kann der Client (sofern er in der Lage ist, den Link aufzulösen) erweiterte Informationen über besagtes Element finden und entscheiden, ob die angebotenen Daten seinen Ansprüchen gerecht werden. Der folgende Auszug eines XML-Schemata stellt ein Beispiel für die Verwendung von „*Model Reference*“ dar.

```

...
<xs:complexType>
  <xs:sequence minOccurs="1" maxOccurs="unbounded">
    <xs:element name="Ozon" type="xs:double"
sawsdl:modelReference="http://www.katzengeist.de/attachments/File/CAFE_Phenomena.owl#O3"/>
    <xs:element name="Stickstoffdioxid" type="xs:double"
sawsdl:modelReference="http://www.katzengeist.de/attachments/File/CAFE_Phenomena.owl#NO2"/>
  </xs:sequence>
</xs:complexType>
...

```

SAWSDL erlaubt neben der Annotation von *Complex Types* (siehe Beispiel) auch die Annotation von *Single Types*, Elementen und Attributen. Neben diesen ermöglicht sie noch die Funktion des „Schema Mappings“, welches im konkreten Kontext nicht relevant ist. Eine ausführliche Auseinandersetzung mit den technischen Details von SAWSDL ist im Rahmen dieser Arbeit jedoch nicht möglich. Durch SAWSDL wird aufgezeigt, dass die semantische Annotation auf Ebene der Datenmodelle bereits als W3C - Standard verankert ist.

4.3.4 Semantische Annotation von Dateninstanzen

Die letzte Ebene, auf welcher die semantische Annotation sinnvoll erscheint, ist die Ebene der Dateninstanz. Daten sind dabei die Träger der Information und die essentielle Ebene der Geodateninfrastruktur. Zugleich sind sie in den meisten Fällen nicht selbsterklärend, sondern bedürfen umfassender Beschreibung, um die Kommunikation und den Einsatz von Daten zu ermöglichen.

Bezogen auf die Spezifikationen des *Sensor Web Enablement* ist die Ebene der Dateninstanz der hauptsächliche Ort der semantischen Annotation. Auf dieser Ebene

finden sich (im Gegensatz zu sehr applikationsspezifischen Schemen basierend auf GML und XML) die eigentlichen semantischen Aussagen. Die Bedeutung von Semantik auf der Datenebene wurde bereits in den Abschnitten Semantische Annotation von Datenschemen und Validierung von Prozessketten verdeutlicht und ist bei den Datenstandards von SWE (O&M und SensorML) besonders ausgeprägt. Auf dieser – neben den Metadaten – wichtigsten Ebene der semantischen Annotation lassen sich durch Referenzen auf externe Quellen verschiedene Aussagen abbilden.

- Semantische Aussagen über die verwendeten Begrifflichkeiten
- Aussagen und Modelle über die eventuell zugrunde liegenden Transformationen und Prozesse, auf denen Begrifflichkeiten beruhen

Beide Varianten der semantischen Erweiterung lassen sich durch Verlinken von Elementen auf externe semantische Modelle realisieren. Diese Modelle können dabei je nach Anwendungsfall applikationsspezifische oder Domänenontologien darstellen.

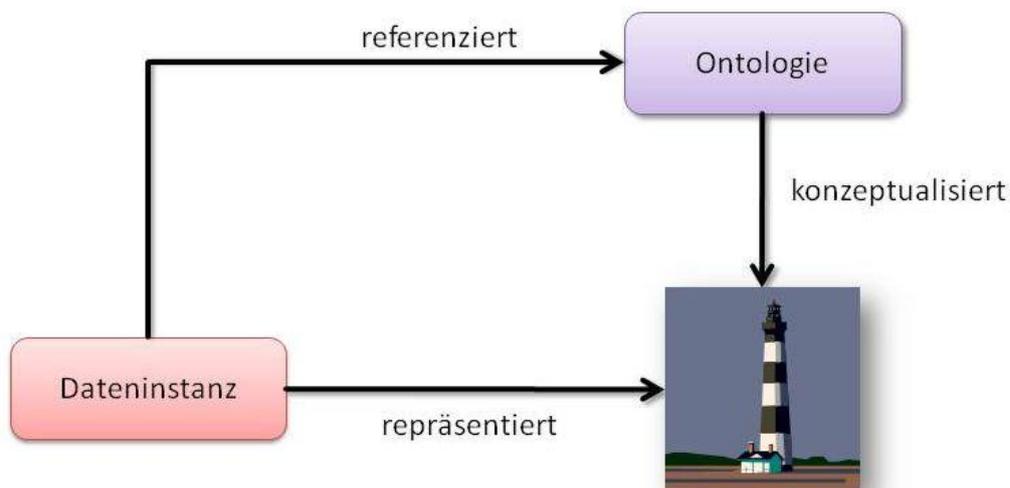


Abbildung 28: Semantische Annotation von Dateninstanzen – Schema

Eine der technisch einfachsten Methoden, die semantische Aussagekraft von Elementen zu verbessern, ist die Verwendung von URN⁹⁴. Diese URN, welche zu den URI⁹⁵ (mit

⁹⁴ Uniform Resource Name (URN)

⁹⁵ Uniform Resource Identifier (URI)

dem Schema „urn:“) zählt, ist ein ortsunabhängiger Bezeichner, welcher (im Gegensatz zu URL) nicht eindeutig sein muss. Eine URN bezeichnet eine Ressource, die über einen langen Zeitraum besteht und unabhängig von ihrer virtuellen Existenz im Web ist. Obwohl diese Bezeichnungen auf Ressourcen im Web verweisen können, ist ihr Einsatz als Metainformation von Elementen ebenso möglich. Durch sie können Elemente unter Anderem thematisch klassifiziert werden. Dies ist kein Verweis auf eine externe semantische Quelle im eigentlichen Sinne, da die URN zwar den Pfad und das Konzept der Klassifikation (z.B. eines Thesaurus) angibt, aber nicht die Quelle derselben. Ein Beispiel für den Einsatz von URN bietet folgender Auszug:

```
...  
<om:observedProperty xlink:href = "urn:ogc:def:phenomenon:OGC:mass"/>  
...
```

Abbildung 29: Semantische Annotation von Dateninstanzen – URN

Durch dieses Beispiel werden sofort die Vor- und Nachteile der Verwendung von URNs augenscheinlich. URNs sind für Klassifikationen einsetzbar, welche für den menschlichen Betrachter verständlich sind. Zudem bieten sie eine einfache Möglichkeit, die semantische Aussagekraft einer Bezeichnung zu verstärken. Allerdings ist ihre Klassifikation keineswegs geregelt und transparent. Da es theoretisch und praktisch möglich ist, individuelle Klassifikationen als URN einzuführen und es kein übergeordnetes Reglement oder System gibt, kann die Verwendung von solchen Bezeichnern undurchsichtig werden. Sofern diese URNs nicht auf genormten Klassifikationen oder Thesauri beruhen, kann diese Form der Klassifikation nicht sinnvoll genutzt werden. Zudem gibt es keine Möglichkeit, diese Form der Bezeichner maschinell aufzulösen. Hierfür können in geschlossenen Systemen und Applikationen zwar sogenannte URN-Resolver verwendet werden, jedoch fehlen solche Möglichkeiten auf globaler Ebene. Es sind in den letzten Jahren Bestrebungen der OGC zu URN-Resolovern zu beobachten, welche aber bislang zu keinem durchgreifenden Erfolg geführt haben.⁹⁶

⁹⁶ OGC Naming Authority (<http://www.opengeospatial.org/ogcna>, 2009)

Zu den Möglichkeiten der eigentlichen semantischen Annotation von Dateninstanzen ist einleitend zu erwähnen, dass es bislang keinen Standard gibt, der diese Form der Referenzierung regelt. Innerhalb der OGC-Spezifikation von *SensorML* und *Observation and Measurement* wurde die semantische Annotation erstmals ergänzend zur XML-Dateninstanz vorgeschlagen. Die dort verwendeten Annotationen beziehen sich jedoch auf keinen Standard und sind von SAWSDL abgeleitet.

Grundsätzlich besteht bei der semantischen Annotation von XML-Instanzen eine direkte Abhängigkeit von den entsprechenden XML-Schemata, auf denen die notwendigen Strukturen die Annotation geschaffen werden müssen. Innerhalb von SAWSDL erfolgt die semantische Annotation durch das Attribut *modelreference* und einem entsprechenden URI (unter Anderem erkennbar am Beispiel von http://www.katzengeist.de/attachments/File/CAFE_Phenomena.owl#O3). Innerhalb individueller Applikationsschemen können semantische Annotationen demnach wie folgt realisiert werden.

```
...
<complexType>
  <sequence minOccurs = "1" maxOccurs = "unbounded">
    <element name = "Pollutant" type = "anyURI"/>
    <element name = "Value" type = "double"/>
  </sequence>
</complexType>
...
```

Abbildung 30: Semantische Annotation von Dateninstanzen – XSD

Im Schema können wir erkennen, dass das Element *<Pollutant>* den Datentyp „*anyURI*“ aufweist. Dies ist die Voraussetzung für die semantische Annotation auf der Ebene der Dateninstanz. Die entsprechende XML-Instanz könnte wie folgt aussehen.

```
...
<STATION tns:id = "AAAAC">
  <Pollutant>"http://www.katzengeist.de/attachments/File/CAFE_Phenomena.owl#O3"</Pollutant>
  <Value>12,3</Value>
</STATION>
...
```

Abbildung 31: Semantische Annotation von Dateninstanzen - XML

Im Kontext der Datenstandards SWE, welche bereits als XML-Schemata *Observation and Measurement* und *SensorML* vorliegen, ist diese Form der semantischen Annotation allerdings weniger sinnvoll, da jegliche individuelle Änderung das jeweilige Standard (Applikations-) Schema unterhöhlen würde. Dies würde den Sinn der Bestrebungen um verstärkte Interoperabilität unterwandern, weshalb innerhalb dieser bereits festgelegten

Standards eine semantische Annotation basierend auf einem eigens dafür implementierten Attribut oder Element nicht möglich ist.

Für die semantische Annotation in *Observation and Measurement* wie auch in *SensorML* kann die generische und offene Struktur der beiden OGC - Standards genutzt werden. In den Beschreibungen der XML - Schemata wird an zahlreichen Stellen bereits die Verwendung von semantischer Annotation vorgeschlagen durch die Autoren vorgeschlagen.⁹⁷ Zahlreiche Elemente (wie beispielsweise *ObservedProperty* in *O&M* oder *UOM* in *SensorML*) verweisen auf Schemenebene auf Typen (wie den *AbstractDataComponentType* oder den *gml:DefinitionType*), welche es erlauben, semantische Annotationen innerhalb der entsprechenden Elemente der XML - Dateninstanz unterzubringen. Eines der großen Probleme der semantischen Annotationen, welche letztendlich auf Datentypen wie „*anyURL*“ oder „*String*“ basiert, ist, dass sie keinerlei Syntax unterliegen. Bis jetzt existieren durch die OGC und das W3C lediglich verschiedene Vorschläge, wie solche Annotationen umzusetzen sind. Für eine weitere erfolgreiche Implementierung von semantischen Annotationen im Sinne des Semantic Web und der semantischen Interoperabilität müssen diese unterschiedlichen Notationen einer Normung unterzogen werden, wie dies mit SAWSDL bereits erfolgreich durchgeführt wurde.

Wie bereits erwähnt finden sich in verschiedenen Projekten des Sensorwebs unterschiedliche Formen der semantischen Annotation in Dateninstanzen. Im Folgenden werden verschiedene Varianten der semantischen Annotation vorgestellt, ohne dass dabei eine persönliche Wertung vorgenommen wird.

In den meisten Fällen wird diese Form der Annotation mittels XLink realisiert. Bei XLink handelt sich um eine Form der Syntax, um Links und Verweise innerhalb von XML-Dokumenten auszudrücken. XLink wird dabei in Kombination mit dem Attribut „*href*“ benutzt, welches einen URI enthält. Dieser URI stellt wie bei der Annotation von Schemata den Pfad der Ontologie und die entsprechende Klasse (bzw. die Instanz) dar. Diese Version der Annotation wird unter Anderem von OGC in den Spezifikationen von *Observation and Measurement* und *SensorML* vorgeschlagen.

⁹⁷ Vgl. dazu Cox, 2007 und Mauè, 2009

```

...
<swe:Phenomenon gml:id = "AtmosphericPressure">
  <gml:description xlink:href = "http://sweet.jpl.nasa.gov/ontology/property.owl#AtmosphericPressure">
    fluid pressure exerted due to the gravitational effect on the column of atmosphere above the position of interest
  </gml:description>
  <gml:name codeSpace = "urn:ogc:tc:arch:doc-rp(05-010)"> urn:ogc:def:property:OGC:AtmosphericPressure</gml:name>
  <gml:name>Atmospheric Pressure</gml:name>
</swe:Phenomenon>
...

```

Abbildung 32: Semantische Annotation von Dateninstanzen – XLink in SensorML

Im obigen Beispiel, welches der OGC - Spezifikation von SensorML⁹⁸ entnommen ist, wurde die angesprochene *xlink:href* - Variante der Annotation gewählt. Durch das SensorML - Schema begünstigt, ist die semantische Annotation tatsächlich eine Erweiterung der bestehenden Inhalte der Dateninstanz. In diesem Beispiel kann zudem die parallele Nutzung von der soeben besprochenen Annotation mittels XLink und der Nutzung von URNs beobachtet werden.

Da allerdings bisher kein Standard existiert, welcher die Verwendung von Annotationen regelt, finden sich in anderen OGC - Spezifikation wiederum verschiedene Formen der Notation. Folgendes Beispiel, welches der Spezifikation von Observation and Measurement⁹⁹ entstammt, verzichtet auf die Verwendung von XLink.

```

...
<om:parameter>
  <swe:Quantity definition = "http://sweet.jpl.nasa.gov/ontology/property.owl#Temperature">
    <swe:uom xlink:href = "urn:ogc:def:uom:UCUM:Cel"/>
    <swe:value>22.3</swe:value>
  </swe:Quantity><!--example of optional soft-typed parameter -->
</om:parameter>
...

```

Abbildung 33: Semantische Annotation von Dateninstanzen – O&M

Da alle diese Beziehungen zwischen Ausdrücken innerhalb von XML - Spezifikationen und der Wissensrepräsentationen, in welchen sie konzeptualisiert sind, von Clients abhängig sind, kann hierbei keinerlei Bewertung vorgenommen werden. Beide

⁹⁸ Botts, 2007

⁹⁹ Cox, 2007

Varianten (mit und ohne xlink:href) können von einem Client aufgelöst werden, welcher die Aufgabe hat, die entsprechenden Informationen innerhalb der Ontologie auszuwerten.

4.4 Bewertung der Annotationsmöglichkeiten und Thesen

Als erstes muss festgehalten werden, dass semantische Annotationen auf allen für eine (auf SWE basierende) Geodateninfrastruktur identifizierten Ebenen umsetzbar ist. Diese Tatsache bezieht sich zudem auf die erste der in Thesen formulierten Aussagen. Zudem wird in den Kapiteln Ontologie und

Semantische Annotation dargelegt, wie externes Wissen mittels Ontologien abgebildet und in Geodateninfrastrukturen eingebunden werden kann. Damit können drei der vier genannten Thesen bestätigt werden. Einzig die Aussage, dass weitere Bestrebungen zu Einbindung und Normierung von semantischen Verweisen „notwendig“ sind, lässt sich nur indirekt bestätigen. Dies kann aus dem immer stärkeren Interesse an Geosemantik und dem damit verbundenen Problembewusstsein abgeleitet werden. Eine zwingende Notwendigkeit kann zum heutigen Stand jedoch nicht abgeleitet werden, da die großen angesprochenen Geodateninfrastrukturvorhaben INSPIRE, SEIS und GMES sich noch im Stadium der Entwicklung und Konzeption befinden. Es bleibt zu erwarten, dass die angesprochenen Probleme der semantischen Interoperabilität erst mit ihrem Auftreten deutlich ins Bewusstsein rücken.

Im Folgenden wird versucht die dargestellten Methoden der semantischen Annotation zu bewerten. Besonderes Augenmerk soll bei der Bewertung der verschiedenen Ansätze auf die Ebene der Metadaten und Dateninstanz gelegt werden.

Semantische Annotationen unterliegen derzeit einem wesentlichen Nachteil, welcher unabhängig von der Implementierungsebene der Annotation ist. Es gibt bisher nur wenige Applikationen, welche in der Lage sind, semantische Verweise aufzulösen und diese zu Nutzen. Dies ist ein besonderer Nachteil, da ohne eine stärkere Nutzung dieser semantischen Komponente nur wenig Druck für weitere Entwicklungen, Normungen und Nutzungskonzepte entsteht.

Semantische Annotation von Metadaten (auf XML basierend) ist die am einfachsten zu implementierende Form der Referenzierung auf externe Quellen. Dies wird vor allem dadurch gefördert, dass innerhalb des betrachteten Metadatenstandard ISO19115 bereits eine solche Annotation vorgesehen ist. Auch wenn die Intention, auf kontrollierte Vokabulare innerhalb der Sektion der Schlüsselwörter zu verweisen, vermutlich auf die Regulierung der verwendeten Ausdrücke beschränkt war, lässt sich dieser Bereich hervorragend für die semantische Annotation von Metadaten nutzen. Dies hat den Vorteil, dass für die Implementierung von semantischen Annotationen innerhalb der Schlüsselwortsektion keine Änderungen am zugrunde liegenden Schema notwendig

sind. Bis jetzt sind bereits in verschiedenen Geodateninfrastrukturprojekten semantische Verweise in Metadaten und die entsprechenden Applikationen, welche diese Verweise konsumieren und verwerten können, umgesetzt worden. Dies deutet darauf hin, dass diese Methode einfach zu implementieren ist und zudem einen Mehrwert im Prozess der Suche nach Daten und Diensten darstellt.

Natürlich beinhaltet die semantische Annotation von Schlüsselwörtern in Metadaten auf negative Aspekte (im Vergleich zu anderen Ebenen). Durch die Annotation von Schlüsselwörtern lassen sich die entsprechenden Daten jedoch nur sehr beschränkt beschreiben. Zudem ist eine Referenz von Schlüsselwörtern untereinander nicht möglich, wodurch die Zusammenhänge nur indirekt durch die Ontologie geklärt werden müssen. Ein weiterer Nachteil entsteht durch die Heterogenität von komplexen Datensätzen, wodurch nicht mehr gewährleistet werden kann, dass die gewählten Schlüsselwörter jeweils den gesamten Datensatz beschreiben. Eine mögliche Folge wären eingeschränkte und ungenaue Schlussfolgerungen (Maué, 2009).

Die Bewertung der semantischen Annotation auf Ebene der Metadaten lässt sich auch auf die Ebene der Prozessbeschreibung übertragen, da diese eng verwandt sind.

Besonders im Kontext von SWE ist die **semantische Annotation auf Ebene der Datenmodelle** weder notwendig noch sinnvoll. Darüber hinaus wurde die semantische Annotation mittels SAWSDL in herkömmlichen XML basierten Applikationsschemen diskutiert. Durch die Verwendung dieses Standards für Verweise auf externe Ressourcen lassen sich Konzepte und Individuen konkreter referenzieren als auf Ebene der Metadaten. Durch die semantische Annotation aus der inneren Struktur heraus können alle Elemente des Datenmodelles direkt annotiert werden. Zudem bietet die Verwendung eines Standards einen hohen Grad an Interoperabilität der Verweise, wodurch die Abhängigkeit von spezifischen Applikationen zum Auflösen der Verweise verbessert wird.

Zum Nachteil der semantischen Annotation auf Ebene der Datenmodelle gehört als erstes die Komplexität der Modelle und internen Strukturen genannt. Zudem kommt es durch diese konkreten und komplexen Strukturen zu einem Verlust an Flexibilität, wie sie auf der Ebene der Metadaten vorliegt (Mauè, 2009).

Die **semantische Annotation auf Ebene der Dateninstanz** ist im Kontext von SWE besonders wichtig. Hier profitieren wir davon, dass diese Form von Verweisen bereits im XML Schemata vorgesehen und vorgeschlagen wird. Dies ermöglicht einfache und flexible Möglichkeiten der Implementierung innerhalb des definierten Standards. Diese Variante der Annotation eignet sich besonders für XML Dateninstanzen, welche einem generischen und weniger flexiblen Schema unterliegen.

Besonders in der Flexibilität der Annotation auf Datenebene liegt auch der große Nachteil dieser Methode. Bis jetzt existiert kein Standard (wie SAWSDL für Schemata und WSDL), welcher die verschiedenen Arten der Notation dieser Verweise regelt. Ein weiterer Kritikpunkt ist die durch Redundanz der Datenpunkte verursachte hohe Menge an Annotationspunkten. Dies kann unter Anderem ein deutlich erhöhtes Datenvolumen verursachen. Da sowohl SWE als auch die semantische Annotation im Bereich Geodateninfrastrukturen, vergleichsweise neue Unterthemen der Geoinformatik sind, existieren bis jetzt nur wenige Applikationen (Server wie auch Clients), welche diese Methoden und Techniken entsprechend unterstützen (Mauè, 2009).

5 Zusammenfassung, Kritik und Ausblick

Semantische Annotationen sind ein Grundbestandteil und eine Schlüsseltechnologie des Semantic Webs. Die Probleme und Ideen, welche zur Formulierung des Konzeptes des Semantic Webs im Jahre 2001 führten, lassen sich direkt auf den Bereich der Geodateninfrastrukturen übertragen. Dazu gehört vor allem die mangelnde Interoperabilität.

Interoperabilität ist zwar im Bereich der Geoinformatik ein oft genutztes Schlagwort, allerdings beziehen sich fast alle Aktivitäten und Ansätze zur Verbesserung der Interoperabilität auf den Teilaspekt der syntaktischen Interoperabilität. Diese Facette ist mittlerweile tatsächlich zu einem hohen Grad sichergestellt. Ein Indikator hierfür ist das Entstehen und erfolgreiche Operieren von zahlreichen nationalen und internationalen Geodateninfrastrukturprojekten. Der zweite Aspekt der Interoperabilität, die semantische Interoperabilität, wurde bisher jedoch weitgehend vernachlässigt. Semantische Interoperabilität ist die zweite Ebene der Interoperabilität, welche sich mit der Bedeutung von Daten bzw. Informationen und ihrer Kommunikation befasst.

Die semantische Annotation ist eine Möglichkeit, dem drohenden Verlust an Inhalten und Informationen in immer stärker zusammenwachsenden Informationsnetzwerken entgegen zu wirken. Der Begriff der semantischen Annotation lässt sich auch mit Links und Verweisen auf externe semantische Modelle gleichsetzen. Die Grundidee dabei ist, dass die bisherigen Inhalte von Dateninstanzen erweitert werden, indem von bestimmten Punkten der Instanz aus auf Wissensrepräsentationen referenziert wird. Diese externen Wissensmodelle können zusätzliche Informationen zur Bedeutung von Begriffen aber auch zu relevanten Prozessen und Transformationen bereitstellen. Dadurch kann dem möglichen Verlust an Semantik in Informationsnetzwerken entgegengewirkt werden und die interoperable Zusammenarbeit von Diensten deutlich gefördert werden.

Im Rahmen dieser Arbeit wurden vier Ebenen innerhalb der Geodateninfrastruktur identifiziert, auf welchen die semantische Annotation sinnvoll erscheint. Dazu gehören die Ebene der Metadaten, der Dienstbeschreibung, der Datenschemata als auch die Ebene der eigentlichen Dateninstanzen. Zu den jeweiligen Ebenen wurden verschiedene

Implementierungsmöglichkeiten vorgeschlagen, welche teilweise bereits in Geodateninfrastrukturprojekten umgesetzt werden. Da der Fokus der Arbeit auf semantischen Annotationen innerhalb von *Sensor Web Enablement* liegt, konzentrieren sich die Diskussion der Implementierung auf die Bereiche der Metadaten und der Dateninstanzen.

Zu den generellen Kritikpunkten an bisherigen Umsetzungen der semantischen Annotation gehört vor allem der geringe Grad an Normung und Standards. Derartige Standards existieren nur für bestimmte Teilaspekte (wie SAWSDL für XSDL und XML-Schemata) der Geodateninfrastruktur. Hier sind vor allem durch das W3C als auch durch die OGC weitere Initiativen und Regelungen notwendig. Ein weiterer Kritikpunkt ist, dass es bisher nur sehr wenige Projekte innerhalb der Geoinformatik gibt, welche sich dem Thema der semantischen Interoperabilität widmen. Hier sind eine stärkere Entwicklung von Applikationen und Diensten notwendig, welche diese Technologie unterstützen und konsumieren.

Eine Regulierung der verschiedenen Formen der semantischen Annotation als auch eine verstärkte softwaretechnische Entwicklung ist die Grundvoraussetzung für den erfolgreichen Einsatz von semantischer Annotation (auch) in internationalen Geodateninfrastrukturvorhaben wie INSPIRE, GEOSS oder SEIS. Mit ihr kann dem Verlust an Semantik und der ständig wachsenden semantischen Lücke vorgebeugt werden. Zudem kann die Verwendung von externen Wissensrepräsentationen die bisherige Ebene der Daten und Informationen deutlich erweitern.

6 Literatur

G. Babitski, S. Bergweiler, J. Hoffmann, D. Schön, C. Stasch, A. Walkowski, Ontology-based Integration of SensorWeb Services in Disaster Management In: TerraCognita 2009 Workshop In conjunction with the 8th International Semantic WebConference (ISWC 2009), Mexico City, 2009

J. Behrendt, K. Zeppenfeld, Web 2.0: Informatik im Fokus, Springer, Berlin, 2008

L. Bernard, H. Pundt, Semantikverlust in integrierten Systemen – Ein Fallbeispiel aus der Umweltplanung, Proceedings: 12. Internationales Symposium "Informatik für den Umweltschutz" der Gesellschaft für Informatik (GI), Bremen, 1998

T. Berners-Lee, J. Hendler, O. Lassila: The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In: Scientific American, 284 (5), S. 34–43, May 2001

R. Bommersbach, Semantic Annotation, Project Discussion Paper, S@NY IP – Sensors Anywhere, SANY Consortium, Wien, 2009

M. Botts, OpenGIS® Implementation Standard OGC® 07-000: OpenGIS® Sensor Model Language (SensorML) Implementation Specification , OGC, 2007.

R. Brachman, H. Levesque, Knowledge representation and reasoning, Morgan Kaufmann, San Francisco, 2004

S. Cox, OpenGIS® Implementation Standard OGC 07-022r1: Observations and Measurements – Part 1 - Observation schema , OGC, 2007.

M. Egenhofer, Toward the semantic geospatial web, In GIS '02: Proceedings of the 10th ACM international symposium on Advances in geographic information systems (2002), pp. 1-4.

J. Farrell, H. Lausen, Semantic Annotations for WSDL and XML Schema, W3C Recommendation, 2007

-
- T. Gill, Building Semantic Bridges Between Museums, Libraries and Archives: The CIDOC Conceptual Reference Model, In First Monday, volume 9, number 5, <http://firstmonday.org>, 2004
- M. F. Goodchild, Interoperating geographic information systems, Kluwer Academic Publishers, Massachusetts, 1999
- T. R. Gruber, Toward Principles for the Design of Ontologies Used for Knowledge Sharing in International Journal Human-Computer Studies 43, S.907-928
- C. Hensos, J. Pschorr, A. Sheth, k. Thirunarayan, SemSOS: Semantic Sensor Observation Service, In: Proceedings of the 2009 International Symposium on Collaborative Technologies and Systems, IEEE Computer Society, Whashington, 2009
- R. Heutschi, Serviceorientierte Architektur: Architekturprinzipien und Umsetzung in die Praxis, Springer, Berlin, 2007
- N. Guarino, Formal ontology in information systems: proceedings of the First International Conference (FOIS'98), June 6-8, Trento, Italy, IOS Press, Amsterdam ,1998
- P. Hitzler, M. Krötzsch, S. Rudolph, Y. Sure, Semantic Web, Springer
- K. Janowicz, S. Schade, A. Bröring, C. Kesler, C. Stasch, P. Maué, T. Diekhof, A Transparent Semantic Enablement Layer for the Geospatial Web, In: TerraCognita 2009 Workshop In conjunction with the 8th International Semantic WebConference (ISWC 2009), Mexico City, 2009
- E. Klien, Semantic Annotation of Geographic Information, PhD Thesis, Institute for Geoinformatics, University of Muenster. Muenster, 2008
- W. Kuhn, Geospatial Semantics: Why, of What, and How? In: Journal on Data Semantics III, Springer, Berlin, 2005
- J. Lieberman, Geospatial Semantic Web Interoperability Experiment Report, 2006
- S. Löbner, Semantik – Eine Einführung, Walter de Gruyter GmbH & Co. KG, Berlin, 2003

P. Maué, S. Schade and P. Duchesne. OGC Discussion Paper 08-167r1: Semantic annotations in OGC standards. Technical report, OGC, 2009.

P. Mertens, F. Bodendorf, W. König, A. Picot, M. Schumann, T. Hess, Grundzüge der Wirtschaftsinformatik, Springer, Berlin, 2004

Object Management Group, Ontology Definition Metamodel – Version 1.0, OMG Document, 2009

L. Obrst, The Ontology Spectrum and Semantic Models. MITRE Information Semantics Group, Information Discovery & Understanding, Center for Innovative, Computing & Informatics, Bratislava, 2006

M. Polanyi, The tacit dimension, The University of Chicago Press, 1966

U. Roumois, Studienbuch Wissensmanagement, Orell Füssli Verlag AG, Zürich, 2007

C. Schmitz, B. Zucker, Wissensmanagement, Metropolitan Verlag, Berlin 2003

C. Stasch, K. Janowicz, A Stimulus-Centric Algebraic Approach to Sensors and Observations, In: Proceedings of the 3rd International Conference on GeoSensor Networks, Springer, Berlin, 2009

W. Stock, M. Stock, Wissensrepräsentation: Auswerten und bereitstellen von Informationen, Oldenbourg Wissenschaftsverlag, München, 2008

H. Stuckenschmidt, Ontologien: Konzepte, Technologien und Anwendungen, Springer, Berlin, 2009

T. Usländer, D. Hilbring, Catalogue Services Enabling Syntactical and Semantic

Interoperability in Environmental Risk Management Architectures, in „Managing Environmental Knowledge“, Proceedings of 20th International Conference on Informatics for Environmental Protection, Shaker Verlag, Aachen, 2006

A. Whiteside, OGC® Implementation Specification OGC 06-121r3: OGC Web Services Common Specification, OGC, 2007

H. Wilke, Systemisches Wissensmanagement, Lucius & Lucius, Stuttgart, 2001

7 Anhang

7.1 Beispielontologie

Das anschließende XML-Dokument gibt die verwendete Beispielontologie dar, welche unter Anderem auch unter der Referenz http://www.katzengeist.de/attachments/File/CAFE_Phenomena.owl zu finden ist.

```
<rdf:RDF
  xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:protege = "http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:xsp = "http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:assert = "http://www.owl-ontologies.com/assert.owl#"
  xmlns:owl = "http://www.w3.org/2002/07/owl#"
  xmlns:xsd = "http://www.w3.org/2001/XMLSchema#"
  xmlns = "http://www.umweltbundesamt.at/fileadmin/site/daten/Ontologien/SERONTO/Units_Dimensions20080701.owl#"
  xmlns:swrl = "http://www.w3.org/2003/11/swrl#"
  xmlns:swrlb = "http://www.w3.org/2003/11/swrlb#"
  xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"
  xml:base = "http://www.umweltbundesamt.at/fileadmin/site/daten/Ontologien/SERONTO/Units_Dimensions20080701.owl">
  <owl:Ontology rdf:about = ""/>
  <owl:Class rdf:ID = "Particulate">
    <rdfs:subClassOf>
      <owl:Class rdf:ID = "SuspendedSubstance"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about = "#SuspendedSubstance">
    <rdfs:subClassOf>
      <owl:Class rdf:ID = "physical"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID = "chemical_compound_or_ion">
    <rdfs:subClassOf>
      <owl:Class rdf:ID = "chemical"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID = "Matter"/>
  <owl:Class rdf:ID = "Substance">
    <rdfs:subClassOf rdf:resource = "#Matter"/>
  </owl:Class>
  <owl:Class rdf:ID = "Monitoring">
    <rdfs:subClassOf>
      <owl:Class rdf:ID = "EnvironmentalLaw"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID = "chemical_element">
    <rdfs:subClassOf>
      <owl:Class rdf:about = "#chemical"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about = "#chemical">
    <rdfs:subClassOf rdf:resource = "#Substance"/>
  </owl:Class>
  <owl:Class rdf:about = "#EnvironmentalLaw">
    <rdfs:subClassOf>
      <owl:Class rdf:ID = "HumanLaw"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about = "#HumanLaw">
    <rdfs:subClassOf>
      <owl:Class rdf:ID = "Law"/>
    </rdfs:subClassOf>
  </owl:Class>
  </owl:Ontology>
</rdf:RDF>
```

```

    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about = "#physical">
    <rdfs:subClassOf rdf:resource = "#Substance"/>
  </owl:Class>
  <owl:Class rdf:ID = "allotrope">
    <rdfs:subClassOf rdf:resource = "#chemical_compound_or_ion"/>
  </owl:Class>
  <owl:ObjectProperty rdf:ID = "hasComponent">
    <rdfs:range>
      <owl:Class>
        <owl:unionOf rdf:parseType = "Collection">
          <owl:Class rdf:about = "#physical"/>
          <owl:Class rdf:about = "#chemical"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:range>
    <owl:inverseOf>
      <owl:ObjectProperty rdf:ID = "hasReport"/>
    </owl:inverseOf>
    <rdfs:domain rdf:resource = "#EnvironmentalLaw"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about = "#hasReport">
    <rdfs:domain>
      <owl:Class>
        <owl:unionOf rdf:parseType = "Collection">
          <owl:Class rdf:about = "#physical"/>
          <owl:Class rdf:about = "#chemical"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
    <rdfs:range rdf:resource = "#EnvironmentalLaw"/>
    <owl:inverseOf rdf:resource = "#hasComponent"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:ID = "hasPart">
    <rdfs:domain rdf:resource = "#chemical_compound_or_ion"/>
    <rdfs:range rdf:resource = "#chemical"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:ID = "hasConstituent">
    <rdfs:range rdf:resource = "#chemical"/>
    <rdfs:domain rdf:resource = "#allotrope"/>
    <rdf:type>
      <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#FunctionalProperty">
        <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
      </rdf:Description>
    </rdf:type>
  </owl:ObjectProperty>
  <owl:DatatypeProperty rdf:ID = "hasOffset">
    <rdf:type rdf:resource = "http://www.w3.org/2002/07/owl#FunctionalProperty"/>
    <rdfs:range rdf:resource = "http://www.w3.org/2001/XMLSchema#float"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID = "hasSymbol">
    <rdfs:range rdf:resource = "http://www.w3.org/2001/XMLSchema#string"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID = "hasMultiple">
    <rdfs:domain rdf:resource = "#allotrope"/>
    <rdfs:range rdf:resource = "http://www.w3.org/2001/XMLSchema#int"/>
    <rdf:type rdf:resource = "http://www.w3.org/2002/07/owl#FunctionalProperty"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID = "hasMultiplier">
    <rdf:type rdf:resource = "http://www.w3.org/2002/07/owl#FunctionalProperty"/>
    <rdfs:range rdf:resource = "http://www.w3.org/2001/XMLSchema#float"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID = "hasDefinition">
    <rdfs:range rdf:resource = "http://www.w3.org/2001/XMLSchema#string"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID = "hasPower">
    <rdfs:range rdf:resource = "http://www.w3.org/2001/XMLSchema#float"/>
    <rdf:type rdf:resource = "http://www.w3.org/2002/07/owl#FunctionalProperty"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID = "hasValue">
    <rdfs:range rdf:resource = "http://www.w3.org/2001/XMLSchema#string"/>
  </owl:DatatypeProperty>
  <owl:FunctionalProperty rdf:ID = "hasLocation">
    <rdf:type>
      <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#DatatypeProperty">

```

```

    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:type>
    <rdfs:range rdf:resource = "http://www.w3.org/2001/XMLSchema#float"/>
  </owl:FunctionalProperty>
  <rdf:Description rdf:about = "http://www.w3.org/1999/02/22-rdf-syntax-ns#List">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <owl:AllDifferent/>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#Restriction">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2000/01/rdf-schema#Literal">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#AllValuesFromRestriction">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#MinCardinalityRestriction">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2000/01/rdf-schema#Datatype">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2000/01/rdf-schema#Container">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/1999/02/22-rdf-syntax-ns#Statement">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#HasValueRestriction">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <owl:AllDifferent/>
  <rdf:Description rdf:about = "http://www.w3.org/1999/02/22-rdf-syntax-ns#Seq">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#AnonymousClass">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#Thing">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <owl:AllDifferent/>
  <owl:distinctMembers rdf:parseType = "Collection">
    <Particulate rdf:ID = "PM10">
      <hasReport>
        <Monitoring rdf:ID = "CleanAirForEurope">
          <hasComponent>
            <chemical_compound_or_ion rdf:ID = "NO2">
              <hasReport rdf:resource = "#CleanAirForEurope"/>
              <hasPart>
                <chemical_element rdf:ID = "N"/>
              </hasPart>
              <hasPart>
                <allotrope rdf:ID = "O2">
                  <hasMultiple rdf:datatype = "http://www.w3.org/2001/XMLSchema#int">2</hasMultiple>
                  <hasConstituent>
                    <chemical_element rdf:ID = "O"/>
                  </hasConstituent>
                </allotrope>
              </hasPart>
            </chemical_compound_or_ion>
          </hasComponent>
          <hasComponent>
            <allotrope rdf:ID = "O3">
              <hasConstituent rdf:resource = "#O"/>
              <hasReport rdf:resource = "#CleanAirForEurope"/>
              <hasMultiple rdf:datatype = "http://www.w3.org/2001/XMLSchema#int">3</hasMultiple>
            </allotrope>
          </hasComponent>
          <hasComponent>
            <chemical_compound_or_ion rdf:ID = "SO2">
              <hasPart>
                <chemical_element rdf:ID = "S"/>
              </hasPart>
            </chemical_compound_or_ion>
          </hasComponent>
        </Monitoring>
      </hasReport>
    </Particulate>
  </owl:distinctMembers>

```

```

    <hasPart rdf:resource = "#O2"/>
    <hasReport rdf:resource = "#CleanAirForEurope"/>
  </chemical_compound_or_ion>
</hasComponent>
<hasComponent rdf:resource = "#PM10"/>
<hasComponent>
  <Particulate rdf:ID = "PM25">
    <hasReport rdf:resource = "#CleanAirForEurope"/>
  </Particulate>
</hasComponent>
<rdfs:comment rdf:datatype = "http://www.w3.org/2001/XMLSchema#string">CleanAirForEurope (CAFE)

```

Directive address (among others) several pollutants and thresholds which has to be monitored and reported to the European Environment Agency (EEA) </rdfs:comment>

```

  </Monitoring>
  </hasReport>
</Particulate>
  <Particulate rdf:about = "#PM25">
    <owl:distinctMembers>
</owl:AllDifferent>
<rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#TransitiveProperty">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<rdf:Description rdf:about = "http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<owl:AllDifferent>
<rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#IntersectionClass">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#AnnotationProperty">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#LogicalClass">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#CardinalityRestriction">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<owl:AllDifferent>
  <owl:distinctMembers rdf:parseType = "Resource">
    <rdf:rest rdf:parseType = "Resource">
      <rdf:rest rdf:parseType = "Resource">
        <rdf:first rdf:resource = "#S"/>
      </rdf:rest>
      <rdf:first rdf:resource = "#O"/>
    </rdf:rest>
    <rdf:first rdf:resource = "#N"/>
  </owl:distinctMembers>
</owl:AllDifferent>
<rdf:Description rdf:about = "http://www.w3.org/1999/02/22-rdf-syntax-ns#Alt">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#DeprecatedClass">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<owl:AllDifferent>
  <owl:distinctMembers rdf:parseType = "Collection">
    <Monitoring rdf:about = "#CleanAirForEurope"/>
  </owl:distinctMembers>
</owl:AllDifferent>
<rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#InverseFunctionalProperty">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#Nothing">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#DeprecatedProperty">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<rdf:Description rdf:about = "http://www.w3.org/2000/01/rdf-schema#Class">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#SymmetricProperty">
  <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
<rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#DataRange">

```

```

    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#ComplementClass">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#MaxCardinalityRestriction">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <owl:AllDifferent/>
  <rdf:Description rdf:about = "http://www.w3.org/1999/02/22-rdf-syntax-ns#Property">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#SomeValuesFromRestriction">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#EnumeratedClass">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#UnionClass">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#AllDifferent">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#ObjectProperty">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#Ontology">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
  <rdf:Description rdf:about = "http://www.w3.org/2002/07/owl#Class">
    <rdfs:label rdf:datatype = "http://www.w3.org/2001/XMLSchema#string"/>
  </rdf:Description>
</rdf:RDF>
<!--Created with Protege (with OWL Plugin 3.4.1, Build 536) http://protege.stanford.edu -->

```

7.2 Observation and Measurement

Wiedergegeben wird das Schema von *Observation and Measurement*, welches unter <http://www.opengeospatial.org/standards/om> publiziert ist. Aus Platzgründen, wurde auf die Dokumentation des Schemas verzichtet und an dieser Stelle auf die Spezifikation durch OGC verwiesen (Cox, 2009). Das Schema verweist zudem auf zahlreiche Unterschemen, welche hier ebenfalls nicht wiedergegeben werden können.

```

<?xml version="1.0" encoding="UTF-8"?>
<schema
  xmlns:xlink = "http://www.w3.org/1999/xlink"
  xmlns:gml = "http://www.opengis.net/gml"
  xmlns:gmd = "http://www.isotc211.org/2005/gmd"
  xmlns = "http://www.w3.org/2001/XMLSchema"
  xmlns:om = "http://www.opengis.net/om/1.0"
  xmlns:swe = "http://www.opengis.net/swe/1.0.1"
  xmlns:sml = "http://www.opengis.net/sensorML/1.0.1"
  targetNamespace = "http://www.opengis.net/om/1.0"
  elementFormDefault = "qualified"
  attributeFormDefault = "unqualified"
  version = "1.0.0">
  <annotation>
    <documentation>observation.xsd An XML implementation of the OandM model from OGC 07-022 Copyright (c) 2007 Open
    Geospatial Consortium - see http://www.opengeospatial.org/ogc/software</documentation>
  </annotation>

```

```

<import namespace = "http://www.opengis.net/gml" schemaLocation = "http://schemas.opengis.net/gml/3.1.1/base/gml.xsd"/>
<import namespace = "http://www.opengis.net/swe/1.0.1" schemaLocation =
"http://schemas.opengis.net/sweCommon/1.0.1/swe.xsd"/>
<import namespace = "http://www.opengis.net/sensorML/1.0.1" schemaLocation =
"http://schemas.opengis.net/sensorML/1.0.1/sensorML.xsd"/>
<complexType name = "ObservationType">
  <complexContent>
    <extension base = "gml:AbstractFeatureType">
      <sequence>
        <element
          name = "metadata"
          type = "om:AnyOrReferenceType"
          minOccurs = "0">
        </element>
        <element name = "samplingTime" type = "swe:TimeObjectPropertyType"/></element>
        <element
          name = "resultTime"
          type = "swe:TimeObjectPropertyType"
          minOccurs = "0">
        </element>
        <element name = "procedure" type = "om:ProcessPropertyType"/></element>
        <element
          name = "resultQuality"
          type = "om:AnyOrReferenceType"
          minOccurs = "0">
        </element>
        <element name = "observedProperty" type = "swe:PhenomenonPropertyType">
          <annotation/></annotation>
        </element>
        <element name = "featureOfInterest" type = "gml:FeaturePropertyType"/></element>
        <element
          name = "parameter"
          type = "swe:AnyDataPropertyType"
          minOccurs = "0"
          maxOccurs = "unbounded">
        </element>
        <element name = "result" type = "anyType"/></element>
      </sequence>
    </extension>
  </complexContent>
</complexType>
<element
  name = "Observation"
  type = "om:ObservationType"
  substitutionGroup = "gml:_Feature">
</element>
<complexType name = "ObservationPropertyType">
  <sequence minOccurs = "0">
    <element ref = "om:Observation"/>
  </sequence>
  <attributeGroup ref = "gml:AssociationAttributeGroup"/>
</complexType>
<complexType name = "ObservationCollectionType">
  <complexContent>
    <extension base = "gml:AbstractFeatureType">
      <sequence>
        <element
          name = "member"
          type = "om:ObservationPropertyType"
          maxOccurs = "unbounded"/>
      </sequence>
    </extension>
  </complexContent>
</complexType>
<element
  name = "ObservationCollection"
  type = "om:ObservationCollectionType"
  substitutionGroup = "gml:_Feature">
</element>
<complexType name = "ProcessPropertyType">
  <sequence minOccurs = "0">
    <choice>
      <element ref = "sml:_Process"/></element>
      <element name = "Process"/></element>
    </choice>
  </sequence>

```

```

    <attributeGroup ref = "gml:AssociationAttributeGroup"/>
</complexType>
<complexType name = "AnyOrReferenceType">
  <sequence minOccurs = "0">
    <any/>
  </sequence>
  <attributeGroup ref = "gml:AssociationAttributeGroup"/>
</complexType>
</schema>

```

Zusätzlich zum Schema soll hier ein kurzes Beispiel einer XML Dateninstanz, basierend auf O&M, wiedergegeben werden. Dieses Beispiel (timeSeries1.xml) entstammt wie auch das Schema der offiziellen OGC Spezifikation. (Cox, 2009)

```

<?xml version="1.0" encoding="UTF-8"?>
<om:Observation
  gml:id = "timeSeries1"
  xmlns:cv = "http://www.opengis.net/cv/0.2.1"
  xmlns:swe = "http://www.opengis.net/swe/1.0.1"
  xmlns:om = "http://www.opengis.net/om/1.0"
  xmlns:xsi = "http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xlink = "http://www.w3.org/1999/xlink"
  xmlns:gml = "http://www.opengis.net/gml"
  xsi:schemaLocation = "http://www.opengis.net/om/1.0 ../om.xsd http://www.opengis.net/cv/0.2.1
http://bp.schemas.opengis.net/06-188r1/cv/0.2.1/cv.xsd">
  <gml:description>Observation test instance - time series</gml:description>
  <gml:name>Time series 1</gml:name>
  <om:samplingTime>
    <gml:TimePeriod gml:id = "ts1t">
      <gml:beginPosition>2005-06-17T09:00:00+08:00</gml:beginPosition>
      <gml:endPosition>2005-06-21T09:00:00+08:00</gml:endPosition>
    </gml:TimePeriod>
  </om:samplingTime>
  <om:procedure xlink:href = "urn:ogc:object:feature:Sensor:BOM:t_2a"/>
  <om:observedProperty xlink:href = "http://sweet.jpl.nasa.gov/ontology/property.owl#Temperature"/>
  <om:featureOfInterest xlink:role = "urn:ogc:def:featureType:OGC:Station" xlink:href =
"http://my.big.org/feature?type=station%26name=st1"/>
  <om:parameter>
    <swe:Quantity definition = "http://sweet.jpl.nasa.gov/ontology/property.owl#Elevation">
      <swe:uom xlink:href = "urn:ogc:def:uom:UCUM:m"/>
      <swe:value>3.45</swe:value>
    </swe:Quantity>
  </om:parameter>
  <om:result>
    <cv:CompactDiscreteTimeCoverage>
      <cv:domainExtent xlink:href = "http://my.big.org/feature?type=station%26name=st1#xpointer(/boundedBy)"/>
      <cv:rangeType xlink:href = "http://sweet.jpl.nasa.gov/ontology/property.owl#Temperature"/>
      <cv:element>
        <cv:CompactTimeValuePair>
          <cv:geometry>2005-06-17T09:00:00+08:00</cv:geometry>
          <cv:value xsi:type = "gml:MeasureType" uom = "Cel">19.3</cv:value>
        </cv:CompactTimeValuePair>
      </cv:element>
      <cv:element>
        <cv:CompactTimeValuePair>
          <cv:geometry>2005-06-18T09:00:00+08:00</cv:geometry>
          <cv:value xsi:type = "gml:MeasureType" uom = "Cel">17.1</cv:value>
        </cv:CompactTimeValuePair>
      </cv:element>
      <cv:element>
        <cv:CompactTimeValuePair>
          <cv:geometry>2005-06-19T09:00:00+08:00</cv:geometry>
          <cv:value xsi:type = "gml:MeasureType" uom = "Cel">16.5</cv:value>
        </cv:CompactTimeValuePair>
      </cv:element>
      <cv:element>
        <cv:CompactTimeValuePair>
          <cv:geometry>2005-06-20T09:00:00+08:00</cv:geometry>
          <cv:value xsi:type = "gml:MeasureType" uom = "Cel">25.8</cv:value>
        </cv:CompactTimeValuePair>
      </cv:element>
    </cv:CompactDiscreteTimeCoverage>
  </om:result>
</om:Observation>

```

```

    </cv:CompactTimeValuePair>
  </cv:element>
  <cv:element>
    <cv:CompactTimeValuePair>
      <cv:geometry>2005-06-21T09:00:00+08:00</cv:geometry>
      <cv:value xsi:type = "gml:MeasureType" uom = "Cel">29.2</cv:value>
    </cv:CompactTimeValuePair>
  </cv:element>
</cv:CompactDiscreteTimeCoverage>
</om:result>
</om:Observation>

```

7.3 SensorML

Ergänzend zu O&M soll hiermit das Schema der verwendeten OGC Spezifikation *Sensor Model Language* (SensorML) wiedergegeben werden. Das komplette Schema kann an dieser Stelle nicht wiedergegeben wird und es wird daher auf <http://www.opengeospatial.org/standards/sensorml> für weitere Informationen verwiesen. Das Schema verweist zudem auf zahlreiche Unterschemen, welche hier ebenfalls nicht wiedergegeben werden können.

```

<?xml version="1.0" encoding="UTF-8"?>
<xs:schema
  xmlns:xs = "http://www.w3.org/2001/XMLSchema"
  xmlns:gml = "http://www.opengis.net/gml/3.2"
  xmlns:sml = "http://www.opengis.net/sensorML/1.0/gml32"
  xmlns:swe = "http://www.opengis.net/swe/1.0/gml32"
  targetNamespace = "http://www.opengis.net/sensorML/1.0/gml32"
  elementFormDefault = "qualified"
  attributeFormDefault = "unqualified"
  version = "1">
  <xs:annotation>
    <xs:documentation>
      SensorML document root definition
      Copyright (c) 2008 OGC, All Rights Reserved.
      For conditions, see OGC Software Notice - http://www.opengeospatial.org/legal/
    </xs:documentation>
  </xs:annotation>
  <!--===== -->
  <xs:import namespace = "http://www.opengis.net/gml/3.2" schemaLocation = "http://schemas.opengis.net/gml/3.2.1/gml.xsd"/>
  <xs:import namespace = "http://www.opengis.net/swe/1.0/gml32" schemaLocation = "../sweCommon/1.0.1_gml32/swe.xsd"/>
  <xs:include schemaLocation = "./system.xsd"/>
  <!--===== -->
  <xs:element name = "SensorML">
    <xs:annotation>
      <xs:documentation>SensorML document root</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:group ref = "sml:metadataGroup"/>
        <xs:element name = "member" maxOccurs = "unbounded">
          <xs:complexType>
            <xs:choice>
              <xs:element ref = "sml:AbstractProcess"/>
              <xs:element ref = "sml:DocumentList"/>
              <xs:element ref = "sml:ContactList"/>
            </xs:choice>
          </xs:complexType>
        </xs:element>
        <xs:attributeGroup ref = "gml:AssociationAttributeGroup"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>

```

```

    </xs:complexType>
  </xs:element>
</xs:sequence>
<xs:attribute
  name = "version"
  type = "xs:token"
  use = "required"
  fixed = "1.0.1"/>
</xs:complexType>
</xs:element>
</xs:schema>

```

Als ergänzendes Beispiel ist hiermit die bereits im Manuskript auszugsweise verwendete *Phenomena* Beschreibung aufgeführt. (Botts, 2007) Weitere Beispiele und Informationen finden sich in den bereits angesprochenen Spezifikationen der OGC.

```

<?xml version="1.0" encoding="UTF-8"?>
<gml:Dictionary
  xmlns:gml = "http://www.opengis.net/gml"
  xmlns:swe = "http://www.opengis.net/swe/1.0"
  xmlns:xlink = "http://www.w3.org/1999/xlink"
  xmlns:xsi = "http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation = "http://www.opengis.net/swe/1.0 ../swe.xsd"
  gml:id = "phenomena_swe">
  <gml:description>A dictionary of phenomena, compiled through OWS-1, OWS-1.2 OWS-3.SJDC 2005-10-03</gml:description>
  <gml:name codeSpace = "urn:ogc:tc:arch:doc-rp(05-010)">urn:ogc:def:property:OGC:ALL</gml:name>
  <gml:name>OWS Phenomena</gml:name>
  <!--=====-->
  <!--==== Simple Phenomenon examples =====>
  <!--=====-->
  <gml:dictionaryEntry>
    <swe:Phenomenon gml:id = "AtmosphericPressure">
      <gml:description xlink:href = "http://sweet.jpl.nasa.gov/ontology/property.owl#AtmosphericPressure">
        fluid pressure exerted due to the gravitational effect on the column of atmosphere above the
        position of interest
      </gml:description>
      <gml:name codeSpace = "urn:ogc:tc:arch:doc-rp(05-010)">urn:ogc:def:property:OGC:AtmosphericPressure</gml:name>
      <gml:name>Atmospheric Pressure</gml:name>
    </swe:Phenomenon>
  </gml:dictionaryEntry>
  <!--=====-->
  <gml:dictionaryEntry>
    <swe:Phenomenon gml:id = "Temperature">
      <gml:description xlink:href = "http://sweet.jpl.nasa.gov/ontology/property.owl#Temperature">
      </gml:description>
      <gml:name codeSpace = "urn:ogc:tc:arch:doc-rp(05-010)">urn:ogc:def:property:OGC:Temperature</gml:name>
      <gml:name>Temperature</gml:name>
    </swe:Phenomenon>
  </gml:dictionaryEntry>
  <!--=====-->
  <!--==== Constrained Phenomenon example =====>
  <!--=====-->
  <gml:dictionaryEntry>
    <swe:ConstrainedPhenomenon gml:id = "WaterTemperature">
      <gml:name codeSpace = "urn:ogc:tc:arch:doc-rp(05-010)">urn:ogc:def:property:OGC:WaterTemperature</gml:name>
      <gml:name>Water Temperature</gml:name>
      <swe:base xlink:href = "#Temperature"/>
      <swe:singleConstraint>
        <swe:TypedValue>
          <swe:property codeSpace = "/">Medium</swe:property>
          <swe:value xsi:type = "gml:CodeType" codeSpace = "http://www.opengis.net/ows/material">Water</swe:value>
        </swe:TypedValue>
      </swe:singleConstraint>
    </swe:ConstrainedPhenomenon>
  </gml:dictionaryEntry>
  <!--=====-->
  <gml:dictionaryEntry>
    <swe:ConstrainedPhenomenon gml:id = "SurfaceWaterTemperature">

```

```

    <gml:name codeSpace = "urn:ogc:tc:arch:doc-rp(05-
010)">urn:ogc:def:property:OGC:SurfaceWaterTemperature</gml:name>
    <gml:name>Surface Water Temperature</gml:name>
    <swe:base xlink:href = "#WaterTemperature"/>
    <swe:singleConstraint>
      <swe:TypedValue>
        <swe:property codeSpace = "/">Depth</swe:property>
        <swe:value>
          <swe:Interval>
            <swe:lowerBound xsi:type = "gml:MeasureType" uom = "/units.xml#m">0.0</swe:lowerBound>
            <swe:upperBound xsi:type = "gml:MeasureType" uom = "/units.xml#m">1.5</swe:upperBound>
          </swe:Interval>
        </swe:value>
      </swe:TypedValue>
    </swe:singleConstraint>
  </swe:ConstrainedPhenomenon>
</gml:dictionaryEntry>
<!--===== -->
<!--==== Composite Phenomenon example ===== -->
<!--===== -->
<gml:dictionaryEntry>
  <gml:name>Earthquake Parameters</gml:name>
  <swe:component>
    <swe:CompositePhenomenon gml:id = "EarthquakeLocation" dimension = "3">
      <gml:name codeSpace = "urn:ogc:tc:arch:doc-rp(05-010)">urn:ogc:def:property:OGC:EarthquakeLocation</gml:name>
      <gml:name>Earthquake Location</gml:name>
      <swe:component xlink:href = "#Epicentre"/>
      <swe:component xlink:href = "#Depth"/>
      <swe:component xlink:href = "#OriginTime"/>
    </swe:CompositePhenomenon>
  </swe:component>
  <swe:component>
    <swe:CompositePhenomenon gml:id = "MomentTensor" dimension = "6">
      <gml:name codeSpace = "urn:ogc:tc:arch:doc-rp(05-010)">urn:ogc:def:property:OGC:MomentTensor</gml:name>
      <gml:name>Earthquake Moment Tensor</gml:name>
      <swe:component xlink:href = "#Mrr"/>
      <swe:component xlink:href = "#Mtt"/>
      <swe:component xlink:href = "#Mff"/>
      <swe:component xlink:href = "#Mrt"/>
      <swe:component xlink:href = "#Mrf"/>
      <swe:component xlink:href = "#Mtf"/>
    </swe:CompositePhenomenon>
  </swe:component>
</gml:dictionaryEntry>
</gml:Dictionary>

```