



Repräsentation von Wissensorganisations- systemen (KOS) im Semantic Web: Ein Best Practice Guide

Verfasser: Klaas Dellschaft und Christian Hachenberg, Universität Koblenz-Landau
Veröffentlicht am: 30. September 2011

Inhaltsverzeichnis

1	Einleitung	1
1.1	Eigenschaften von Wissensorganisationssystemen (KOS)	2
1.2	Modellierungsebenen eines Wissensorganisationssystems (KOS)	3
1.3	Verwendete Wissensorganisationssysteme (KOS)	4
1.3.1	Die Schlagwortnormdatei (SWD)	4
1.3.2	Der Thesaurus Technik und Management (TEMA)	5
1.3.3	Die Dewey Dezimalklassifikation (DDC)	5
2	Linked Data und das Semantic Web	5
2.1	Die Linked Data-Prinzipien	6
2.2	Standards im Semantic Web: RDF, RDFS und OWL	6
2.3	Benutzerinterfaces zur Erkundung von Linked Data	7
2.3.1	Generelle Benutzerinterfaces	7
2.3.2	Anwendungsspezifische Benutzerinterfaces	9
3	Repräsentation von Wissensorganisationssystemen im Semantic Web	10
3.1	Transformation des KOS-Schemas	10
3.2	Transformation der KOS-Daten	11
3.3	Transformation der Annotierungen	11
3.4	Verwendung von Metamodellierung	12
4	Vorgehensmodelle für Linked Data-Projekte	13
4.1	Allgemeines Phasenmodell für Wissensmanagement-Applikationen	14
4.1.1	Machbarkeitsstudie	14
4.1.2	Kickoff	15
4.1.3	Verfeinerung und Anpassung	15
4.1.4	Evaluation	16
4.1.5	Anwendung und Weiterentwicklung	16
4.2	Konkrete Schritte zur Transformation eines KOS	17
4.2.1	Analyse des ursprünglichen KOS-Schemas	17
4.2.2	Mapping auf das Zielschema im Semantic Web	17
4.2.3	Transformation der KOS-Daten	19
5	Umsetzung der Linked Data-Prinzipien	19
5.1	Entwurfsmuster für den URI-Namespace	20
5.2	Bereitstellung unterschiedlicher Repräsentationen	21
5.3	Abfragen über Linked Data	23
6	Transformation der Schlagwortnormdatei	23
6.1	Einordnung in das allgemeine Phasenmodell	24
6.2	Analyse des Schemas der SWD	24
6.3	Mapping auf das Zielschema im Semantic Web	24
6.4	Transformation der Daten der SWD	25
7	Transformation des Thesaurus Technik und Management	27
7.1	Analyse des Schemas von TEMA	27
7.2	Mapping auf das Zielschema im Semantic Web	28
7.3	Transformation der Daten von TEMA	29
8	Zusammenfassung	33

1 Einleitung

In diesem Dokument sollen Begriffe, Prinzipien und Methoden vorgestellt werden, die sich als hilfreich bei der Erstellung von Semantic Web konformen Repräsentationen von Wissensorganisationssystemen (KOS) erwiesen haben, wie z. B. Thesauri und Klassifikationssysteme. Das Dokument richtet sich an Organisationen wie z. B. Bibliotheken, die ihre traditionellen Wissensorganisationssysteme im Rahmen des Semantic Web veröffentlichen wollen. Die in diesem Dokument beschriebenen Vorgehensweisen und Prinzipien sind nicht als normativ anzusehen. Sie sollen nur dabei helfen, von bisher gemachten Erfahrungen zu profitieren und einen leichteren Einstieg in die wichtigsten Begrifflichkeiten und Techniken des Semantic Web zu bekommen. An vielen Stellen wird zudem auf weiterführende Literatur zum Thema und auf relevante Standards und Spezifikationen aus dem Bereich des Semantic Web hingewiesen.

Im Zusammenhang mit Semantic Web konformen Repräsentationen von Datenbeständen hat sich in den letzten Jahren der Begriff von *Linked Data* eingeprägt. Unter diesem Begriff werden einige Entwurfsprinzipien zusammengefasst, die bei der Bereitstellung von Datenbeständen im Semantic Web beachtet werden sollten (siehe Abschnitt 2.1). Insbesondere ist dabei hervorzuheben, dass alle Elemente in einem solchen Datenbestand durch eine dereferenzierbare URI identifiziert werden und nach Möglichkeit Verbindungen zu anderen Datenbeständen bereitgestellt werden sollten. Auf Grund dieser Verbindungen zwischen den Datenbeständen spricht man häufig auch von der *Linked Data-Wolke* [22], in der die verschiedenen Datenbestände und deren Verbindungen untereinander visualisiert werden.

Im Folgenden wollen wir zuerst einige wichtige Begriffe klären, die für das bessere Verständnis des Dokuments hilfreich sind. In Abschnitt 1.1 werden unterschiedliche Eigenschaften von Wissensorganisationssystemen (KOS) geklärt, die u. a. einen Einfluss auf die spätere Transformation haben können. Anschließend werden kurz die unterschiedlichen Modellierungsebenen eines KOS verdeutlicht. Die Unterscheidung zwischen diesen Modellierungsebenen findet sich auch später bei den Vorgehensweisen wieder, z. B. in Abschnitt 4.2.

Der Rest dieses Dokuments ist wie folgt strukturiert: In Abschnitt 2 werden die grundlegenden Prinzipien von Linked Data erläutert und es wird eine kurze Einführung in Standards für das Semantic Web gegeben, wie z. B. RDF, RDF Schema und die Web Ontology Language (OWL). Außerdem werden weitere Anforderungen an Linked Data-Repräsentationen von Datenbeständen abgeleitet, indem die unterschiedlichen Zugriffsmethoden durch Benutzerinterfaces für Linked Data analysiert werden.

Abschnitt 3 gibt ein Überblick, wie die einzelnen Modellierungsebenen eines Wissensorganisationssystems (KOS) in eine Semantic Web konforme Repräsentation überführt werden können. Die SKOS-Spezifikation [12] spielt dabei eine besondere Rolle, es wird aber auch auf alternative Spezifikationen eingegangen, die ebenfalls für die Repräsentation von KOS genutzt werden können.

In Abschnitt 4 wird zum einen ein allgemeines Phasenmodell vorgestellt, welches sich für die Durchführung von Projekten zur Erstellung von Wissensmanagement-Applikationen als hilfreich erwiesen hat. Zum anderen werden drei konkrete Schritte vorgestellt, die bei der Transformation eines KOS durchlaufen werden sollten.

In Abschnitt 5 werden Best Practices vorgestellt, die bei der Umsetzung der Linked Data-Prinzipien aus Abschnitt 2.1 helfen. Diese Best Practices betreffen (1) den Entwurf des URI-Namespaces zur Identifizierung von Elementen im Datenbestand und (2) Handlungsoptionen wie neben der durch Maschinen lesbaren Repräsentation der Daten auch eine für Menschen lesbare Repräsentation zur Verfügung gestellt werden kann.

Abschließend werden diese Prinzipien und Methoden anhand von zwei konkreten Projekten verdeutlicht. In Abschnitt 6 wird das an der Deutschen Nationalbibliothek durchgeführte Projekt beschrieben, in dessen Verlauf u. a. für die Schlagwortnormdatei (SWD) eine Linked Data-Repräsentation erstellt wurde. Dieses Projekt ist geeignet um sowohl das allgemeine Phasenmodell als auch die drei konkreten Schritte zur Transformation von Wissensorganisationssystemen (KOS) zu verdeutlichen. Weitere Details zu diesem Projekt sind auch den Dokumenten [17, 27] zu entnehmen. Ergänzend wird in Abschnitt 7 die Transformation des Thesaurus Technik und Management sowie der Fachordnungsklassifikation der WTI Frankfurt eG beschrieben. Es werden die drei konkreten Schritte zur Transformation von KOS noch einmal genauer ausgeführt und angewendet. Für den Thesaurus Technik und Management fehlt aber die Einordnung in ein übergeordnetes Projekt. Beispiele für weitere konkrete Projekte sind zu finden in [32, 33, 35, 39, 43].

1.1 Eigenschaften von Wissensorganisationssystemen (KOS)

Grundsätzlich haben Wissensorganisationssysteme unterschiedliche Eigenschaften, die eine Transformation in eine Linked Data-Repräsentation beeinflussen. Neben dem konkreten Typ des Wissensorganisationssystems, wie z. B. ob es ein Thesaurus oder ein Klassifikationssystem ist, ist noch die Unterscheidung wichtig, ob das Wissensorganisationssystem ein hohes Maß an Präkombination aufweist oder ob es eher post-koordiniert eingesetzt wird. Ein weiterer wichtiger Punkt ist die Unterscheidung zwischen termbasierten und konzeptbasierten Wissensorganisationssystemen.

Ein hohes Maß an **Präkombination** liegt vor, wenn bereits innerhalb des Wissensorganisationssystems viele der bestehenden Terme oder Konzepte miteinander kombiniert werden, um so z. B. komplexe Themen darzustellen. Ein Beispiel für ein Wissensorganisationssystem mit einem hohen Maß an Präkombination ist die Dewey Dezimalklassifikation (siehe Abschnitt 1.3). Sie enthält ein genaues Regelwerk, wie verschiedene Notationen miteinander kombiniert und zu neuen Notationen zusammengefasst werden dürfen.

Ein Wissensorganisationssystem (KOS) wird eher **post-koordiniert** eingesetzt, wenn die Kombination von mehreren Konzepten oder Termen zur Repräsentation von komplexen Themen nicht bereits im KOS selber durchgeführt wird, sondern erst im Kontext der Verwendung bei z. B. den Titeldaten. Die post-koordinierte Verwendung hat den Vorteil, dass während des Retrievals auch mehrere Terme oder Konzepte mit Hilfe von logischen Operatoren miteinander verknüpft werden können, um so eine höhere Retrievalgenauigkeit zu erzielen. Ein Beispiel für ein post-koordiniert eingesetztes Wissensorganisationssystem ist die Schlagwortnormdatei (siehe Abschnitt 1.3), deren Schlagwörter in Form von Schlagwortfolgen in Titeldaten verwendet werden.

Bei **termbasierten Wissensorganisationssystemen** sind Terme bzw. Wörter oder Phrasen die Grundelemente, die jeweils in Bezug zueinander gesetzt werden. Zum Beispiel können hierarchische Beziehungen wie *broader term* existieren oder Verknüpfungen zwischen Deskriptoren und Nicht-Deskriptoren wie *use* oder *use for*.

Bei **konzeptbasierten Wissensorganisationssystemen** bilden die Konzepte die Grundelemente. Konzepte entsprechen tendenziell abstrakten Themen, die jeweils einen künstlich generierten, eindeutigen Namen haben. Jedem Konzept sind jedoch auch ein oder mehrere Terme bzw. Label zugeordnet. Zwischen den Konzepten existieren Beziehungen wie z. B. *broader concept* oder *related concept*, während zwischen Termen Beziehungen wie z. B. *use*, *use for* oder *translation of* existieren können.

Die genaue Ausprägung von Beziehungen und Elementen hängt stark vom konkreten Wissensorganisationssystem ab. Aus diesem Grund sind wichtige Punkte bei der Transfor-

mation von Wissensorganisationssystemen auch immer die Erfassung des ursprünglichen KOS-Schemas (siehe Abschnitt 4.2.1, 6.2 und 7.1) und die Auswahl eines passenden Ziel-schemas im Semantic Web (siehe die Diskussion in Abschnitt 3.1 und 4.2.2).

1.2 Modellierungsebenen eines Wissensorganisationssystems (KOS)

Bei der Transformation von KOS und den hierdurch annotierten Titeldaten können grundsätzlich verschiedene Modellierungsebenen unterschieden werden. Die Unterscheidung zwischen diesen Modellierungsebenen wird später relevant, wenn es um die Abbildung von KOS auf Semantic Web Standards wie RDF, RDFS und OWL geht (siehe Abschnitt 2.2):

Das **KOS-Schema** beschreibt die dem KOS zugrundeliegenden syntaktischen Elemente und ihre Relationen untereinander. Im Fall eines konzeptbasierten KOS können z. B. Konzepte durch verschiedene Relationen wie Ober-/Unterkonzept in Beziehung zueinander gesetzt werden. Weiterhin kann noch im Schema eines KOS definiert sein, dass Konzepte eine eindeutige Vorzugsbenennung haben. Auf der Grundlage dieses Schemas werden dann die eigentlichen Daten eines KOS organisiert, wobei das Schema je nach KOS-Typ (z. B. Thesaurus oder Klassifikationssystem) und Ausprägung größere Unterschiede aufweisen kann.

Die **KOS-Daten** entsprechen dem eigentlichen Inhalt von z. B. einem Thesaurus oder einem Klassifikationssystem. Beispielsweise beinhaltet der Thesaurus *Technik und Management* (siehe Abschnitt 1.3) u. a. Sachschlagwörter aus dem Bereich Technik. Diese konkreten Sachschlagwörter, wie z. B. *Kathode* oder *Ausfallrate*, und deren Beziehungen untereinander sind Bestandteil der Datenebene des Thesaurus.

Häufig werden die KOS-Daten verwendet, um Ressourcen entsprechend zu annotieren. Im Falle der Katalogisierung in Bibliotheken können z. B. in den Titeldaten Schlagwörter aus einem Thesaurus zur inhaltlichen Beschreibung des Titels verwendet werden. Auch im Fall der **Annotierungen** kann grundsätzlich zwischen dem Schema für die Annotierungen und den eigentlichen Daten unterschieden werden. Auf der Schema-Ebene der Annotierungen könnte z. B. festgelegt sein, dass Schlagwörter mit Hilfe von `dc:subject` mit den Titeldaten verknüpft werden. Auf der Datenebene befände sich dann die Information, dass der *Varta-Führer Deutschland* durch das Schlagwort *BRD* beschrieben wird.

Im Fall der post-koordinierten Verwendung von z. B. einem Thesaurus könnte es auch sein, dass auf der Daten-Ebene der Annotierungen einzelne Schlagwörter zur Darstellung von komplexen Themen kombiniert werden. Ein Beispiel hierfür sind die Schlagwortketten, wie sie in den Regeln für den Schlagwortkatalog beschrieben sind [16]. Die Kombination von Schlagwörtern zu komplexen Themen kann aber auch bereits auf der Modellierungsebene der KOS-Daten geschehen. In diesem Fall spricht man von präkombinierten Schlagwörtern. Diese Repräsentation von komplexen Themen auf unterschiedlichen Modellierungsebenen wird jedoch problematisch, wenn man ein Mapping herstellen will zwischen KOS, die hauptsächlich post-koordiniert verwendet werden, und KOS, die einen hohen Anteil an präkombinierten Termen und Konzepten aufweisen (siehe [23]).

Der Zusammenhang zwischen KOS-Schema, KOS-Daten und den z. B. in den Titeldaten vorgenommenen Annotierungen ist noch einmal in Abbildung 1 dargestellt. In diesem Dokument werden wir uns vornehmlich mit KOS-Schemata und -Daten befassen, die zur Darstellung von Sachschlagwörtern geeignet sind. Insbesondere das KOS-Schema müsste aber für andere Arten von Daten, wie z. B. Personen oder Körperschaften, angepasst werden oder auch für verschiedene Typen von KOS, wie z. B. Klassifikationssysteme.

Es ist ausdrücklich nicht Ziel dieses Best Practice-Guides konkrete Muster vorzuschlagen, wie ein bestimmtes Konstrukt auf der Schema-Ebene eines KOS in z. B. SKOS oder

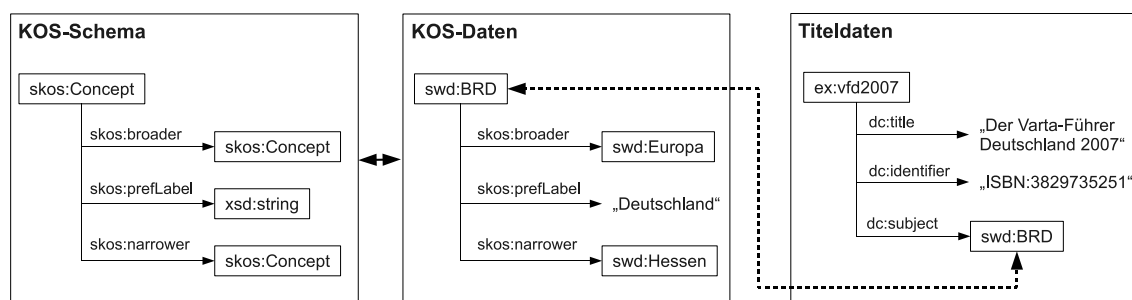


Abbildung 1: Zusammenhang zwischen KOS-Schema, KOS-Daten und der Verknüpfung mit den Titeldaten.

OWL abgebildet werden kann! Für diese Art von Muster sei auf bereits bestehende Literatur verwiesen (siehe z. B. [32, 33, 35, 39, 43]). Es ist stattdessen Ziel dieses Best Practice-Guides generelle Anforderungen, Vorgehensweise und Prinzipien zusammenzufassen, die bei der Transformation beliebiger KOS beachtet werden sollten (siehe Abschnitt 2, 3, 4 und 5). Diese generellen Anforderungen, Vorgehensweisen und Prinzipien werden dann in Abschnitt 6 und 7 anhand von zwei konkreten Projekten verdeutlicht.

1.3 Verwendete Wissensorganisationssysteme (KOS)

Die in diesem Dokument vorgestellten Best Practices werden in Abschnitt 6 angewendet auf die Schlagwortnormdatei (SWD) und in Abschnitt 7 auf den Thesaurus Technik und Management. In Abschnitt 6 wird kurz die Problematik des Mappings zwischen hauptsächlich post-kodiert verwendeten KOS wie der SWD und KOS mit einem hohen Grad der Präkombination wie der Dewey Dezimalklassifikation verdeutlicht. Weitere Details zu dieser Problematik sind [23] zu entnehmen. Im Folgenden sollen kurz alle drei Datenbestände vorgestellt werden:

1.3.1 Die Schlagwortnormdatei (SWD)

Die Schlagwortnormdatei (SWD)¹ ist ein normierter, terminologisch kontrollierter Wortschatz. Sie enthält Schlagwörter aus allen Fachgebieten und Schlagwortkategorien, die durch die beteiligten Bibliotheken zur verbalen Inhaltserschließung genutzt und dabei täglich aktualisiert werden. Unter anderem wird die Schlagwortnormdatei auch zur verbalen Erschließung der Titel verwendet, die im Rahmen der Deutschen Nationalbibliografie veröffentlicht werden.

Bei der verbalen Erschließung wird jedem Titel mindestens eine Schlagwortfolge zugewiesen. Schlagwortfolgen werden dazu benutzt, um die komplexen Themen eines Dokuments zu beschreiben. Die Katalogisierung erfolgt also durch Post-Koordination von Schlagwörtern zu präkoordinierten Schlagwortfolgen gemäß den Regeln für Schlagwortkatalog (RSWK). Erst das Zusammenführen der einzelnen Schlagwörter in der Schlagwortfolge entspricht dem speziellen Thema eines Werkes. Das Auffinden dieser Themen erfolgt über eine präkoordinierte Suche der Schlagwortfolgen. Die üblichen Retrievalumgebungen erlauben auch die Postkoordination von Schlagwörtern bei der Suche. Dabei geht aber der spezifische Sinnzusammenhang, der durch die Schlagwortfolge generiert wurde verloren. Zusätzlich wird bei der Erschließung im Rahmen des Neuerscheinungsdiensts der Deutschen Nationalbibliografie jedem Titel noch mindestens eine Sachgruppe zugewiesen.

¹<http://www.d-nb.de/standardisierung/normdateien/swd.htm>

Seit dem Jahrgang 2004 orientieren sich diese Sachgruppen weitestgehend an der zweiten Ebene der Dewey-Dezimalklassifikation.

1.3.2 Der Thesaurus Technik und Management (TEMA)

Der Thesaurus Technik und Management (TEMA)² der WTI Frankfurt eG ist ein Thesaurus, der über 150.000 Fachbegriffe aus verschiedenen technisch orientierten Anwendungsdomänen enthält. Der Thesaurus ist termbasiert. Insgesamt wird in dem Thesaurus unterschieden zwischen Vorzugsbenennungen, Synonymen, versteckten Synonymen und englischen Synonymen. Alle diese verschiedenen Synonymarten werden in Bezug zu der Vorzugsbenennung gesetzt.

Während der Erschließung werden mehrere Begriffe aus dem Thesaurus einem Dokument zugeordnet. Im Gegensatz zu der Schlagwortfolge bei der SWD sind aber die einzelnen Begriffe nicht in einen bestimmten Bezug zueinander gesetzt, sondern die einzelnen Begriffe stehen gleichberechtigt nebeneinander. Zusätzlich zu den Begriffen aus dem Thesaurus werden jedem Dokument noch Fachgebiete zugeordnet, unter denen dieses Dokument klassifiziert werden kann. Diese Fachgebiete sind vergleichbar mit den Sachgruppen bei der Schlagwortnormdatei.

1.3.3 Die Dewey Dezimalklassifikation (DDC)

Die Dewey-Dezimalklassifikation (DDC)³ ist eine international weitverbreitete Universalklassifikation. Von ihr existieren zahlreiche Übersetzungen, u. a. ins Deutsche. Seit Januar 2006 wird die DDC neben der SWD zur Inhaltserschließung in der Deutschen Nationalbibliografie verwendet.

Bei der Inhaltserschließung mit Hilfe der DDC wird jedem Inhalt eine oder mehrere DDC-Notationen zugewiesen. Facetten eines Themas werden durch die Synthese von zwei oder mehr Notationen zu einer neuen Notation dargestellt. Somit ist die DDC-Notation sowohl vergleichbar mit einzelnen Schlagwörtern aus der SWD als auch mit den Schlagwortfolgen, wobei letztere zur Darstellung von komplexen Themen verwendet werden und somit ihre Entsprechung in den synthetischen Notationen haben.

Die DDC ist eine enumerative Klassifikation, die aber auch in festgelegten Bereichen eine Facettierung vorsieht. D. h. ein Thema kann durch die Synthese von zwei oder mehr Notationen zu einer neuen synthetischen Notation dargestellt werden. DDC und SWD sind nicht vergleichbar. In der DDC repräsentiert die Notation einer Klasse häufig ein Thema, das nach RSWK mit Hilfe der SWD nur durch die Kombination verschiedener Schlagwörter in einer Schlagwortfolge wiedergegeben werden kann, weil der ganze Hierarchiebaum die Bedeutung der Klasse ausmacht. Schlagwörter in der SWD sind zunächst nicht auf einen bestimmten Sinnzusammenhang festgelegt. Dieser wird erst in der Kombination mit anderen Schlagwörtern in der Schlagwortfolge erzeugt. Andererseits kann aber auch ein einzelnes Schlagwort der SWD den Inhalt einer synthetischen Notation wiedergeben.

2 Linked Data und das Semantic Web

In diesem Abschnitt soll eine kurze Einführung in die Grundlagen von Linked Data und dem Semantic Web gegeben werden. In Abschnitt 2.1 werden die vier Prinzipien kurz zu-

²<http://www.wti-frankfurt.de/index.php/thesaurus>

³<http://www.oclc.org/dewey/> und <http://www.ddc-deutsch.de/>

sammengefasst, die bei der Veröffentlichung von Linked Data beachtet werden sollten und in Abschnitt 2.2 werden die im Semantic Web verwendeten Standards wie RDF, RDF Schema und die Web Ontology Language vorgestellt. Danach wird in Abschnitt 2.3 diskutiert, wie auf Linked Data durch einen Endbenutzer zugegriffen werden kann und welche zusätzlichen Anforderungen an eine Transformation von Datenbeständen nach Linked Data sich dadurch ergeben.

2.1 Die Linked Data-Prinzipien

Der Begriff *Linked Data* wurde 2006 durch Tim Berners-Lee in einem Artikel geprägt [20]. In diesem Artikel wurden die vier Prinzipien beschrieben, die bei der Veröffentlichung von maschinenlesbaren Daten im Semantic Web beachtet werden sollten:

1. Für die Bezeichnung einzelner Entitäten (Schlagwörter, Personen, Orte, ...) sollen URIs (Uniform Resource Identifier) verwendet werden.
2. Die URIs sollen auf dem HTTP-Protokoll basieren, um diese auf einfache Art dereferenzierbar zu machen. Dieser Mechanismus kann dazu benutzt werden, um weitere Informationen über eine Entität nachzuschlagen.
3. Wenn eine URI dereferenziert wird, sollen weiterführende Informationen über die dadurch bezeichnete Entität zurückgeliefert werden. Diese Informationen sollen in standardisierten Datenformaten wie z. B. RDF, RDFS oder OWL vorliegen und nach Möglichkeit auch durch Anfragesprachen wie z. B. SPARQL abrufbar sein.
4. Die Informationen über eine Entität sollen nach Möglichkeit Links zu anderen Entitäten enthalten. Diese anderen Entitäten müssen nicht notwendigerweise aus der gleichen Datenquelle stammen. Es soll auch zu externen Datenquellen verlinkt werden.

Die Berücksichtigung der angeführten Prinzipien soll dazu führen, dass nicht nur einfach Daten im Semantic Web veröffentlicht werden, sondern dass sowohl Menschen als auch Maschinen durch das Verfolgen von Links das Semantic Web erkunden können.

2.2 Standards im Semantic Web: RDF, RDFS und OWL

Neben dieser Verlinkung von Informationen im Semantic Web ist es auch wichtig, dass standardisierte Austauschformate verwendet werden. Als Standardformat für Daten im Semantic Web ist hierbei RDF anzusehen (Resource Description Framework). Es ermöglicht die Beschreibung beliebiger Daten mittels dreiteiliger Tupel [9]. Diese Tupel bestehen immer aus einem Subjekt S , einem Prädikat P und einem Objekt O .

Das Subjekt ist immer eine durch eine URI identifizierte Entität, über die weitergehende Informationen bereitgestellt werden. Das Prädikat entspricht einer Eigenschaft des Subjekts und wird selbst ebenfalls durch eine URI repräsentiert. Das Objekt enthält den Wert dieser Eigenschaft, wobei das Objekt sowohl auf eine andere Entität mittels einer URI verweisen als auch ein einfaches Literal beinhalten kann (z. B. einen String, ein Datum oder einen Integer-Wert).

Prinzipiell können beliebige URIs zur Bezeichnung der Subjekte, Prädikate und Objekte verwendet werden. Zu beachten ist jedoch, dass ein Großteil des Potentials im Semantic Web darin liegt, dass nicht in allen Fällen eigene Prädikate verwendet werden, sondern dass man sich auf ein gemeinsames Schema einigt. Dies kann beispielsweise die Frage betreffen, wie die Daten über Thesauri abgelegt werden sollen. Ein solches Schema kann im

Rahmen von RDF Schema (RDFS) oder der Web Ontology Language (OWL) beschrieben werden, die beide auf der RDF-Spezifikation aufbauen.

In einem solchen Schema werden u.a. alle im Datenbestand verwendeten Prädikate genauer beschrieben. So kann einem Prädikat eine natürlichsprachige Beschreibung zugewiesen werden, es kann von anderen Prädikaten durch Spezialisierung abgeleitet werden oder es kann als transitiv definiert werden. Es ist außerdem möglich in RDFS bzw. OWL eine Klassenhierarchie zu definieren. Weitere Details über die Möglichkeiten von RDFS und OWL sind den entsprechenden Spezifikationen zu entnehmen (siehe [6, 10]).

Insgesamt kann man also bei der Repräsentation von Daten im Semantic Web zwischen der Schema-Ebene und der Daten-Ebene unterscheiden. Bei der Repräsentation von Thesauri könnte z. B. erstens auf der Schema-Ebene definiert sein, dass ein Thesaurus Konzepte enthält und dass Konzepte über `skos:prefLabel`⁴ mit einem natürlichsprachigen Label verknüpft sind (siehe Abbildung 1 in Abschnitt 1.2). Zweitens könnte noch im Schema definiert sein, dass Konzepte über die Eigenschaft `skos:broader` miteinander verknüpft sind. Auf der Daten-Ebene des Thesaurus ist dann z. B. definiert, dass das Konzept `swd:BRD` das Label *Deutschland* hat und ein Unterkonzept von `swd:Europa` ist.

2.3 Benutzerinterfaces zur Erkundung von Linked Data

Generell kann zwischen zwei verschiedenen Typen von Benutzerinterfaces für die Erkundung von Daten aus dem Semantic Web unterschieden werden:

Zum einen gibt es die generellen Benutzerinterfaces, die unabhängig vom Schema der Daten arbeiten und keinerlei Annahmen über die zu erkundenden Datenbestände machen. Mit diesen Benutzerinterfaces kann prinzipiell die komplette Linked Data-Wolke erkundet werden. Allerdings sind diese Interfaces nur sehr eingeschränkt für den Endbenutzer geeignet. Sie richten sich eher an professionelle Benutzer.

Zum anderen gibt es die anwendungsspezifischen Benutzerinterfaces, welche in der Regel nur mit Datenbeständen umgehen, die mit Hilfe eines oder mehrerer bestimmter Schemata abgelegt wurden. Diese sind meist auf wenige, miteinander verbundene Datenquellen eingeschränkt. Solche spezialisierten Benutzerinterfaces können prinzipiell die komplette Semantik der Datenbestände nutzen und sind in der Regel auch für den Endbenutzer geeignet.

2.3.1 Generelle Benutzerinterfaces

Beispiele für generelle Benutzerinterfaces zur Erkundung von Linked Data sind die beiden Suchmaschinen *Sig.ma* [41] und *VisiNav* [28]. Bei derartigen Suchmaschinen startet der Suchprozess entweder mit einer allgemeinen Stichwortsuche oder der Eingabe einer bestimmten Ressource (identifiziert durch eine URL), zu dem weitere Informationen gesucht werden sollen.

Anschließend kann das so erzielte Suchergebnis durch zwei verschiedene Arten der Navigation verändert oder eingeschränkt werden (siehe [28]):

- **Path Traversal:** In diesem Fall folgt man den Links zu weiteren Informationsquellen bzw. -objekten. Im Fall von *VisiNav* kann man z. B. den Fokus der Suche auf alle Informationsquellen verschieben, die über ein bestimmtes Prädikat mit einer der

⁴Die Notation `skos:prefLabel` bezeichnet die URL <http://www.w3.org/2004/02/skos/core#prefLabel>. Dabei steht das Präfix `skos:` für den Namespace mit der URL <http://www.w3.org/2004/02/skos/core#>.

Informationsquellen aus dem aktuellen Suchfokus verbunden sind. In Sig.ma hingegen kann der Fokus immer nur zu einer bestimmten Ressource verschoben werden, zu der dann alle weiteren Informationen angezeigt werden. Dabei ist zu beachten, dass die momentanen Implementierungen von generellen Benutzerinterfaces wie z. B. Sig.ma oder VisiNav in der Regel nur explizit vorhandenen Relationen zwischen Ressourcen folgen können. In der Regel erfolgt keine Inferenz von implizit vorhandenen Relationen, wie z. B. inversen oder symmetrische Relationen, bei denen nur die eine Richtung explizit modelliert ist.

- **Facet Selection:** Hier wird der aktuelle Fokus der Suche nicht verschoben, sondern nur weiter eingeschränkt. Die Auswahl einer Facette entspricht der Einschränkung der Suchergebnisse auf die Subjekte, die zu einem bestimmten Prädikat einen bestimmten Wert als Objekt stehen haben.

Insgesamt bleibt festzuhalten, dass die momentan verfügbaren, generellen Benutzerinterfaces nicht oder nur sehr wenig Semantik der Daten auswerten und in der Regel keine Inferenz von implizit vorhandenen Relationen betreiben. Sie stellen nur generelle Navigationsmöglichkeiten für Graphen zur Verfügung, die benannte Kanten besitzen und deren Knoten Attributwerte zugewiesen werden können. Dies wird sich vermutlich auch in Zukunft nicht ändern, da eine viel zu große Heterogenität in den Daten vorliegt. Es ist höchstens zu erwarten, dass besonders populäre Schemata wie z. B. FOAF⁵, SIOC⁶ oder Dublin Core⁷ rudimentär durch solche generellen Benutzerinterfaces unterstützt werden.

Trotzdem können einige Anforderungen abgeleitet werden, die bei der Erzeugung von Linked Data beachtet werden sollten, um die passende Verarbeitung durch entsprechende generelle Benutzerinterfaces sicherzustellen:

- **Verwendung von etablierten Schemata:** Nach Möglichkeit sollten für die Darstellung der Daten etablierte Schemata verwendet werden. Als relevante Schemata für KOS wie z. B. Thesauri, Klassifikationen oder Personennormdateien können z. B. SKOS⁸ und FOAF identifiziert werden. Für die Katalogisierung und die Annotation von Titeldaten mit Schlagwörtern bietet sich z. B. Dublin Core an. Insbesondere SKOS wurde bereits in mehreren Fällen eingesetzt, um KOS im Semantic Web zugänglich zu machen (siehe [43, 33, 39, 35]).
- **Natürlichsprachige Beschreibungen:** Wo immer dies möglich und sinnvoll erscheint, sollte an ein Informationsobjekt ein Label annotiert werden. Dies stellt sicher, dass der natürlichsprachige Sucheinstieg möglich ist. Zudem werden diese Label benutzt, um die Daten für den Benutzer leserlich im Interface aufzubereiten. Nach Möglichkeit sollte das Prädikat, mit dem solche Label annotiert werden, vom Prädikat `rdfs:label` abgeleitet sein, damit generelle Benutzerinterfaces diese Information entsprechend verarbeiten können.
- **Auflösbare URLs:** Die URL einer Entität sollte immer auflösbar sein, da sich die generellen Benutzerinterfaces auf diese Weise weiterführende Informationen über diese Entität selbstständig erschließen können. Dies entspricht auch einem der Linked Data-Prinzipien aus [20].

⁵<http://www.foaf-project.org/>

⁶<http://sioc-project.org/>

⁷<http://dublincore.org/documents/dc-rdf/>

⁸<http://www.w3.org/2004/02/skos/>

2.3.2 Anwendungsspezifische Benutzerinterfaces

Beispiele für anwendungsspezifische Benutzerinterfaces sind *Museum Finland* [30], *SemaPlover* [37] oder die integrierte Ansicht von zwei verschiedenen Sammlungen von illuminierten Handschriften [24]. Diese Benutzerinterfaces haben gemeinsam, dass sie nur auf manuell selektierte und für den Anwendungsfall relevante Datenbestände angewendet werden. Dadurch ist eine Anpassung der Navigationsmöglichkeiten auf den konkreten Datenbestand, die darin verwendeten Schemata und die konkreten Informationsbedürfnisse der Benutzer in dem Anwendungsfall möglich.

Prinzipiell unterscheiden sich die Navigationstechniken in den anwendungsspezifischen Benutzerinterfaces nicht von denen in den generellen Benutzerinterfaces. Durch die Beschränkung auf einen bestimmten Anwendungsfall und bestimmte Datenbestände ist es aber möglich, vordefinierte Navigationspfade mit Hilfe der generellen Navigationstechniken zu definieren. Die generellen Benutzerinterfaces stellen im Prinzip eine interaktive Möglichkeit zur Verfügung, um beliebige Anfragen für beliebige Schemata zu formulieren (eine Art „interaktives SQL“). Die anwendungsspezifischen Benutzerinterfaces stellen hingegen bereits vordefinierte Anfragen mit entsprechenden Platzhaltern zur Verfügung, die nach Meinung von Domänenexperten dem Informationsbedürfnis der Benutzer am besten entsprechen.

Im Grundsatz haben anwendungsspezifische Benutzerinterfaces keinerlei Anforderungen bezüglich der Transformation von KOS. Es sind jedoch einige Dinge zu beachten, um den Aufwand für die Entwicklung von anwendungsspezifischen Benutzerinterfaces möglichst gering zu halten und um möglichst einfach weitere Datenbestände in die Anwendung zu integrieren:

- **Verwendung von etablierten Schemata:** Nach Möglichkeit sollten die gleichen Schemata verwendet werden wie in den bereits existierenden Datenbeständen, zu denen man beitragen will. Dadurch sinkt der Aufwand während der Integrationsphase von weiteren Datenbeständen in die Anwendung. Es bieten sich insbesondere etablierte Schemata wie SKOS, FOAF und Dublin Core an um KOS und die damit annotierten Titeldaten für das Semantic Web zugänglich zu machen.
- **Verwendung von wohldefinierten Schemata:** Die Semantik der verwendeten Schemata sollte wohldefiniert sein. Dadurch ist sichergestellt, dass jede Anwendung das Schema auf gleiche Weise interpretiert. Analog bieten sich hier wieder die bereits oben aufgeführten Schemata an.
- **Verlinkung mit und Mapping auf andere Datenbestände:** Die Integration von zwei Datenbeständen wird erleichtert, wenn entweder eine direkte Verbindung zwischen diesen Datenbeständen vorliegt oder wenn sie beide auf einen dritten, zentralen Datenbestand verlinken. Aus dieser Anforderung lässt sich ableiten, dass KOS idealerweise auf andere relevante KOS aus dem gleichen Fachbereich verlinken sollten. Auch die Verlinkung mit dem DBPedia-Datenbestand⁹ kann sinnvoll sein, da dieser sich zu einem zentralen Datenbestand in der Linked Data-Wolke entwickelt hat, auf den die meisten anderen Datenbestände ebenfalls wiederum verlinken (siehe [22]). Dies folgt aus einem der zentralen Prinzipien von Linked Data (siehe Abschnitt 2.1).

⁹<http://dbpedia.org/>

3 Repräsentation von Wissensorganisationssystemen im Semantic Web

In Abschnitt 1 wurde bereits erläutert, dass man bei Wissensorganisationssystemen (KOS) generell zwischen drei verschiedenen Modellierungsebenen unterscheiden kann. Zum einen existiert das KOS-Schema, das generell beschreibt, welche Art von Elementen und Beziehungen in einem KOS enthalten sein können. Zum anderen gibt es die Ebene mit den KOS-Daten, in der z. B. die konkreten Schlagwörter und Beziehungen zwischen Schlagwörtern abgelegt sind. Und als letztes gibt es die Annotierungen, durch die bestimmte Schlagwörter z. B. einem Buchtitel zugewiesen werden. Im Folgenden stellen wir Ansätze vor, wie diese Modellierungsebenen eines KOS im Rahmen des Semantic Webs dargestellt werden können.

3.1 Transformation des KOS-Schemas

Wie bereits in Abschnitt 1 erklärt, enthält das KOS-Schema die Information, nach welchem Schema die Daten in einem KOS abgelegt werden. In diesem Schema kann zum Beispiel definiert sein, dass ein Konzept genau eine Vorzugsbenennung und beliebig viele Synonyme hat, und dass Konzepte prinzipiell über eine hierarchische **broader**- oder eine assoziative **related**-Relation in Beziehung zueinander gesetzt werden können. Es hängt vom konkreten KOS ab, welche Elemente und Beziehungen zwischen den Elementen im Rahmen des KOS-Schemas definiert werden.

Das KOS-Schema wird bei der Transformation nach RDF/OWL auf der Schema-Ebene der erzeugten Ontologie abgelegt (siehe Abschnitt 2.2). Auch wenn jedes KOS im Prinzip ein einzigartiges KOS-Schema hat, so gibt es doch eine große Überschneidung zwischen verschiedenen KOS. Deswegen bietet es sich an für diesen gemeinsamen Kern verschiedener KOS-Schemata auf ein bereits vordefiniertes Schema zurückzugreifen. In den letzten Jahren hat sich dafür die SKOS-Spezifikation etabliert [12]. Die SKOS-Spezifikation ist insbesondere zur Darstellung von Sachschlagwörtern in Thesauri und deren Beziehungen untereinander geeignet, kann aber auch z. B. für die Darstellung von Klassifikationssystemen genutzt werden. Durch die Verwendung eines solchen etablierten Schemas erfüllt die Repräsentation eines KOS eine der wichtigen Anforderungen an Linked Data, die im Rahmen der Analyse in Abschnitt 2.3 identifiziert wurden.

Allerdings wird man bei der Verwendung von SKOS in der Regel auch an die Grenzen dieser Spezifikation stoßen und nicht alle Informationen im eigenen KOS auf dieses Schema abbilden können, da SKOS z. B. nicht für die genaue Repräsentation von Personennormdaten oder Geographika ausgelegt ist. Auch bei der Abbildung von Klassifikationssystemen wurde über größere Probleme berichtet (siehe die Diskussion in Abschnitt 4.2.2 und [35]). In einem solchen Fall bietet es sich an nach ergänzenden oder alternativen Spezifikationen zu suchen, in deren Rahmen die fehlenden Elemente definiert sind, bevor man selber Erweiterungen eines bestehenden Schemas wie SKOS vornimmt.

Eine Liste weiterer in Frage kommender Spezifikationen ist zum Beispiel der Beschreibung des Linked Data Services der Deutschen Nationalbibliothek zu entnehmen [17]. Eine weitere Möglichkeit bietet der Service *Linked Open Vocabularies*¹⁰. Auf dieser Webseite werden Schemata von anderen Datenbestände aus der Linked Data-Wolke aufgelistet, sowie deren Verbreitung. Die Schemata sind kategorisiert nach ihrem Anwendungsgebiet, beispielsweise der Darstellung von Informationen aus der Bibliotheks-Welt¹¹ oder der Dar-

¹⁰<http://labs.mondeca.com/dataset/lov/index.html>

¹¹http://labs.mondeca.com/dataset/lov/details/vocabularySpace_Library.html

stellung von geographischen Informationen¹². Als Beispiele für alternative oder ergänzende Spezifikationen für Personennormdaten und Geographika sind insbesondere die DCMI Metadata Terms¹³, die RDA Element Sets¹⁴ oder FOAF¹⁵ zu nennen.

3.2 Transformation der KOS-Daten

Während der Transformation der KOS-Daten wird der eigentliche Inhalt eines KOS nach RDF/OWL transformiert. Die KOS-Daten werden dabei auf der Daten-Ebene der erzeugten RDF/OWL-Ontologie abgelegt und sind entsprechend den Vorgaben des KOS-Schemas auf der Schema-Ebene der Ontologie organisiert. Demzufolge hängt dieser Schritt maßgeblich von den Entscheidungen ab, die man bei der Transformation des KOS-Schemas im vorherigen Schritt getroffen hat.

Bei dieser Art der Aufteilung eines KOS auf die Schema- und die Daten-Ebene einer RDF/OWL-Ontologie wird hauptsächlich die Semantik des KOS-Schemas explizit ausgedrückt. Diese Semantik kann dafür genutzt werden um aus den bestehenden, explizit gemachten Aussagen in einer Ontologie neue, implizit vorhandene Aussagen abzuleiten. Die Ableitung von neuen Aussagen basiert dabei auf der verwendeten Ontologie-Sprache und deren zugrundeliegender Logik. OWL2 basiert beispielsweise auf der Beschreibungslogik, während im Fall von OWL DL dies z. B. die Beschreibungslogik *SR**O**I**Q* wäre. Im Folgenden werden zwei Beispiele gegeben, wie implizit vorhandene Aussagen mit Hilfe eines Reasoners abgeleitet werden können:

In Abbildung 2 ist auf der Schema-Ebene definiert, dass die `skos:broader`-Beziehung immer zwischen zwei Instanzen der Klasse `skos:Concept` besteht. Falls wir nun die Information vorliegen haben, dass `swd:BRD` in einer `skos:broader`-Beziehung zu der Instanz `swd:Europa` steht, könnte die Information von der Schema-Ebene benutzt werden um zu schlussfolgern, dass `swd:Deutschland` und `swd:Europa` jeweils eine Instanz der Klasse `skos:Concept` sind. Es müsste somit nicht explizit angegeben werden, dass `swd:BRD` und `swd:Europa` als `rdf:type` die Klasse `skos:Concept` haben.

In einem anderen Beispiel könnte auf der Schema-Ebene in Abbildung 2 modelliert sein, dass die `skos:broader`-Beziehung die inverse Beziehung zu `skos:narrower` ist. Diese Information kann durch einen RDF/OWL-Reasoner benutzt werden um zu schlussfolgern, dass nicht nur die `skos:broader`-Beziehung zwischen `swd:BRD` und `swd:Europa` besteht, sondern auch eine `skos:narrower`-Beziehung zwischen `swd:Europa` und `swd:BRD`.

Zusammenfassend kann man sagen, dass ein Reasoner dazu genutzt werden kann aus den in einer Ontologie enthaltenen Aussagen neue Aussagen abzuleiten, welche auf den Regeln der verwendeten Ontologiesprache basieren.

3.3 Transformation der Annotierungen

Die eigentliche Transformation der Annotierungen ist nicht Gegenstand dieses Best Practice-Guides. Prinzipiell ist das Vorgehen aber ähnlich zu der Transformation des eigentlichen KOS. Zuerst muss auf der Schema-Ebene der RDF/OWL-Ontologie definiert werden, nach welchem Schema Annotationen vorgenommen werden. Auch hier sollte nach Möglichkeit ein bereits existierendes Schema wiederverwendet werden. So könnte man die Relation `dc:subject` aus dem Dublin Core Metadata Element Set¹⁶ verwenden um einen Buchtitel

¹²http://labs.mondeca.com/dataset/lov/details/vocabularySpace_Space.html

¹³<http://dublincore.org/documents/dcmi-terms/>

¹⁴<http://metadataregistry.org/rdabrowse.htm>

¹⁵<http://xmlns.com/foaf/spec/>

¹⁶<http://dublincore.org/documents/dces/>

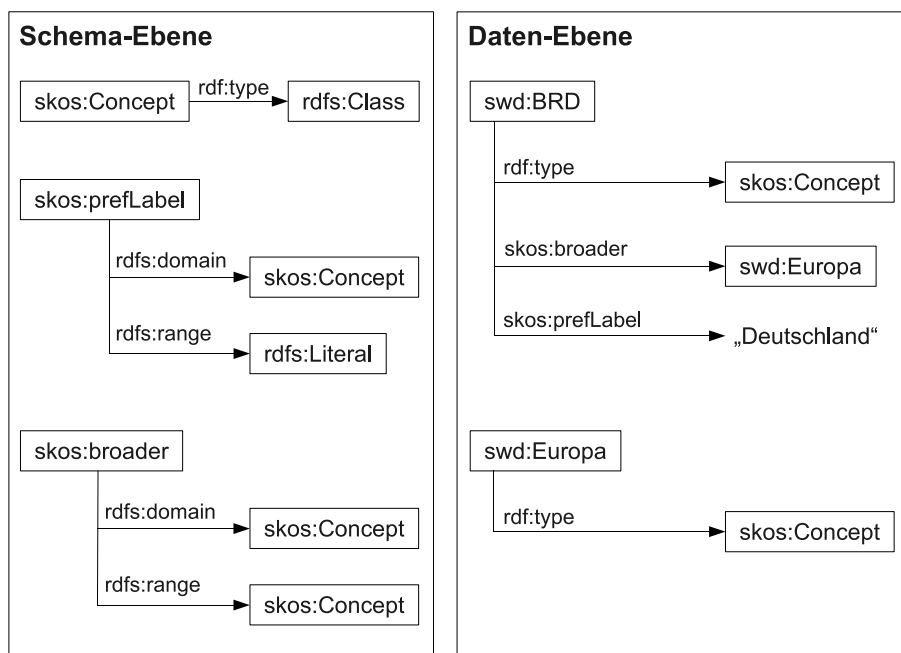


Abbildung 2: Beispiel für die Aufteilung eines Thesaurus auf die Schema- und die Daten-Ebene einer Semantic Web Repräsentation.

mit dem dazugehörigen Konzept aus dem Thesaurus in Beziehung zu setzen. Die konkreten Daten, welche Titel mit welchen Konzepten annotiert sind, befinden sich dann wieder auf der Daten-Ebene der erzeugten RDF/OWL-Ontologie.

3.4 Verwendung von Metamodellierung

Bei der oben beschriebenen Art der Transformation von KOS in eine Semantic Web Repräsentation wird hauptsächlich die Semantik des KOS-Schemas explizit ausgedrückt. Dies ermöglicht u.a. das automatische Schlussfolgern, wie es in Abschnitt 3.2 beschrieben ist. Die Art und Weise, wie jedoch die auf der Daten-Ebene modellierten KOS-Daten zu interpretieren sind, ist nicht explizit ausgedrückt. Was bedeutet es zum Beispiel im Rahmen des Retrievals, wenn ein bestimmter Buchtitel mit dem Konzept `swd:Wiesbaden` annotiert ist und wir wissen, dass `swd:Hessen` ein Oberkonzept von `swd:Wiesbaden` ist? Soll der Buchtitel zurückgeliefert werden, wenn ein Benutzer nach Titeln sucht, die unter dem Konzept `swd:Hessen` klassifiziert sind?

Diese Frage kann im Prinzip nur im Rahmen einer spezifischen Anwendung beantwortet werden, die die Semantik eines KOS-Schemas, wie z. B. der SKOS-Spezifikation, interpretiert. Im Rahmen einer solchen Anwendung *kann* es daher Sinn machen, die Konzepte in einem Thesaurus selbst wie Klassen zu benutzen und die `skos:broader`-Beziehung zwischen den Konzepten wie eine hierarchische Beziehung zwischen Klassen zu interpretieren (siehe auch [19]). In diesem Fall würden die Thesaurus-Konzepte nicht nur als Instanzen der Klasse `skos:Concept` modelliert, sondern würden selbst wieder als OWL-Klassen modelliert. Im Falle, dass zwischen zwei Konzepten die `skos:broader`-Beziehung besteht, würde diese zusätzlich noch durch eine `rdfs:subClassOf`-Beziehung modelliert.

Bei diesem Vorgehen würde die Interpretation der SKOS-Spezifikation bereits durch den Modellierer des KOS übernommen, indem nicht nur die Semantik des KOS-Schemas sondern auch die der KOS-Daten mit Hilfe von RDFS/OWL formalisiert wird. Eine SKOS-

Anwendung bräuchte dementsprechend die KOS-Daten nicht mehr selber zu interpretieren, sondern könnte dies einem RDFS/OWL-Reasoner überlassen. Das durch SKOS formalisierte KOS-Schema würde in diesem Fall zu einem Metamodell für das durch die KOS-Daten ausgedrückte Modell.

Auch wenn Metamodellierung auf den ersten Blick sehr praktisch erscheint, weil auf diese Weise existierende RDFS/OWL-Reasoner genutzt werden können um eine SKOS-Anwendung zu implementieren, sind doch mindestens zwei Probleme damit verbunden, die die Metamodellierung nur eingeschränkt für eine Best Practice geeignet erscheinen lassen:

1. Im obigen Beispiel wird ein Thesaurus-Konzept als Instanz der Klasse `skos:Concept` definiert, gleichzeitig ist das Konzept aber auch eine Klasse im Sinne von RDFS/OWL. Ein Thesaurus-Konzept befindet sich also sowohl auf der Daten-Ebene der erzeugten Ontologie als auch auf der Schema-Ebene. Dies führt dazu, dass die erzeugte Ontologie als OWL Full Ontologie klassifiziert wird und somit nicht mehr entscheidbar ist. Erst in OWL 2 wurde mit Punning eine Möglichkeit für Metamodellierung eingeführt, die unter bestimmten Voraussetzungen die Entscheidbarkeit der so erzeugten Ontologie erhält (siehe [6]). Dies wird dadurch ermöglicht, dass die Klasse und die Instanz logisch als zwei unterschiedliche Entitäten betrachtet werden, obwohl sie die gleiche URI besitzen.
2. Eine Abbildung der Konzepte und Beziehungen eines KOS auf z. B. `rdfs:subClassOf` und `rdf:Class` entspricht immer einer semantischen *Interpretation* des Thesaurus, die so nicht in jedem Fall durch den ursprünglichen Thesaurus bzw. seine Dokumentation abgedeckt ist. Es vermischen sich häufig bei der `skos:broader`-Beziehung aus ontologischer Sicht mehrere Beziehungstypen: (1) Sie kann eine Subklassen-Beziehung darstellen (`=rdfs:subClassOf`), (2) sie kann eine Klassen-Instanz-Beziehung darstellen (`=rdf:type`) oder (3) eine Teil-Ganzes-Beziehung (siehe auch [25]). Auch wenn sich alle 3 Beziehungstypen in einigen Anwendungen ähnlich verhalten und daher eine gemeinsame Abbildung auf z. B. `rdfs:subClassOf` gerechtfertigt zu sein scheint, können sich die unterschiedlichen Beziehungstypen in anderen Anwendungen sehr unterschiedlich verhalten. Eine einheitliche Abbildung auf die `rdfs:subClassOf`-Beziehung würde also eine große Ungenauigkeit herbeiführen. Eine genaue Unterscheidung der verschiedenen Beziehungstypen führt hingegen zu einem großen Arbeitsaufwand und wäre eine nicht zu unterschätzende Quelle für Fehler während der Transformation, da z. B. auch für menschliche Evaluatoren im Einzelnen nicht immer klar ersichtlich ist, welcher Beziehungstyp im Falle einer `skos:broader`-Beziehung vorliegt (siehe auch [25]). Trotz dieser Schwierigkeiten bei der Transformation der Semantik der KOS-Daten in RDFS oder OWL gibt es einige Beispiele, bei denen die damit verbundenen Anstrengungen unternommen wurden (siehe z. B. [29, 38]). Diese Anstrengungen lohnen sich aber in der Regel nur, wenn ein konkreter Anwendungsfall existiert, für den die explizite Repräsentation der Semantik der KOS-Daten benötigt wird.

4 Vorgehensmodelle für Linked Data-Projekte

In diesem Abschnitt stellen wir zwei verschiedene Vorgehensmodelle vor. Beide sind hilfreich bei der Durchführung eines Projekts, während dessen ein existierender Thesaurus in eine Linked Data-Repräsentation überführt werden soll.

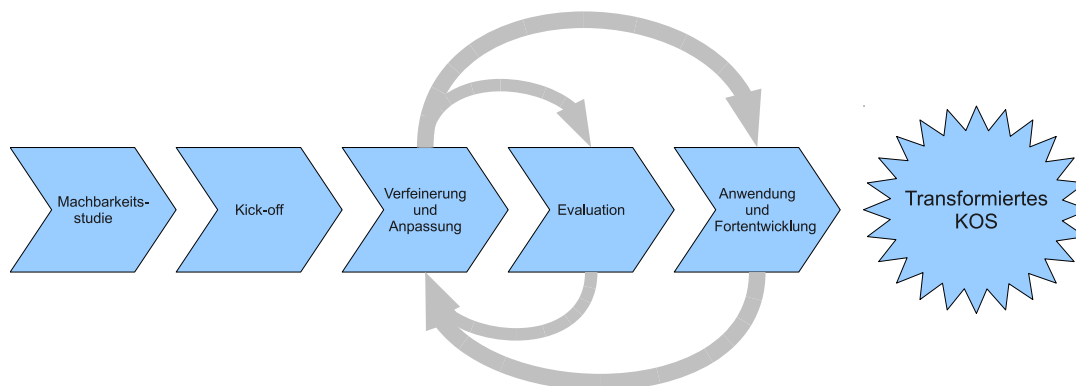


Abbildung 3: Allgemeines Phasenmodell zur Transformation von KOS in eine Linked Data-Repräsentation (nach [40]).

Das erste Vorgehensmodell betrifft den allgemeinen Projektablauf. Es enthält die Phasen, während derer unter anderem geklärt wird, warum der Thesaurus überhaupt nach Linked Data transformiert werden soll, wer die späteren Anwender der Linked Data-Repräsentation sein sollen und unter welcher Lizenz der Thesaurus zur Verfügung gestellt werden soll. Mehr Details zu diesem allgemeinen Vorgehensmodell sind in Abschnitt 4.1 zu finden.

Das zweite Vorgehensmodell befasst sich hingegen nur mit der Durchführung der konkreten Transformation in eine Linked Data-Repräsentation. Es besteht im Prinzip aus den drei Phasen (1) Analyse des existierenden KOS-Schemas, (2) Mapping auf ein Schema im Semantic Web wie z. B. SKOS und (3) Entwicklung eines Programms zur Transformation der KOS-Daten. Dieses Vorgehensmodell ist in Abschnitt 4.2 genauer beschrieben.

Die beiden Vorgehensmodelle sind nicht als Alternativen zueinander zu sehen, vielmehr beziehen sie sich auf unterschiedliche Phasen während der Durchführung des Projekts bzw. sie beschreiben die Prozesse auf unterschiedlichen Granularitätsebenen.

4.1 Allgemeines Phasenmodell für Wissensmanagement-Applikationen

Das allgemeine Phasenmodell zur Transformation von Wissensorganisationssystemen (KOS) in eine Linked Data-Repräsentation orientiert sich am in [40] beschriebenen iterativen, rückgekoppelten Phasenmodell für Wissensmanagement-Applikationen. Es besteht aus 5 aufeinanderfolgenden Phasen. Im Folgenden werden diese 5 Phasen genauer erläutert und auf den Kontext der Transformation von KOS nach Linked Data bezogen. Ein Überblick über die 5 Phasen ist in Abbildung 3 dargestellt.

4.1.1 Machbarkeitsstudie

Am Anfang eines Linked Data-Projekts steht die Analyse der Gegebenheiten und Anforderungen. Im Folgenden listen wir zwei beispielhafte Aktivitäten auf, die u.a. während dieser Phase durchgeführt werden sollten:

Als erstes sollte überlegt werden, für welchen Anwendungsbereich das KOS als Linked Data zur Verfügung gestellt werden soll und wer die anvisierte Zielgruppe ist. Dies schließt mit ein sich Gedanken zur beabsichtigten Nutzung von entsprechenden Benutzeroberflächen zu machen (siehe Abschnitt 2.3). Daraus folgt weiterhin, wie in [17] beschrieben, eine Auswahl der Datenbestände, mit denen das eigene KOS angereichert und verknüpft werden kann.

In dieser Phase sollte man sich weiterhin auch Gedanken darüber machen, unter welcher Art von Lizenz die Daten des KOS zur Verfügung gestellt werden. Dies schließt mit ein zu klären, welche Rechte man selber an welchen Teilen des KOS hat (siehe [27]). Unter Umständen ist es auch nicht möglich das komplette KOS zur Verfügung zu stellen, sondern nur die Teile an denen man selbst die dafür benötigten Rechte besitzt bzw. bei denen nicht der Schutz persönlicher Informationen einer Veröffentlichung entgegensteht (teilweise bei Personennormdaten gegeben).

Für die Veröffentlichung der Daten gibt es diverse Lizenzmodelle. Beispiele für solche Lizenzmodelle sind die verschiedenen Arten der Creative Commons Lizenz¹⁷ und die Open Database License¹⁸ im Zusammenspiel mit der Database Contents License¹⁹. Weitere Informationen und Diskussionen zu diesem Thema werden in Zukunft auch im Kontext der Lizenzen-Gruppe der DINI AG KIM gesammelt und geführt.²⁰

4.1.2 Kickoff

Nachdem in der vorherigen Phase die Anforderungen und Gegebenheiten analysiert wurden, empfiehlt es sich einen ersten Entwurf der Linked Data-Repräsentation zu erstellen. In dieser Phase können dementsprechend einmal die in Abschnitt 4.2 vorgeschlagenen Schritte zur Analyse des existierenden KOS angewendet werden, sowie dem anschließenden Mapping auf z. B. das SKOS-Schema und der Durchführung der Transformation der KOS-Daten.

Diese erste transformierte Version des KOS kann danach in der folgenden Phase genutzt werden um mit den für das KOS zuständigen Fachabteilungen ins Gespräch zu kommen oder Rückmeldungen von den zukünftigen Anwendern der Linked Data-Repräsentation zu bekommen. Die Kickoff-Phase soll mögliche Problemfelder bei der Transformation aufzeigen, die dann in den folgenden Phasen behoben werden.

4.1.3 Verfeinerung und Anpassung

In dieser Phase soll das bereits transformierte KOS weiter verfeinert und angepasst werden. Es sollen die identifizierten Probleme behoben und das KOS für die nachfolgende Evaluation vorbereitet werden. Als typisches Problem kann beispielsweise auftreten, dass nicht alle Elemente des ursprünglichen KOS auf entsprechende Elemente in z. B. SKOS gemappt werden können bzw. dass dieses Mapping nicht eindeutig ist. In Rücksprache mit den Fachabteilungen und den Anwendern müssen diese Probleme nachfolgend in dieser Phase behoben werden.

Ein möglicher Lösungsweg um die oben angesprochenen Probleme zu beheben, könnte zum Beispiel die Suche nach weiteren relevanten Schemata sein, die die benötigten Elemente zur Verfügung stellen. In Abschnitt 3.1 sind einige Schemata aufgelistet, die neben SKOS noch für die Transformation von KOS relevant sind. Sollte diese Lösung nicht ausreichend sein, kann auch darüber nachgedacht werden eigene Elemente in das KOS-Schema einzuführen. Nach Möglichkeit sollten diese aber von bestehenden Elementen in etablierten Schemata abgeleitet werden. Wenn man eigene Elemente einführt bzw. von bestehenden Elementen ableitet, sollte auch unbedingt auf eine Veröffentlichung des entsprechenden Schemas geachtet werden. Denn erst dadurch wird die Semantik der neuen

¹⁷<http://creativecommons.org/>

¹⁸<http://www.opendatacommons.org/licenses/odbl/>

¹⁹<http://www.opendatacommons.org/licenses/dbcl/>

²⁰<https://wiki.d-nb.de/display/DINIAGKIM/Lizenzen+Gruppe>

Elemente zugänglich und die Interoperabilität mit anderen Datenbeständen in der Linked Data-Wolke ermöglicht. Diese Interoperabilität mit anderen Datenbeständen ist eine wichtige Anforderung an Linked Data-Repräsentationen (siehe Abschnitt 2.3).

4.1.4 Evaluation

In dieser Phase wird der bisherige Stand der Transformation bewertet. Dies geschieht sowohl durch die für das KOS zuständigen Fachabteilungen als auch durch die zukünftigen Anwender, die als Zielgruppe während der Machbarkeitsstudie identifiziert wurden. Neben dieser Evaluation durch Anwender und Fachabteilungen kann jedoch auch noch eine automatische Überprüfung und Validierung des transformierten KOS vorgenommen werden. Ein Beispiel wäre die Korrektheit des erzeugten RDF zu testen oder auch weitere Integritätsbedingungen zu überprüfen, wie sie auch durch die SKOS-Spezifikation vorgegeben werden.

Die Überprüfung des erzeugten RDF kann mithilfe des RDF-Validators des W3C durchgeführt werden.²¹ Eine Auflistung weiterer zu überprüfender Kriterien für Linked Data im Allgemeinen ist z. B. [7] zu entnehmen. Qualitätskriterien speziell aufgelistet für SKOS-Datenbestände sind [8] zu entnehmen. [14] führt genauer auf, inwiefern sich die verschiedenen Datenbestände in der Linked Data-Cloud an die Best Practices zur Bereitstellung von Linked Data halten.

In [26] wird darauf hingewiesen, dass ein besonderes Augenmerk auch auf die korrekte Integration/Verknüpfung mit anderen Datenbeständen gelegt werden sollte. Für KOS bezieht sich das beispielsweise auf Verknüpfungen zu anderen KOS im Linked Data-Format oder zu z. B. DBPedia. Nicht vergessen werden sollte in dieser Phase auch die technische Infrastruktur zu überprüfen, mit deren Hilfe die Linked Data-Repräsentation bereitgestellt wird, ob sie u.a. die in sie gesetzten Erwartungen erfüllt.

Bei Fehlern oder auftretenden Problemen wird in einem iterativen Schritt zurück zur Verfeinerungs-Phase gegangen um das Problem zu beheben. Wenn keine weiteren Probleme auftreten, kann mit der Phase der Anwendung und Weiterentwicklung des Linked Data-Services fortgesetzt werden.

4.1.5 Anwendung und Weiterentwicklung

Während der Anwendung und Weiterentwicklung wird die Linked Data-Repräsentation eines KOS endgültig öffentlich gemacht. Dies sollte in der Regel auch eine Registrierung des Datenbestandes auf CKAN²² mit einschließen um die Veröffentlichung des Datenbestandes bekanntzugeben. Auch danach kann es zu einer kontinuierlichen Anpassung und Erweiterung der KOS-Daten sowie deren Transformation in eine Linked Data-Repräsentation kommen, welches einem iterativen Zurückgehen in die vorherigen Phasen entspricht.

Spätestens nach der Veröffentlichung des KOS als Linked Data sollte hingegen darauf geachtet werden, dass die verwendeten URIs zur Bezeichnung von Entitäten möglichst änderungsarm bzw. faktisch änderungsfrei sind. Durch eine nachträgliche Änderung von URIs würden andernfalls Verknüpfungen von anderen Datenbeständen auf das veröffentlichte KOS ungültig (siehe auch [21, 36]).

Wie bereits mehrfach betont wurde, ist dieses Phasenmodell iterativ zu durchlaufen um eine kontinuierliche Verbesserung unter Einbeziehung der Anwender zu erreichen. Dies umfasst alle angesprochenen Fragestellungen wie verwendete Lizenzen, Datenmodellierung,

²¹<http://www.w3.org/RDF/Validator/>

²²<http://ckan.net/>

Integration und Erweiterung mit neuen Datenbeständen oder auch technische Infrastruktur etc.

4.2 Konkrete Schritte zur Transformation eines KOS

Das im vorherigen Abschnitt beschriebene Vorgehensmodell bezieht sich auf den allgemeinen Ablauf eines Linked Data-Projekts. In diesem Abschnitt hingegen wollen wir konkrete Schritte vorstellen, wie das KOS von seiner bisherigen Repräsentation in eine Linked Data-Repräsentation transformiert werden kann. Laut dem Bericht von van Assem et al. in [18] haben sich dabei drei aufeinanderfolgende Schritte als nützlich erwiesen: (1) Analyse des existierenden KOS-Schemas, (2) Mapping auf ein Zielschema im Semantic Web wie z. B. SKOS und (3) Entwicklung eines Programms zur Transformation der KOS-Daten. Diese Vorgehensweise hat bereits breite Anwendung in diversen anderen Linked Data-Projekten gefunden (siehe [32, 33, 39, 43]).

Im Folgenden werden diese drei Schritte noch einmal genauer betrachtet. Konkrete Beispiele, wie diese Schritte umgesetzt werden können, sind in [18] zu finden. In Abschnitt 6 und 7 wenden wir dieses Vorgehen dann auf die Transformation der Schlagwortnormdatei und des Thesaurus Technik und Management an.

4.2.1 Analyse des ursprünglichen KOS-Schemas

Im ersten Schritt wird das existierende KOS-Schema analysiert. Neben dem KOS selbst wird dafür auch die verfügbare Dokumentation herangezogen, Experten für das KOS befragt und die auf dem KOS aufbauenden Anwendungen analysiert. Ergebnis dieses Schrittes ist eine Liste der im KOS verwendeten Elemente und Beziehungen. Diese Liste sollte auch jeweils eine Beschreibung beinhalten, wie die Elemente und Beziehungen verwendet werden bzw. welche Bedeutung sie haben. Im Prinzip wird also in diesem Schritt das existierende KOS-Schema und dessen Semantik erfasst.

Dieser Schritt wird erleichtert, wenn sich das bestehende KOS-Schema an einem existierenden Standard orientiert, wie z. B. den ISO-Normen 2788, 5964 oder 25964 [2, 3, 4]. In so einem Fall sind die Elemente und Beziehungen des Wissensorganisationssystems bereits in dem jeweiligen Standard aufgelistet und beschrieben. Es ist jedoch darauf zu achten, inwiefern das vorliegende Wissensorganisationssystem von dem verwendeten Standard abweicht.

4.2.2 Mapping auf das Zielschema im Semantic Web

Die im vorherigen Schritt gesammelten Informationen über das ursprüngliche KOS-Schema werden anschließend genutzt um es auf ein entsprechendes Zielschema im Semantic Web abzubilden. Wie bereits in Abschnitt 2.3 analysiert wurde, ist dabei ein wichtiges Ziel die Wiederverwendung von bereits existierenden KOS-Schemata. Durch die Wiederverwendung wird die Interoperabilität mit anderen KOS in der Linked Data-Wolke sichergestellt.

Bei vielen KOS wird sich dabei ein Mapping auf das SKOS-Schema anbieten. Es ist allerdings insbesondere bei Thesauri, die sich an existierende Thesaurus-Standards wie die ISO-Normen 2788, 5964 oder 25964 [2, 3, 4] halten, zu erwarten, dass große Teile des Thesaurus auf SKOS gemappt werden können.²³ Bei Klassifikationssystemen jedoch wurde teilweise schon über größere Probleme beim Mapping auf das SKOS-Schema berichtet. In

²³Der Anhang des SKOS Primer [31] enthält explizit eine Auflistung, wie die in ISO 2788 und 5964 aufgelisteten Elemente und Beziehungen auf SKOS abgebildet werden können und an welcher Stelle SKOS von diesen beiden Standards abweicht.

[35] werden sechs Probleme aufgelistet, die beim Mapping von Klassifikationen auf SKOS auftreten können. Zwei dieser Probleme beziehen sich auf die Beziehung zwischen Klassen und Themen, die in SKOS nicht richtig modelliert werden können. Das kann zu einer inkonsistenten Verwendung der Beziehungstypen in SKOS führen. Außerdem existiert keine Beziehung in SKOS, mit der auf eine bevorzugt zu verwendende Klasse hingewiesen werden kann, falls ein Thema interdisziplinär behandelt wird und somit zwei unterschiedlichen Klassen in einer hierarchisch organisierten Klassifikationen zugeordnet werden kann (für mehr Details siehe [35]). Ein weiteres Problem von Klassifikationen kann die nicht eindeutige Benennung von Klassen sein (siehe auch [5]). Erst durch die Einordnung einer Klasse in die Klassenhierarchie werden manche Benennungen eindeutig. Dadurch stellt sich die Frage, wie die Klassenbenennungen auf `skos:prefLabel` abgebildet werden, da letztere innerhalb des selben `skos:ConceptScheme` eindeutig sein sollen.

Bei größeren Unterschieden zwischen dem eigenen KOS-Schema und dem SKOS-Schema bietet es sich zumeist an, auf zusätzliche Elemente aus anderen Schemata zurückzugreifen oder komplett andere Schemata zu verwenden (siehe auch die Diskussion in Abschnitt 5.2). Eine Liste in Frage kommender Schemata ist Abschnitt 3.1 zu entnehmen.

Ein weiterer Punkt bei diesem Schritt liegt darin existierende Schemata wie SKOS an die eigenen Bedürfnisse anzupassen. Davon wird u. a. in [35] Gebrauch gemacht. Um die Interoperabilität mit anderen Datenbeständen zu erhalten, müssen die eigenen Erweiterungen auf bestehende Elemente oder Beziehungen des bereits existierenden Schemas zurückgeführt werden. Im Fall von SKOS bieten sich dafür zwei verschiedene Entwurfsmuster an, die aber auch auf andere existierende Schemata übertragen werden können:

1. Zum einen kann man die eigenen Elemente oder Beziehungen von bestehenden Elementen oder Beziehungen mit Hilfe von `rdfs:subClassOf` oder `rdfs:subPropertyOf` ableiten. Dieses Muster bietet sich z. B. an, wenn man eine bestehende Beziehung wie `skos:broadMatch` weiter differenzieren will. In [23] wurde dieses Muster angewendet um die CrissCross-Mappingrelationen auf die bestehenden SKOS-Mappingrelationen zurückzuführen. Eine Anwendung, die mit den CrissCross-Erweiterungen umzugehen weiß, kann dadurch die genaue Semantik der Mappingrelationen verwenden, während allgemeine SKOS-Anwendungen die unspezifischere SKOS-Mappingrelation verwenden.
2. Zum anderen kann man Regeln in der verwendeten Ontologiesprache (z. B. OWL DL) angeben, mit deren Hilfe von eigenen Elementen und Beziehungen auf bestehende Elemente und Beziehungen geschlussfolgert werden kann. Ein Beispiel hierfür sind die Property Chains in OWL (siehe [6]). Diese werden beispielsweise bei der SKOS-Erweiterung SKOS-XL [11] verwendet um von der `skosxl:prefLabel` und der `skosxl:literalForm`-Beziehung auf die bestehende `skos:prefLabel`-Beziehung schlussfolgern zu können (für Details siehe [11]).

Solch eine eigene Erweiterung von existierenden Schemata sollte aber immer vorsichtig vorgenommen werden, da dadurch potentiell die Interoperabilität mit anderen Datenbeständen gefährdet wird. Eine zwingende Voraussetzung, um die eigenen Erweiterungen anderen Benutzern zugänglich zu machen, ist immer die maschinenlesbare Veröffentlichung des erweiterten Schemas und die für den Menschen bestimmte Dokumentation. Wenn man eine der beiden obigen Entwurfsmuster für eigene Erweiterungen verwendet, sollte man sich bewusst sein, dass deren Interpretation die Verwendung eines Reasoners voraussetzt. Dies ist nicht immer der Fall, wie bereits in 2.3 angemerkt wurde.

4.2.3 Transformation der KOS-Daten

In diesem Schritt wird ein konkretes Programm entwickelt, mit dem die KOS-Daten in eine Linked Data-Repräsentation überführt werden, nachdem im vorherigen Schritt das KOS-Schema des existierenden KOS auf ein entsprechendes KOS-Schema im Semantic Web abgebildet wurde. Die Komplexität dieses Schritts hängt stark davon ab, wie groß die Übereinstimmungen zwischen den beiden gemappten KOS-Schemata sind und in welchem Format das KOS ursprünglich vorliegt. Prinzipiell kann in diesem Schritt jede beliebige Programmiersprache und jedes beliebige Werkzeug zum Einsatz kommen.

Um aber insbesondere in der Kick-Off Phase eines Projekts schnell Ergebnisse produzieren zu können, auf deren Basis man dann z. B. die Diskussion mit den Fachabteilungen und Anwendern führen kann, bietet sich die Entwicklung eines Prototyps an. Bei einem Prototyp muss im Gegensatz zum Produktivsystem u.a. nicht auf besondere Performanz geachtet werden und es muss auch nicht auf die Live-Daten des zu transformierenden KOS zugegriffen werden.

Für die Erstellung eines solchen Prototyps hat sich in der Vergangenheit die Verwendung von Transformationssprachen wie D2R Map²⁴ bewährt. Um eine solche Transformationssprache prototypisch verwenden zu können, müssen die KOS-Daten aber in der Regel zuerst in eine relationale Datenbank importiert werden. Aus diesem Grund ist dieser Ansatz für Produktivsysteme mit Zugriff auf die Live-Daten eines KOS in der Regel nicht geeignet, wenn die ursprünglichen Daten nicht bereits in einer relationalen Datenbank vorgehalten werden. Eine Liste weiterer Werkzeuge zur Transformation des Inhalts von relationalen Datenbanken in eine RDF Repräsentation ist der Webseite <http://www.w3.org/wiki/RdfAndSql> zu entnehmen. Es soll hier betont werden, dass dies nur eine mögliche Vorgehensweise unter vielen anderen ist und dass bei anderen Projekten sich eine andere Vorgehensweise möglicherweise besser eignet. So wurde z. B. in [18] in zwei Fällen SWI-Prolog für die Transformation benutzt und in einem anderen Fall Perl.

5 Umsetzung der Linked Data-Prinzipien

Der folgende Abschnitt geht näher auf einzelne Punkte bezüglich der Bereitstellung der technischen Infrastruktur ein, sowie die dabei zu beachtenden Entwurfsmuster um eine möglichst reibungslose Einbettung des transformierten KOS in die Linked Data-Wolke bzw. das Semantic Web zu gewährleisten. Wie bereits in Abschnitt 2.1 genauer erläutert wurde, sollen hierbei die folgenden 4 Prinzipien bei der Veröffentlichung von Daten in der Linked Data-Wolke bzw. dem Semantic Web beachtet werden:

1. Für die Bezeichnung der einzelnen Entitäten (Personen, Gegenstände, Orte, ...) in den Daten sollen URIs verwendet werden.
2. Die URIs sollen auf dem HTTP-Protokoll basieren, um sie so auf einfache Art dereferenzierbar zu machen.
3. Wenn eine URI dereferenziert wird, sollen weiterführende Informationen in standardisierten Datenformaten wie z. B. RDF, RDFS oder OWL vorliegen und nach Möglichkeit durch Anfragesprachen wie z. B. SPARQL abrufbar sein.
4. Die Informationen über eine Entität sollen nach Möglichkeit Links zu anderen Entitäten enthalten. Diese anderen Entitäten müssen nicht notwendigerweise in der

²⁴<http://www4.wiwiss.fu-berlin.de/bizer/d2rmap/D2Rmap.htm>

gleichen Datenquelle beschrieben sein. Es soll auch zu externen Datenquellen verlinkt werden.

5.1 Entwurfsmuster für den URI-Namespaces

Ein zentraler Aspekt bei der Veröffentlichung von Linked Data ist die richtige Wahl von URIs. Dabei ist besonders wichtig, dass die URIs änderungsarm bzw. änderungsfrei sind um so bestehende Verlinkungen zwischen Datenbeständen zu erhalten (siehe Abschnitt 4.1). Ein weiterer wichtiger Gesichtspunkt beinhaltet weiterhin zu unterscheiden, nach welchem Schema einzelne Entitäten, Relationen und Typen mit Hilfe einer URI identifiziert werden. Generell kann hier unterschieden werden zwischen den sogenannten *Hash-Namespaces* und den *Slash-Namespaces* für URIs. Als Beispiel wollen wir das SWD-Schlagwort „Violine“ nehmen, das die interne ID 4019791-8 hat:

Als eine Best Practice hat sich in diesem Zusammenhang erwiesen die Domäne der URI aus dem Namespace des Herausgebers abzuleiten. Im Beispiel eines SWD-Schlagworts ist dies z. B. <http://d-nb.info/>. Anschließend wird noch ein Pfad angehängt, der den Datenbestand identifiziert. In unserem Fall werden die Informationen der Schlagwortnormdatei (SWD) und u. a. der Personennormdatei (PND) innerhalb der Gemeinsamen Normdatei (GND) zusammengefasst. Als letzter Schritt wird an die vorherige Domäne noch der Pfad zur Gemeinsamen Normdatei angehängt: <http://d-nb.info/gnd/>.

Als nächstes stellt sich noch die Frage, wie innerhalb dieser Gemeinsamen Normdatei die einzelnen Entitäten referenziert werden sollen. Hier bieten sich zwei unterschiedliche Entwurfsmuster an. Zum einen kann das Muster der Hash-Namespaces verwendet werden, bei dem die ID einer Entität mit Hilfe eines Hashs an die Basis-URI des Datenbestandes angehängt wird: <http://d-nb.info/gnd#4019791-8>. Zum anderen kann das Muster der Slash-Namespaces verwendet werden, bei dem die Basis-URI des Datenbestandes und die ID der Entität durch einen Slash getrennt werden: <http://d-nb.info/gnd/4019791-8>.

Auch wenn sich die beiden URIs auf den ersten Blick nicht groß voneinander unterscheiden, so hat diese Entscheidung doch größere Auswirkungen auf die Art und Weise, wie ein Client auf die Informationen zu einer Entität zurückgreift:

- Im Fall der **Hash-Namespaces** liegen die Informationen zu allen Schlagwörtern in einem gemeinsamen Dokument vor (z. B. <http://d-nb.info/gnd>) und der Teil hinter dem Hash wird genutzt um zu einer bestimmten Position in diesem großen Dokument zu springen. Falls der Client anschließend noch Informationen zu einem weiteren Schlagwort aus der Gemeinsamen Normdatei zugreifen will, muss er nicht eine weitere Anfrage an den Server schicken sondern nur an eine andere Stelle im Gesamtdokument springen. Insgesamt kann so die Anzahl der Anfragen beim Webserver und die damit verbundene Verzögerung reduziert werden. Allerdings hat dies nicht nur Vorteile, denn hierdurch werden potentiell sehr große Dokumente zurückgeliefert, die immer den kompletten Datenbestand enthalten. Daher empfiehlt sich dieses Entwurfsmuster im Prinzip nur bei kleinen, überschaubaren Datenbeständen, die voraussichtlich auch in Zukunft nicht signifikant wachsen werden.
- Im Fall der **Slash-Namespaces** muss für jede Entität eine separate Anfrage beim Webserver gestellt werden um das dazugehörige Dokument zu erhalten. Dieser kann somit für die jeweilige URI zu einem entsprechenden Dokument weiterleiten, z. B. mit Hilfe der HTTP-Umleitung 303 **See Other** (siehe auch Abschnitt 5.2). Insgesamt wird dadurch die Anzahl der Anfragen beim Webserver im Vergleich zu der Verwendung von Hash-Namespaces erhöht. Die Verzögerung bei der Auflösung von

mehreren URIs ist hier nicht zu vernachlässigen. Gleichzeitig kann durch die Verwendung von Slash-Namespaces aber auch die übertragene Menge an Daten reduziert werden, da der Client gezielt die von ihm benötigten Informationen anfordern kann und nicht immer den kompletten Datenbestand herunterladen muss. Dieses Entwurfsmuster bietet sich daher insbesondere bei besonders großen Datenbeständen an.

Speziell bei der Definition von eigenen Elementtypen und Relationen auf der Schema-Ebene der Ontologie bietet sich die Verwendung eines Hash-Namespaces jedoch an. So haben z. B. auch die URIs der Elemente des SKOS-Vokabulars einen Hash-Namespace. Für die URIs auf der Daten-Ebene einer Ontologie bietet sich hingegen eher die Verwendung des Slash-Namespaces an. Weiterführende Informationen zu Hash- und Slash-Namespaces und dazu passenden Konfigurationen eines Webservers wie z. B. Apache kann [15] entnommen werden.

5.2 Bereitstellung unterschiedlicher Repräsentationen

Als Best Practice bei der Bereitstellung von Linked Data gilt nicht nur die maschinenlesbare Repräsentation der Daten anzubieten, sondern gleichzeitig auch parallel eine durch Menschen lesbare Repräsentation. Beide Repräsentation sollten mit Hilfe der URI einer Entität erreichbar sein. Als Entwurfsmuster zur Realisierung dieser Best Practice hat sich dabei Content Negotiation etabliert. Dabei gibt der Client die bevorzugte Repräsentationsform innerhalb des `Accept`-Datenfelds des HTTP-Headers während seiner Anfrage beim Server an.

Der Server unterscheidet dann mit Hilfe einer sogenannten „Weiche“, ob der Client zur maschinenlesbaren Repräsentation des Datenbestandes weitergeleitet wird oder zu der durch Menschen lesbaren Repräsentation. Um beispielsweise die maschinenlesbare Repräsentation zurückgeliefert zu bekommen würde ein Linked Data-Client im `Accept`-Feld des HTTP-Headers als bevorzugten Datentyp `application/rdf+xml` angeben. Webbrowser hingegen senden im `Accept`-Header standardmäßig `text/html` beziehungsweise `application/xhtml+xml` als bevorzugte Datentypen.

Die Weiche des Web-Servers leitet entsprechend des Eintrags im `Accept`-Feld den Client mittels der HTTP-Umleitung `303 See Other` dann von der die Entität bezeichnenden URI weiter zu einer spezifischen Repräsentation der zu dieser Entität vorliegenden Informationen. Zum Beispiel könnte für den Thesaurus-Begriff mit der URI <http://d-nb.info/gnd/4019791-8/> die HTML-Repräsentation unter <http://d-nb.info/gnd/4019791-8/about/html> abgelegt sein und die RDF-Repräsentation unter <http://d-nb.info/gnd/4019791-8/about/rdf>.

Detailliertere Informationen zur Content Negotiation sind zu finden in [36]. Beispielhafte Konfigurationen zur Realisierung der Content Negotiation im Rahmen eines Webservers wie z. B. Apache sind zu finden unter [15].

Bei der Bereitstellung von maschinenlesbaren Repräsentationen kann es sinnvoll sein, die Daten mit unterschiedlichen Schemata anzubieten. Zum Beispiel hat sich im allgemeinen Semantic Web die FOAF-Spezifikation zur Darstellung von personenbezogenen Daten etabliert. Diese wird aber nicht allen Anforderungen zur Erfassung solcher Daten im Bibliothekswesen gerecht. Die Konzepte für die Namensvergabe sind oft suboptimal, da es z. B. nicht erlaubt ist zwischen einer bevorzugten und einer alternativen Namensansetzung zu unterscheiden (siehe die Diskussion in [17]).

Die Unterscheidung zwischen bevorzugten und alternativen Namensansetzungen wird

```

<rdf:Description rdf:about="http://viaf.org/viaf/77390479">
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  <rdf:type rdf:resource="http://RDVocab.info/uri/schema/FRBRentitiesRDA/Person"/>
  <foaf:name>Tillet, Barbara B.</foaf:name>
  <foaf:name>Tillet, Barbara B. 1946-</foaf:name>
  <foaf:name>Tillet, Barbara Barnett 1946-...</foaf:name>
  <foaf:name>Tillet, Barbara Ann Barnett 1946-</foaf:name>
  <foaf:name>Tillet, Barbara 1946-</foaf:name>
  <rdaGr2:dateOfBirth>1946</rdaGr2:dateOfBirth>
  <owl:sameAs rdf:resource="http://d-nb.info/gnd/129329347"/>
  <owl:sameAs rdf:resource="http://libris.kb.se/resource/auth/233803"/>
</rdf:Description>

```

Abbildung 4: Ausschnitt der RDF/XML-Repräsentation für den VIAF-Datensatz für *Barbara Tillet*: <http://viaf.org/viaf/77390479/rdf.xml>

aber durch die *RDA Group 2 Elements*²⁵ unterstützt. Dieses Schema ist jedoch nicht so verbreitet im Semantic Web wie FOAF und wird aus diesem Grund durch viele potentielle Benutzer nicht unterstützt. Diese Problematik in der Wahl von einem genauem, unbekannteren Schema oder einem ungenauen, weiter verbreiteten Schema kann u. a. dadurch gelöst werden, dass man die Daten in beiden Formaten anbietet.

Ein Beispiel, bei dem konkurrierende Versionen des gleichen Datensatzes angeboten werden, ist der VIAF-Datenbestand. Dort hat man sich dafür entschieden für beide Versionen eines Datensatzes die gleiche URI zu verwenden [1]. Beispielsweise bezeichnet die URI <http://viaf.org/viaf/77390479/> den VIAF-Datensatz über Barbara Tillet. Die durch diese URI bezeichnete Ressource hat sowohl den Typ `foaf:Person` als auch `rdaEnt:Person` (siehe Abbildung 4). Diese URI wird auch als Subjekt für Aussagen verwendet, die mit Hilfe von FOAF und der RDA Group 2 Elements über diese Person gemacht werden. So kann der Name einer Person sowohl als `foaf:name` als auch als `rdaGr2:preferredNameForThePerson` oder `rdaGr2:variantNameForThePerson` vorliegen.

In so einem Fall ist es dem Benutzer des Datensatzes überlassen, welches der beiden verwendeten Schemata er bevorzugt verarbeitet. Genauso ist es möglich, dass ein Benutzer die mithilfe des RDA-Schemas bereitgestellten Informationen ignoriert, wenn er z. B. nur FOAF verarbeiten kann. Durch die Bereitstellung alternativer Repräsentationen wird man also beiden Seiten gerecht. Allerdings hat dieser Ansatz den Nachteil, dass er komplexere Transformationen benötigt und zu einem höheren Aufwand führt.

Will man diesen erhöhten Aufwand vermeiden, existiert noch die Möglichkeit, dass man die Information jeweils in dem Vokabular abspeichert, welches die ursprüngliche Semantik am besten erhält, sowie Mappings zu anderen Schemata anbietet. Dies betrifft das Forschungsgebiet des Ontology Mappings (siehe [34] für einen Überblick). Generell gibt es zwei verschiedene Ansätze, wie solche Mappings repräsentiert werden können:

1. Eine Möglichkeit besteht darin die Klassen und Properties aus zwei verschiedenen Schemata mit Hilfe von OWL-Konstrukten wie `owl:equivalentClass` oder `rdfs:subClassOf` aufeinander zu matchen. Dies hat den Vorteil, dass die Mappings durch einen OWL-Reasoner interpretiert werden können und somit prinzipiell durch jeden Benutzer eines Linked Data-Datenbestands, der einen Reasoner verwendet. Allerdings hat dieser Ansatz den Nachteil, dass Konstrukte wie `owl:equivalentClass` oder `rdfs:subClassOf` bestimmte logische Schlussfolgerungen ermöglichen, die in der Regel bei Mappings zwischen Schemata oder auch Instanzdaten nicht erwünscht sind (siehe Beispiel 74 in [12]). Daher ist dieser Ansatz nur bedingt für Mappings

²⁵<http://metadataregistry.org/schema/show/id/15.html>

zwischen unabhängigen Schemata geeignet und nur mit äußerster Vorsicht einzusetzen.

2. Eine weitere Möglichkeit nutzt die Tatsache, dass eigenständige Mapping-Sprachen existieren, die mittels verschiedener Tools genutzt werden können um zum Beispiel Instanzdaten von einem Schema in ein anderes zu übersetzen. Einige Beispiele für solche Mapping-Sprachen sind in [34] aufgeführt. Der Nachteil ist jedoch, dass diese Sprachen und Tools keine allgemeine Verbreitung gefunden haben und somit in der Regel nicht durch die Benutzer eines Linked Data-Datenbestands interpretiert werden können. Daher existiert auch der Vorschlag, die Mapping-Relationen aus SKOS wie z. B. `skos:closeMatch` für diesen Zweck zu verwenden. In diesem Fall werden aber statt OWL-Klassen oder -Properties aus den Vokabularen zwei OWL-Instanzen vom Typ `skos:Concepts` aufeinander gemappt. Demzufolge würde dieser Lösungsvorschlag erfordern, dass ein Reasoner die so aufeinander gemappten OWL-Klassen gleichzeitig auch als OWL-Instanzen interpretiert. Dies ist jedoch nur in OWL-Full Ontologien möglich.²⁶ Daraus folgt, dass das Mapping von OWL-Klassen mit Hilfe von `skos:closeMatch` nicht als Best Practice empfohlen werden kann.

5.3 Abfragen über Linked Data

Ein weiterer Baustein zur Bereitstellung von Linked Data ist ein sogenannter SPARQL-Endpunkt. SPARQL [13] ist eine Anfragesprache für RDF-Daten, vergleichbar mit SQL für relationale Datenbanken. Mithilfe von SPARQL kann ein kompletter Datenbestand nach bestimmten Mustern durchsucht werden und falls diese Muster an einer oder mehreren Stellen im RDF-Graphen des Datenbestandes gefunden werden, wird dieser Teilgraph für die Generierung des Ergebnisses verwendet. SPARQL erlaubt dabei die Rückgabe einzelner Tupel-Werte genauso wie die Generierung von Ergebnis-RDF-Graphen und hat sich im Semantic Web als Standardmethode etabliert Linked Data automatisiert auf einem Server abzufragen und zu durchsuchen. Der Vorteil liegt darin, dass der Client nicht zuerst alle Daten heruntergeladen haben muss um sie zu durchsuchen. Zudem kann auch über die Grenzen von Datenquellen hinweg abgefragt werden, sofern für jede der Datenquellen ein SPARQL-Endpunkt zur Verfügung steht.

Als weitere Alternative bzw. Ergänzung zu SPARQL-Endpunkten gibt es noch den ONKI SKOS Server [42] von Tuominen et. al.. Er ermöglicht zum Beispiel die Abfrage von Thesaurus-Bestandteilen mithilfe eines Web Services sowie den Zugriff mittels eines auf AJAX basierenden Benutzerinterfaces. Der ONKI SKOS Server erlaubt sowohl das Publizieren, als auch das Verarbeiten und Anzeigen von SKOS Thesauri und basiert auf anerkannten Webstandards wie Java-Servlets, Jena RDF-Store und Lucene Indexing.

6 Transformation der Schlagwortnormdatei

Die Transformation der Schlagwortnormdatei (SWD) der Deutschen Nationalbibliothek wurde bereits in einer vorläufigen Version durchgeführt. Die Erfahrungen und Vorgehensweisen sind in [17, 27] beschrieben. Es wurde in den vorliegenden Fällen festgestellt, dass

²⁶In OWL2 würde eine solche Ontologie nicht mehr automatisch als OWL Full klassifiziert. Dort ist es möglich, die gleiche URI sowohl für eine Klasse als auch für eine Instanz zu verwenden. Trotzdem würden die Klasse und die Instanz semantisch als zwei unterschiedliche Dinge angesehen. Das `skos:closeMatch` im obigen Beispiel würde sich also aus semantischer Sicht nicht auf die OWL-Klassen beziehen, sondern auf zwei neu eingeführte SKOS-Konzepte, die „zufälligerweise“ die gleiche URI wie die OWL-Klassen haben.

offene Probleme zum einen im Bereich der Integration mit anderen Datenbeständen liegen (mehrgliedrige Oberbegriffe enthalten auch Schlagwörter aus GKD oder PND, welche nicht in der transformierten SWD vorliegen) und zum anderen u. a. in der Notwendigkeit die SWD-Sachgruppen ebenfalls nach SKOS zu transformieren, welches bisher noch nicht geschehen ist.

6.1 Einordnung in das allgemeine Phasenmodell

Die bereits durchgeführte Transformation lässt sich größtenteils im Rahmen des in Abschnitt 4.1 erläuterten allgemeinen Phasenmodells in die Bereiche “Kickoff” und “Verfeinerung und Anpassung” einordnen. Zur Phase **Kickoff** gehört, dass die Transformation der SWD nach Linked Data in einer prototypischen Version zum 31.03.2010 erstmalig vorlag. Diese Version wurde dann gemäß dem Phasenmodell zur Revision in den Fachabteilungen freigegeben. In weiteren Versionen wurde u. a. die SWD stetig ausgebaut, beispielsweise um DDC-Sachgruppen erweitert oder Anpassungen am URI-Schema vorgenommen. Diese Beta-Phase, die in enger Rücksprache mit der Fachcommunity durchgeführt wurde und zu stetigen Verbesserungen in einem iterativen Verfahren führte, ist folglich in die Phase **Verfeinerung und Anpassung** einzuordnen.

6.2 Analyse des Schemas der SWD

Als Grundlage für das zu transformierende Schema wurde das Feldverzeichnis der SWD genutzt.²⁷ Auf dieser Basis wurde entschieden, welche Informationen in die Linked Data-Repräsentation der SWD übertragen werden sollen (siehe auch [17]). Aktuell umfasst die Linked Data-Repräsentation der SWD nur die Sachschlagwörter (topical headings). Aufgrund der Beschränkung auf die Sachschlagwörter eignet sich daher insbesondere die SKOS-Spezifikation für die Darstellung des Schemas der SWD (siehe auch Abschnitt 3.1).

Ein weiterer Bestandteil der Analyse ist die vorgesehene Verwendung der Linked Data Repräsentation. Zum jetzigen Zeitpunkt wird die Linked Data-Repräsentation der Schlagwortnormdatei innerhalb des DNB-Katalogs zur Verfügung gestellt. Sie kann genutzt werden um weiterführende Informationen über die an einem Titelbestand annotierten Sachschlagwörter zu erhalten. Es ist vorgesehen, die Linked Data-Repräsentation zu einem wichtigen Bestandteil des sogenannten *Global Cultural Graph* auszubauen, der hochwertige und zuverlässige Informationen im Rahmen des Semantic Web zur Verfügung stellen soll (siehe [27]).

6.3 Mapping auf das Zielschema im Semantic Web

Prinzipiell setzt sich ein SWD-Eintrag aus einer Kombination von Feldern und speziellen Indikatoren im Format *PICA+* zusammen. Zum Beispiel steht ein Hauptschlagwort in der SWD im Feld 041A und wird durch den Indikator **\$s** identifiziert. Für die Transformation wurde eine Untermenge der Felder aus der Menge der, in der SWD verfügbaren, Felder ausgewählt. Diese Untermenge umfasst insgesamt 12 Felder mit ihren Indikatoren.

Jedem dieser Felder wurde eine Repräsentation in RDF zugeordnet und dabei nach Möglichkeit auf existierende Vokabulare zurückgegriffen. In wenigen Fällen konnte kein existierendes Vokabular gefunden werden, so dass auf eine eigene Definition zurückgegriffen wurde. Für Details siehe [17]. Trotz allem konnte die Mehrzahl der umgesetzten Felder der SWD mit Hilfe von SKOS realisiert werden. So wurde zum Beispiel das Hauptschlagwort als `skos:prefLabel` definiert und `skos:broader` verweist auf das übergeordnete

²⁷http://support.d-nb.de/iltis/feldverzeichnis/Normdaten_SWD_Cross.pdf

Schlagwort. Diese Elemente und die restlichen Felder werden genauer in Tabelle 4 in [17] erläutert.

Im Rahmen des Linked Data-Projekts an der Deutschen Nationalbibliothek wurden weiterhin noch verschiedene andere Vokabulare auf ihre Tauglichkeit für das Projekt untersucht. So musste für die Umsetzung, der in der Gemeinsamen Normdatei enthaltenen Körperschaften und Personen, auf andere Vokabulare zurückgegriffen werden, da diese nicht durch SKOS unterstützt werden. Insgesamt wurden 12 verschiedene Vokabulare auf ihre Tauglichkeit überprüft. Eine Auflistung der untersuchten Vokabulare und in welchem Maße sie für eine Nachnutzung im Bibliothekswesen geeignet sind, kann [17] entnommen werden.

Einen weiteren wichtigen Bestandteil des Linked Data-Services der Deutschen Nationalbibliothek stellen die Mappings auf andere Datenbestände in der Linked Data-Cloud dar. Ein Beispiel hierfür sind die im Rahmen des CrissCross-Projekts²⁸ erstellten Mappings zwischen SWD-Sachschlagwörtern und Notationen der Dewey Dezimalklassifikation. Die CrissCross-Relationen sind ursprünglich nur von einem SWD-Schlagwort in Richtung DDC zu benutzen. In [23] wurde hingegen gezeigt, wie mit Hilfe von *koordinierten Konzepten* diese unidirektionalen CrissCross-Relationen auf die in SKOS definierten bidirektionalen Mapping-Relationen `skos:closeMatch`, `skos:broadMatch` und `skos:relatedMatch` abgebildet werden können. In Abbildung 6 ist ein Beispiel für eine CrissCross-Relation angegeben.

Erst durch die Verknüpfung mit anderen Datenbeständen wird der eigene Datenbestand in die Linked Data-Wolke integriert und ein wichtiges Kriterium von Linked Data wird erfüllt (siehe Abschnitt 2.1). Im Fall der SWD existieren neben den Criss-Cross-Relationen zur Dewey Dezimalklassifikation u. a. noch Verknüpfungen mit den *Library of Congress Subject Headings (LCSH)* und *RAMEAU*. In vielen Fällen bietet sich auch eine Verknüpfung mit dem DBPedia-Datenbestand an, der sich zu einem zentralen Datenbestand in der Linked Data-Wolke entwickelt hat (siehe [22]).

6.4 Transformation der Daten der SWD

In diesem Abschnitt sollen 2 Punkte bei der durchgeführten Transformation der SWD betrachtet werden. Wie bereits in Abschnitt 5.1 ausgeführt, wird der Namespace der SWD über den gemeinsamen Namespace der GND <http://d-nb.info/gnd/> realisiert. Dabei wird pro Sachschlagwort unter Verwendung der internen Identifikationsnummer eine URI mit Slash-Namespace verwendet. Die Gründe hierfür sind, dass zum einen die SWD insgesamt sehr groß ist und zum anderen ein Benutzer in der Regel immer nur an einem kleinen Ausschnitt der SWD interessiert ist (z. B. an einem Schlagwort zusammen mit seinen direkten Ober- und Unterschlagwörtern). Durch die Verwendung des Slash-Namespace kann so der Benutzer gezielt die Informationen abfragen, die ihn interessieren und muss nicht immer die komplette SWD herunterladen.

Bezüglich der unterschiedlichen Repräsentationen eines Sachschlagwortes wird sowohl eine RDF- als auch eine HTML-Repräsentation angeboten. Die HTML-Repräsentation wird beispielsweise im Rahmen des Katalogs der DNB auf <http://www.dnb.de> verwendet (siehe Abbildung 5). Im Katalog kann darüber hinaus auf die RDF-Repräsentation eines Schlagwortes umgeschaltet werden (siehe Abbildung 6). Diese ist selbstverständlich auch unabhängig vom Katalog über das bereits genannte URI-Schema von jedem RDF-fähigen Client aus zugreifbar.

²⁸<http://www.d-nb.de/wir/projekte/crisscross.htm>

SKOS	
Link zu diesem Datensatz	http://d-nb.info/gnd/4042855-2
Schlagwort	Nymphensittich
Quelle	M
Typ	Sachschlagwort (Indikator s)
Synonyme	Nymphicus hollandicus
DDC-Notation	598.71 636.68656
Thema in	50 Publikationen <ol style="list-style-type: none"> 1. <i>Perruches calopsittes</i> Paris : Hachette Pratique, 2010 2. <i>Nymphensittiche</i> Radtke, Georg. - Stuttgart (Hohenheim) : Ulmer, 2009, 4., neubearb. Aufl. 3. <i>Cockatiels</i> Hauppauge, NY : Barron, 2008, 2. ed. 4. <i>Nymphensittiche</i> München : Gräfe und Unzer, 2009, 1. Aufl. 5. <i>Les perruches calopsittes</i> [Alleur] : Marabout, 2008 6. ...

Abbildung 5: Ausschnitt der HTML-Repräsentation für das Sachschlagwort *Nymphensittich*

```

<skos:Concept rdf:about="http://d-nb.info/gnd/4042855-2">
  <skos:closeMatch
    rdf:resource="http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb12233750m"/>
  <skos:prefLabel xml:lang="de">Nymphensittich</skos:prefLabel>
  <dcterms:identifler>(DE-588)040428559</dcterms:identifler>
  <skos:closeMatch rdf:resource="http://id.loc.gov/authorities/sh85027598#concept"/>
  <dnb:hasCoordinatedConcept-of>
    <dnb:CoordinatedConcept>
      <dnb:coordination-of rdf:resource="http://d-nb.info/ddc-sg/630"/>
      <dnb:coordination-of rdf:resource="http://d-nb.info/gnd/4042855-2"/>
      <dnb:det3 rdf:resource="http://d-nb.info/ddc/class/636.68656"/>
    </dnb:CoordinatedConcept>
  </dnb:hasCoordinatedConcept-of>
  <dnb:hasCoordinatedConcept-of>
    <dnb:CoordinatedConcept>
      <dnb:coordination-of rdf:resource="http://d-nb.info/ddc-sg/590"/>
      <dnb:coordination-of rdf:resource="http://d-nb.info/gnd/4042855-2"/>
      <dnb:det2 rdf:resource="http://d-nb.info/ddc/class/598.71"/>
    </dnb:CoordinatedConcept>
  </dnb:hasCoordinatedConcept-of>
  <skos:altLabel xml:lang="de">Nymphicus hollandicus</skos:altLabel>
  <dcterms:identifler>(DE-588c)4042855-2</dcterms:identifler>
</skos:Concept>
    
```

Abbildung 6: Ausschnitt der RDF/XML-Repräsentation für das Sachschlagwort *Nymphensittich*. Das Schlagwort ist im Kontext der Sachgruppe 590 (Tiere, Zoologie) mit Hilfe der CrissCross-Relation `dnb:det2` auf die DDC-Notation 598.71 abgebildet. Die Relation `dnb:det2` ist abgeleitet von `skos:broadMatch` (siehe [23]).

7 Transformation des Thesaurus Technik und Management

Im Folgenden wird das Vorgehen bei der Transformation des Thesaurus Technik und Management (TEMA) zu einer Linked Data-Repräsentation beschrieben. Der Thesaurus Technik und Management wird durch die WTI Frankfurt eG bereitgestellt und enthält über 150.000 Fachbegriffe aus verschiedenen, technisch orientierten Anwendungsdomänen. Er findet Verwendung bei der Verschlagwortung von Literaturnachweisen, die im Rahmen von verschiedenen Fachdatenbanken zur Verfügung gestellt werden (siehe auch Abschnitt 1.3.2). Neben den Schlagwörtern aus dem Thesaurus werden die Literaturnachweise darüber hinaus mithilfe der Fachordnungsklassifikation²⁹ klassifiziert. Ein Beispiel für die Erschließung eines Literaturnachweises mithilfe des Thesaurus und der Klassifikation ist zu finden unter <http://www.wti-frankfurt.de/images/stories/download/de-tema.pdf>.

7.1 Analyse des Schemas von TEMA

Bisher wird der Thesaurus Technik und Management mittels einer Datenbank verwaltet, in welcher vier verschiedene Tabellen enthalten sind, die wie folgt beschrieben werden können:

- In **Tabelle 1** werden die verschiedenen Terme abgespeichert. Jeder Term hat ein Label sowie eine interne Nummer und es ist für jeden Term abgespeichert, ob es sich hierbei um eine Vorzugsbenennung, ein deutsches Synonym, ein englisches Synonym oder ein verstecktes Synonym handelt. Wie die Terme zueinander in Beziehung stehen, ist separat in Tabelle 3 abgespeichert.
- In **Tabelle 2** werden Hinweise zu Termen abgespeichert. Die Information, welcher Hinweis sich auf jeweils welche Vorzugsbenennung bezieht, ist in Tabelle 3 abgespeichert. Außerdem enthält Tabelle 3 noch die Notationen und Label der Fachordnungsklassifikation (siehe oben). Die darin enthaltenen Fachgebiete können anhand ihrer Notation hierarchisch gegliedert werden. So ist beispielsweise das Fachgebiet *3BF Optik* dem Fachgebiet *3BFX nichtlineare Optik* übergeordnet, d. h. auf jeder Hierarchieebene wird ein weiterer Buchstabe an die Notation des übergeordneten Fachgebietes angehängt.
- In **Tabelle 3** werden die Beziehungen zwischen den verschiedenen Termen abgespeichert. Die deutschen, englischen und versteckten Synonyme sowie die Hinweise werden jeweils ihrer Vorzugsbenennung zugeordnet, welche mithilfe der Relationen *Oberbegriff*, *Unterbegriff* und *Verweis (assoziative Relation)* selbst zueinander in Beziehung gesetzt wird.
- In **Tabelle 4** wird die Zuordnung zwischen den Vorzugsbenennungen und deren deutschen Synonymen auf der einen Seite und den dazugehörigen englischen Übersetzungen auf der anderen Seite vorgenommen. Zum Beispiel wird der Vorzugsbenennung *Ausfallrate* in dieser Tabelle die Übersetzung *failure rate* zugeordnet und dem dazugehörigen deutschen Synonym *Ausfallhäufigkeit* die Übersetzung *failure frequency*. In Tabelle 3 sind beide englischen Übersetzungen nur als englische Synonyme für die Vorzugsbenennung abgespeichert, d. h. mit den Informationen aus Tabelle 3 ist die Zuordnung von *failure frequency* als direkte Übersetzung von *Ausfallhäufigkeit* nicht möglich.

²⁹<http://www.wti-frankfurt.de/images/stories/download/WTI-Fachordnung.pdf>

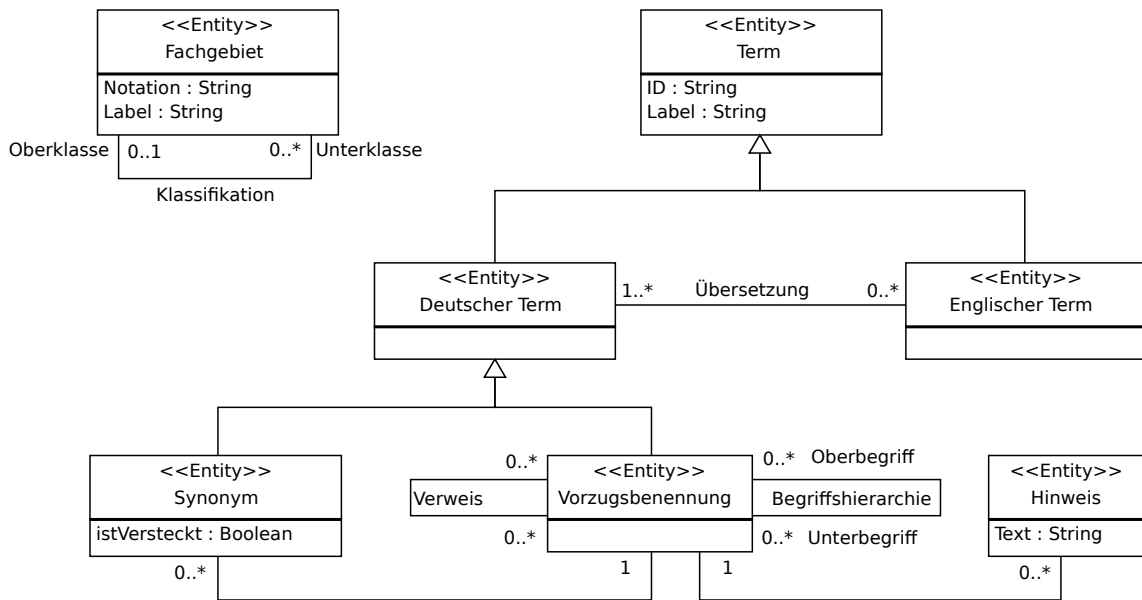


Abbildung 7: Entity-Relationship Diagramm des Schemas des Thesaurus Technik und Management und der Fachordnungsklassifikation.

Basierend auf einer Analyse dieser vier Tabellen des Thesaurus Technik und Management und der Fachordnungsklassifikation kann das in Abbildung 7 gezeigte Entity-Relationship Diagramm des Schemas erstellt werden. Demzufolge ist der Thesaurus Technik und Management ein termbasierter Thesaurus, da alle Beziehungen direkt zwischen den Termen existieren. Der Thesaurus wird postkoordiniert eingesetzt, da im Rahmen der Inhaltserschließung Folgen von Schlagwörtern vergeben werden und eine Kombination von Schlagwörtern nicht bereits im Rahmen des Thesaurus stattfindet. Schließlich ist noch zu erwähnen, dass keinerlei Beziehung besteht zwischen den Fachgebieten aus der Klassifikation und den Termen im Thesaurus. Beide sind daraus folgend voneinander unabhängige Erschließungsinstrumente.

7.2 Mapping auf das Zielschema im Semantic Web

Eine Abbildung auf das, durch SKOS spezifizierte, Thesaurus-Schema ist ohne größere Probleme möglich, da der Thesaurus Technik und Management einen termbasierten Thesaurus repräsentiert. Analog gilt dies für die Abbildung der Fachordnungsklassifikation auf SKOS. In Tabelle 1 ist aufgelistet, wie die verschiedenen Elemente und Beziehungen aus dem Entity-Relationship Diagramm von Abbildung 7 auf die SKOS-Spezifikation konkret abgebildet werden. Hierbei sind zwei Punkte hervorzuheben, die besondere Aufmerksamkeit erfordern:

1. Die SKOS-Spezifikation stellt ein Schema für konzeptbasierte KOS zur Verfügung. Bei der Abbildung von termbasierten Thesauri wie dem Thesaurus Technik und Management auf die SKOS-Spezifikation ist daher ein besonderes Augenmerk darauf zu richten, wie die Konzepte modelliert werden und welchen Identifier die Konzepte bekommen. In unserem Fall haben wir uns dafür entschieden die ID der Vorzugsbenennung als Grundlage für die URI des Konzeptes zu benutzen. Die Details dazu sind der Beschreibung des Entwurfs des URI-Namespaces in Abschnitt 7.3 zu entnehmen.

2. In der SKOS-Spezifikation selbst sind keine Beziehungen zwischen den Labels vorgesehen, wie sie im Fall der **Übersetzung**-Beziehung zwischen deutschen und englischen Termen vorliegt. Um Beziehungen zwischen Labels modellieren zu können muss auf die *SKOS eXtension for Labels* (SKOS-XL) zurückgegriffen werden [11]. Diese Erweiterung der SKOS Spezifikation stellt die Beziehung `skosxl:labelRelation` zur Verfügung, von der die **Übersetzung**-Beziehung abgeleitet werden kann. Anstatt jedoch mithilfe von `skosxl:labelRelation` eine eigene Übersetzungs-Relation zwischen zwei Labels einzuführen, verwenden wir die Relation `ext:hasTranslation`³⁰, die bereits für die Linked Data-Repräsentation des Thesaurus Sozialwissenschaften etabliert wurde (siehe auch [32]).

Im Gegensatz zur Transformation der SWD liegen im Fall des Thesaurus Technik und Management keinerlei Mappings zu anderen Datenbeständen der Linked Data-Wolke vor. Innerhalb des Thesaurus sind nur Mappings auf den DKF-Thesaurus³¹ vorhanden. Da dieser Thesaurus aber nicht in einer Linked Data-Repräsentation vorliegt, sind diese Mappings nicht geeignet, um den Thesaurus Technik und Management in die Linked Data-Wolke zu integrieren. Aus diesem Grund sollte für zukünftige Versionen des Thesaurus überprüft werden, ob in (semi-)automatischen Verfahren Mappings auf andere, bereits in der Linked Data-Wolke vorhandene Datenbestände, bereitgestellt werden können.

Wie bereits in Abschnitt 7.1 erwähnt, wird durch die WTI Frankfurt eG neben dem Thesaurus Technik und Management auch die Fachordnungsklassifikation zur Erschließung von Dokumenten genutzt. Wie in Abbildung 7 dargestellt ist, besteht diese nur aus den hierarchischen Beziehungen zwischen einzelnen Fachgebieten. Dabei ist jedem Fachgebiet weiterhin eine Notation zugewiesen. Auch für die Fachordnungsklassifikation muss berücksichtigt werden, wie die URI eines `skos:Concept`, das jeweils einem Fachgebiet entspricht, im Einzelnen aufgebaut ist. Die Details hierzu sind Abschnitt 7.3 zu entnehmen. Auch für die Fachordnungsklassifikation gilt, dass sie bisher in keinerlei Beziehung zu einem anderen, in der Linked Data-Wolke vorhandenen, Datenbestand steht.

7.3 Transformation der Daten von TEMA

Nachdem im vorherigen Abschnitt das Mapping zwischen dem ursprünglichen Thesaurus-Schema und seiner entsprechenden Repräsentation im Semantic Web angesprochen wurde, muss für die Transformation der Daten noch geklärt werden, wie der URI-Namespace für den Thesaurus Technik und Management und für die Fachordnungsklassifikation aufgebaut werden soll. In Abschnitt 5.1 wurde grundsätzlich unterschieden zwischen einem Hash-Namespace und einem Slash-Namespace. Das wichtigste Kriterium für die Entscheidung zwischen diesen beiden Varianten ist die Größe des Datenbestandes. Nur für kleine Datenbestände sollte die Verwendung eines Hash-Namespace in Erwägung gezogen werden.

Im Fall des Thesaurus Technik und Management können wir demzufolge von vornherein die Verwendung eines Hash-Namespace ausschließen, da dieser Datenbestand momentan Informationen über 52.031 Konzepte und 161.214 Terme enthält. Somit ist dieser Datenbestand deutlich zu groß um ihn in einer einzigen großen Datei anbieten zu können, welches jedoch die Voraussetzung für den Hash-Namespace ist. Im Fall der Fachordnungsklassifikation käme ein Hash-Namespace prinzipiell noch in Frage, da diese nur 415 hierarchisch angeordnete Fachgebiete umfasst. Um aber eine einheitliche Gestaltung

³⁰<http://lod.gesis.org/data/ext/hasTranslation>

³¹http://www.dkf-ev.de/dkf_thes.pdf

Original-Element	RDF-Element	Bemerkung
Vorzugsbenennung	<code>skos:Concept</code> und <code>skosxl:prefLabel</code>	Die URI des <code>skos:Concept</code> und des <code>skosxl:prefLabel</code> basieren beide auf der ID der Vorzugsbenennung. Details sind Abschnitt 7.3 zu entnehmen. Das Attribut <code>xml:lang</code> des <code>skosxl:prefLabel</code> wird auf den Wert <code>de</code> gesetzt.
Synonym (<code>istVersteckt=false</code>)	<code>skosxl:altLabel</code>	Die URI basiert auf der ID des Synonyms (siehe auch Abschnitt 7.3). Das Attribut <code>xml:lang</code> des <code>skosxl:altLabel</code> wird auf den Wert <code>de</code> gesetzt.
Synonym (<code>istVersteckt=true</code>)	<code>skosxl:hiddenLabel</code>	Die URI basiert auf der ID des Synonyms (siehe auch Abschnitt 7.3). Das Attribut <code>xml:lang</code> des <code>skosxl:hiddenLabel</code> wird auf den Wert <code>de</code> gesetzt.
Englischer Term	<code>skosxl:altLabel</code>	Die URI basiert auf der ID des Terms (siehe auch Abschnitt 7.3). Das Attribut <code>xml:lang</code> wird auf den Wert <code>en</code> gesetzt.
Hinweis	<code>skos:note</code>	Der Hinweis steht in Beziehung zu dem <code>skos:Concept</code> , das für die Vorzugsbenennung erzeugt wurde.
Begriffshierarchie	<code>skos:broader</code>	Die Hierarchie wird zwischen den für die Vorzugsbenennung erzeugten Konzepten definiert.
Verweis	<code>skos:related</code>	Der Verweis wird zwischen den für die Vorzugsbenennung erzeugten Konzepten definiert.
Übersetzung	<code>ext:hasTranslation</code>	Die Beziehung wird hergestellt zwischen dem <code>skosxl:prefLabel</code> bzw. <code>skosxl:altLabel</code> des deutschen Terms und dem <code>skosxl:altLabel</code> des englischen Terms.
Fachgebiet	<code>skos:Concept</code> , <code>skos:prefLabel</code> und <code>skos:notation</code>	Die URI des <code>skos:Concept</code> basiert auf der Notation des Fachgebietes (siehe auch Abschnitt 7.3). Das Label wird als <code>skos:prefLabel</code> modelliert und die Notation darüber hinaus als <code>skos:notation</code> .
Klassifikation	<code>skos:broader</code>	

Tabelle 1: Mapping der Elemente aus Abbildung 7 auf die entsprechenden Elemente der SKOS-Spezifikation. Das Präfix `skos` steht für den Namespace <http://www.w3.org/2004/02/skos/core#>, `skosxl` für den Namespace <http://www.w3.org/2008/05/skos-xl#> und `ext` für den Namespace <http://lod.gesis.org/data/ext/>.

der beiden Namespaces zu erhalten, sollte auch hier ein Slash-Namespace verwendet werden.

Nachdem diese Entscheidung für den Slash-Namespace getroffen wurde, muss im Folgenden noch entschieden werden, welche Domäne für den Namespace verwendet wird, wie der Pfad gestaltet wird und wie die IDs der einzelnen Terme, Konzepte und Fachgebiete generiert werden sollen:

- Für den URI-Namespace sollte immer die Domäne des Anbieters des Datenbestandes verwendet werden. Im Fall des Thesaurus Technik und Management und der Fachordnungsklassifikation ist das die Domäne der WTI Frankfurt eG. Ferner bietet es sich an für die Bereitstellung der Daten eine eigene Unterdomäne einzuführen um sie besser von der Webseite des Anbieters unterscheidbar zu machen. Daher wird in unserem Fall als Basis für den URI-Namespace die Adresse <http://lod.wti-frankfurt.de> verwendet, wobei die Unterdomäne `lod` für *Linked Open Data* steht.
- Als nächstes muss der Pfad für die verschiedenen Datenbestände des Anbieters definiert werden. Hier müssen wir in unserem Fall zwischen dem Thesaurus Technik und Management und der Fachordnungsklassifikation unterscheiden. Da es in Zukunft sein kann, dass nicht nur Daten bereitgestellt werden, sondern auch eigene Erweiterungen für SKOS oder andere Spezifikationen, werden die verschiedenen Datenbestände unter der URI <http://lod.wti-frankfurt.de/data/> referenziert und die Schema-Informationen unter <http://lod.wti-frankfurt.de/schema/>. Der Thesaurus Technik und Management wird dementsprechend somit unter <http://lod.wti-frankfurt.de/data/thesaurus/> und die Fachordnungsklassifikation unter <http://lod.wti-frankfurt.de/data/classification/> zugreifbar.
- Bei der Wahl der ID von z. B. einem Konzept oder einem Term ist es wichtig, dass eine Verbindung zu den bisherigen Bereitstellungsformen möglich ist. Es sollten für die Linked Data-Repräsentation eines Datenbestandes in der Regel keine komplett neuen IDs eingeführt werden, sondern tendenziell immer auf die bestehenden zurückgegriffen werden. Im Fall des Thesaurus werden daher die Termnummern als ID verwendet, die bisher intern in der Datenbank vergeben wurden. Bei der Fachordnungsklassifikation werden die Notationen als ID verwendet. Die jeweilige ID wird hier einfach an den Basispfad des Datenbestandes angehängt. Das Fachgebiet *3BX chemische Grundlagen* wird z. B. durch die URI <http://lod.wti-frankfurt.de/data/classification/3BX> bezeichnet (siehe Abbildung 8).

In Abschnitt 7.2 wurde darauf hingewiesen, dass der Thesaurus Technik und Management ursprünglich ein termbasierter Thesaurus ist und bei der Transformation zu einer SKOS-Repräsentation aus diesem Grund Konzepte eingeführt werden müssen. Dabei bietet es sich an die ID der Vorzugsbenennung für die neu eingeführten Konzepte zu verwenden. Um diese von jenen der Vorzugsbenennung unterscheidbar zu machen, verwenden wir jeweils zwei unterschiedliche Teilpfade für die URIs der Konzepte und der Terme. So hat die Vorzugsbenennung *Aufsatzgerät für erdgebundenes Digitalfernsehen* die interne Termnummer *067017* beispielsweise. Das dazugehörige `skos:Concept` erhält die URI <http://lod.wti-frankfurt.de/data/thesaurus/concepts/067017> und das `skosxl:prefLabel` benutzt die URI <http://lod.wti-frankfurt.de/data/thesaurus/terms/067017>. Von dem Label wird dann mittels der Relation `ext:hasTranslation` auf die dazugehörige englische Übersetzung mit der Termnummer *067018* verwiesen (siehe Abbildung 9).

```

<rdf:Description rdf:about="http://lod.wti-frankfurt.de/data/classification/3BX">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:inScheme>http://lod.wti-frankfurt.de/data/classification/</skos:inScheme>
  <skos:notation>3BX</skos:notation>
  <skos:prefLabel xml:lang="de">chemische Grundlagen</skos:prefLabel>
  <skos:broader rdf:resource="http://lod.wti-frankfurt.de/data/classification/3B"/>
</rdf:Description>

<rdf:Description rdf:about="http://lod.wti-frankfurt.de/data/classification/3BXP">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:inScheme>http://lod.wti-frankfurt.de/data/classification/</skos:inScheme>
  <skos:notation>3BXP</skos:notation>
  <skos:prefLabel xml:lang="de">physikalische Chemie</skos:prefLabel>
  <skos:broader rdf:resource="http://lod.wti-frankfurt.de/data/classification/3BX"/>
</rdf:Description>
    
```

Abbildung 8: Ausschnitt der RDF/XML-Repräsentation der Fachordnungsklassifikation der WTI Frankfurt eG.

```

<rdf:Description rdf:about="http://lod.wti-frankfurt.de/data/thesaurus/concepts/067017">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:inScheme>http://lod.wti-frankfurt.de/data/thesaurus/</skos:inScheme>
  <skosxl:prefLabel rdf:resource="http://lod.wti-frankfurt.de/data/thesaurus/terms/067017"/>
  <skosxl:altLabel rdf:resource="http://lod.wti-frankfurt.de/data/thesaurus/terms/067019"/>
  <skosxl:altLabel rdf:resource="http://lod.wti-frankfurt.de/data/thesaurus/terms/067020"/>
  <skos:related rdf:resource="http://lod.wti-frankfurt.de/data/thesaurus/concepts/069410"/>
  <skos:related rdf:resource="http://lod.wti-frankfurt.de/data/thesaurus/concepts/067001"/>
  <skos:broader rdf:resource="http://lod.wti-frankfurt.de/data/thesaurus/concepts/067015"/>
</rdf:Description>

<rdf:Description rdf:about="http://lod.wti-frankfurt.de/data/thesaurus/terms/067017">
  <ext:hasTranslation rdf:resource="http://lod.wti-frankfurt.de/data/thesaurus/terms/067018"/>
  <rdf:type rdf:resource="http://www.w3.org/2008/05/skos-xl#Label"/>
  <skosxl:literalForm xml:lang="de">Aufsatzgerät für erdgebundenes Digitalfernsehen</skosxl:literalForm>
</rdf:Description>

<rdf:Description rdf:about="http://lod.wti-frankfurt.de/data/thesaurus/terms/067018">
  <skosxl:literalForm xml:lang="en">digital video broadcasting terrestrial settop boxes</skosxl:literalForm>
  <rdf:type rdf:resource="http://www.w3.org/2008/05/skos-xl#Label"/>
</rdf:Description>

<rdf:Description rdf:about="http://lod.wti-frankfurt.de/data/thesaurus/terms/067019">
  <ext:hasTranslation rdf:resource="http://lod.wti-frankfurt.de/data/thesaurus/terms/067020"/>
  <rdf:type rdf:resource="http://www.w3.org/2008/05/skos-xl#Label"/>
  <skosxl:literalForm xml:lang="de">DVB-T-Aufsatzgerät</skosxl:literalForm>
</rdf:Description>

<rdf:Description rdf:about="http://lod.wti-frankfurt.de/data/thesaurus/terms/067020">
  <rdf:type rdf:resource="http://www.w3.org/2008/05/skos-xl#Label"/>
  <skosxl:literalForm xml:lang="en">DVB T settop boxes</skosxl:literalForm>
</rdf:Description>
    
```

Abbildung 9: Ausschnitt der RDF/XML-Repräsentation aus dem TEMA-Thesaurus für das Schlagwort *Aufsatzgerät für erdgebundenes Digitalfernsehen*. Für die deutschen Schlagwörter existiert auch jeweils eine englische Übersetzung, die mithilfe der Relation `ext:hasTranslation` verlinkt ist.

8 Zusammenfassung

In diesem Dokument wurden grundlegende Begriffe und Methoden vorgestellt, die sich als hilfreich bei der Erstellung einer Semantic Web konformen Repräsentation von KOS erwiesen haben, wie z. B. Thesauri und Klassifikationssysteme. Besonders wichtig in diesem Zusammenhang sind die Anforderungen und Prinzipien, die eine Semantic Web konforme Repräsentation erfüllen sollte um leicht verarbeitbar zu sein (siehe Abschnitt 2.1 und 2.3).

Die Wiederverwendung bereits etablierter Schemata bildet hier ein weiteres, wichtiges Grundprinzip, um die Interoperabilität des eigenen Datenbestandes mit anderen Datenbeständen zu ermöglichen. Nur wenn in etablierten Schemata keine passenden Elemente gefunden werden, sollten eigene Erweiterungen von etablierten Schemata zum Einsatz kommen. Als Beispiel für diese Ausnahme von der Regel sind die Erweiterungen anzusehen, die im Rahmen der Transformation der SWD in Abschnitt 6.3 vorgenommen wurden. So musste bspw. im Fall der SWD das neue Element `gnd:invalidIdentifierForTheSubject` eingeführt werden. Falls solche eigenen Elemente eingeführt werden, sollten sie gut dokumentiert und nach Möglichkeit von bestehenden Elementen aus z. B. SKOS abgeleitet sein. Neben der Verwendung von etablierten Schemata ist es auch wichtig, dass die erzeugten Datenbestände natürlichsprachige Label und Dokumentation enthalten, sowie mit anderen Datenbeständen verlinkt sind und durch HTTP-Anfragen auflösbare URLs verwenden (siehe Abschnitt 2.1 und 2.3).

Neben diesen Prinzipien wurden auch zwei Vorgehensmodelle vorgestellt, die sich in der Praxis als vorteilhaft erwiesen haben. Dabei ist zu unterscheiden zwischen einem allgemeinen Phasenmodell zur Durchführung des Gesamtprojekts (siehe Abschnitt 4.1) und den konkreten Schritten, die während der eigentlichen Transformation der Daten zu durchlaufen sind (siehe Abschnitt 4.2). Es ist hier insbesondere hervorzuheben, dass sich iterative Ansätze besonders bewährt haben. Damit ist gemeint, dass eine erste Transformation der Datenbestände durchgeführt wird um anschließend diese Transformation immer weiter anzupassen und zu erweitern, basierend auf den Rückmeldungen der Benutzer der Datenbestände und der Fachabteilungen, die den zu transformierenden Datenbestand betreuen.

In diesem Dokument wurden zwei Projekte beispielhaft beschrieben, in denen die in diesem Dokument erläuterten Prinzipien und Vorgehensmodelle umgesetzt wurden bzw. auf deren Erfahrungen sie teilweise beruhen. Zum einen wurde in Abschnitt 6 das Projekt an der Deutschen Nationalbibliothek zusammengefasst, in dessen Rahmen u. a. die Schlagwortnormdatei (SWD) in eine Linked Data-Repräsentation überführt wurde. Weitere Details zu diesem Projekt sind auch den Dokumenten [17, 27] zu entnehmen. Zum anderen wurden die in Abschnitt 4.2 vorgestellten konkreten Schritte auch auf den Thesaurus Technik und Management sowie auf die Fachordnungsklassifikation der WTI Frankfurt eG angewendet. Das in Abschnitt 7 beschriebene Vorgehen ist dabei nicht in ein übergeordnetes Projekt zur Publikation dieser Linked Data-Repräsentation eingebunden, sondern soll nur die in Abschnitt 4.2 und 5 vorgestellten Vorgehensweisen und Entwurfsmuster verdeutlichen. Die in Abschnitt 4.2 vorgestellten konkreten Schritte wurden darüber hinaus aber auch bereits in vielen anderen Projekten eingesetzt, wie z. B. in [32, 33, 39, 43].

Alle in diesem Dokument vorgestellten Vorgehensweisen und Prinzipien sind nicht als normativ anzusehen. Sie sollen nur als Leitfaden für zukünftige, ähnliche Projekte herangezogen werden, damit diese von den bisher gemachten Erfahrungen profitieren können. Es ist aber ausdrücklich nicht Ziel dieses Best Practice-Guides, konkrete Muster vorzuschlagen, wie ein bestimmtes Konstrukt auf der Schema-Ebene eines KOS in z. B. SKOS oder OWL abgebildet werden kann! Für diese Art von Muster sei direkt auf die bestehende Literatur zu konkreten Transformations-Projekten verwiesen, wie z. B. [17, 23, 32, 33, 35, 39, 43].

Literatur

- [1] *Changes to VIAF's RDF*. Webseite, . – <http://outgoing.typepad.com/outgoing/2011/04/changes-to-viafs-rdf.html>
- [2] *Documentation - Guidelines for the establishment and development of monolingual thesauri. [ISO 2788:1986]*. Webseite, . – http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=7776
- [3] *Documentation - Guidelines for the establishment and development of multilingual thesauri. [ISO 5964:1985]*. Webseite, . – http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=12159
- [4] *Information and Documentation – Thesauri and Interoperability with other Vocabularies – Part 1: Thesauri for Information Retrieval [ISO 25964-1:2011]*. Webseite, . – http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail.htm?csnumber=53657
- [5] *Information and Documentation – Thesauri and Interoperability with other Vocabularies – Part 2: Interoperability with other Vocabularies [ISO/CD 25964-2]*. Webseite, . – http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail.htm?csnumber=53658
- [6] *OWL 2 Web Ontology Language Primer*. Webseite, . – <http://www.w3.org/TR/owl2-primer/>
- [7] *Quality Criteria for Linked Data sources*. Webseite, . – http://sourceforge.net/apps/mediawiki/trdf/index.php?title=Quality_Criteria_for_Linked_Data_sources
- [8] *Quality Criteria for SKOS Vocabularies*. Webseite, . – <https://github.com/cmader/qSKOS/wiki/Quality-Criteria-for-SKOS-Vocabularies>
- [9] MANOLA, Frank (Hrsg.) ; MILLER, Eric (Hrsg.): *RDF Primer - W3C Recommendation 10 February 2004*. Webseite, . – <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>
- [10] *RDF Vocabulary Description Language 1.0: RDF Schema*. Webseite, . – <http://www.w3.org/TR/rdf-schema/>
- [11] *SKOS eXtension for Labels*. Webseite, . – <http://www.w3.org/TR/2009/REC-skos-reference-20090818/#x1>
- [12] MILES, Alistair (Hrsg.) ; BECHHOFFER, Sean (Hrsg.): *SKOS Simple Knowledge Organization System - Reference W3C Recommendation 18 August 2009*. Webseite, . – <http://www.w3.org/TR/skos-reference/>
- [13] PRUD'HOMMEAUX, Eric (Hrsg.) ; SEABORNE, Andy (Hrsg.): *SPARQL Query Language for RDF - W3C Recommendation 15 January 2008*. Webseite, . – <http://www.w3.org/TR/rdf-sparql-query/>
- [14] *State of the LOD Cloud – Compliance with Best Practices for Data Provisioning*. Webseite, . – <http://www4.wiwi.fu-berlin.de/lodcloud/state/#best-practices>

- [15] *Best Practice Recipes for Publishing RDF Vocabularies - W3C Working Group Note 28 August 2008*. Webseite, 2008. – <http://www.w3.org/TR/swbp-vocab-pub/>
- [16] *Regeln für den Schlagwortkatalog, 3. überarbeitete und erweiterte Auflage auf dem Stand der 5. Ergänzungslieferung*. Elektronische Resource, April 2009. – http://files.d-nb.de/pdf/rswk_gesamtausgabe.pdf
- [17] *Der Linked Data Service der Deutschen Nationalbibliothek*. Version 3.0, Stand 15. Januar, 2011. – http://files.d-nb.de/pdf/linked_data.pdf
- [18] ASSEM, Mark V. ; MALAISE, Véronique ; MILES, Alistair ; SCHREIBER, Guus: A method to convert thesauri to SKOS. In: *Proceedings der European Semantic Web Conference (ESWC)*, Springer, 2006, S. 95–109
- [19] ASSEM, Mark V. ; MENKEN, Maarten R. ; SCHREIBER, Guus ; WIELEMAKER, Jan ; WIELINGA, Bob: A Method for Converting Thesauri to RDF/OWL. In: *Proceedings der 3. Internationalen Semantic Web Conference (ISWC)*, Springer-Verlag, 2004, S. 17–31
- [20] BERNERS-LEE, Tim: *Linked Data – Design Issues*. Webseite, . – <http://www.w3.org/DesignIssues/LinkedData>
- [21] BERNERS-LEE, Tim: *Cool URIs don't change*. W3C Hypertext Style, 1998. – <http://www.w3.org/Provider/Style/URI>
- [22] CYGANIAK, Richard ; JENTZSCH, Anja: *Linking Open Data cloud diagram*. Webseite, . – <http://lod-cloud.net>
- [23] DELLSCHAFT, Klaas: *Darstellung der CrissCross-Mappingrelationen im Rahmen des Semantic Web*. 2010
- [24] GENDT, Marjolein van ; ISAAC, Antoine ; MEIJ, Lourens van d. ; SCHLOBACH, Stefan: Semantic Web Techniques for Multiple Views on Heterogeneous Collections: A Case Study. In: GONZALO, Julio (Hrsg.) ; THANOS, Costantino (Hrsg.) ; VERDEJO, M. (Hrsg.) ; CARRASCO, Rafael (Hrsg.): *Research and Advanced Technology for Digital Libraries* Bd. 4172. Springer Berlin / Heidelberg, 2006, S. 426–437
- [25] GREEN, Rebecca ; PANZER, Michael: Relationships in the Notational Hierarchy of the Dewey Decimal Classification. In: *Proceedings des International UDC Seminars*, 2011
- [26] HAHN, Udo ; SCHULZ, Stefan: Building a Very Large Ontology from Medical Thesauri. In: *Handbook on Ontologies*. Springer Verlag Berlin Heidelberg, 2004, S. 133–150
- [27] HANNEMANN Jan ; KETT Jürgen: Linked Data for Libraries. In: *World Library and Information Congress: 76th IFLA General Conference and Assembly*. Göteborg, 2010
- [28] HARTH, Andreas: VisiNav: A system for visual search and navigation on web data. In: *Web Semantics: Science, Services and Agents on the World Wide Web* 8 (2010), Nr. 4, S. 348–354
- [29] HYVÖNEN, E. ; VILJANEN, K. ; MÄKELÄ, E. ; KAUPPINEN, T. ; RUOTSALO, T. ; VALKEAPÄÄ, O. ; SEPPÄLÄ, K. ; SUOMINEN, O. ; ALM, O. ; LINDROOS, R. u. a.: Elements of a National Semantic Web Infrastructure – Case Study Finland on the

- Semantic Web. In: *Proceedings der 1. International Semantic Computing Conference (IEEE ICSC 2007)*, 2007
- [30] HYVÖNEN, Eero ; MÄKELÄ, Eetu ; SALMINEN, Mirva ; VALO, Arttu ; VILJANEN, Kim ; SAARELA, Samppa ; JUNNILA, Miikka ; KETTULA, Suvi: MuseumFinland–Finnish museums on the semantic web. In: *Web Semantics: Science, Services and Agents on the World Wide Web 3* (2005), Nr. 2-3, S. 224 – 241
- [31] ISAAC, Antoine ; SUMMERS, Ed: *SKOS Simple Knowledge Organization System Primer*. Webseite, . – <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818>
- [32] MAYR, Philipp ; ZAPILKO, Benjamin ; SURE, York: Ein Mehr-Thesauri-Szenario auf Basis von SKOS und Crosskonkordanzen. In: *Proceedings des 25. Oberhofer Kolloquiums zur Information und Kommunikation*, 2010
- [33] NEUBERT, Joachim: Bringing the Thesaurus for Economics on to the Web of Linked Data. In: *Proceedings des Linked Data on the Web Workshops (LDOW2009)*, 2009
- [34] NOY, Natalya: Ontology Mapping. In: STAAB, Steffen (Hrsg.) ; STUDER, Rudi (Hrsg.): *Handbook on Ontologies*. 2. Ausgabe. Springer-Verlag Berlin Heidelberg, 2009, S. 573–590
- [35] PANZER, Michael ; ZENG, Marcia L.: Modeling Classification Systems in SKOS: Some Challenges and Best-Practice Recommendations. In: *Proceedings der International Conference on Dublin Core and Metadata Applications*, 2009
- [36] SAUERMANN, Leo ; CYGANIAK, Richard: *Cool URIs for the Semantic Web*. W3C Interest Group Note, März 2008. – <http://www.w3.org/TR/cooluris/>
- [37] SCHENK, Simon ; SAATHOFF, Carsten ; STAAB, Steffen ; SCHERP, Ansgar: SemaPlorer–Interactive semantic exploration of data and media based on a federated cloud infrastructure. In: *Web Semantics: Science, Services and Agents on the World Wide Web 7* (2009), Nr. 4, S. 298 – 304
- [38] SOERGEL, Dagobert ; LAUSER, Boris ; LLANG, Anita ; FISSEHA, Frehiwot ; KEIZER, Johannes ; KATZ, Stephen: Reengineering Thesauri for New Applications: the AGROVOC Example. In: *Journal of Digital Information 4* (2004)
- [39] SUMMERS, Ed ; ISAAC, Antoine ; REDDING, Clay ; KRECH, Dan: LCSH, SKOS and Linked Data. In: *Proceedings der International Conference on Dublin Core and Metadata Applications*, 2009
- [40] SURE, York ; STAAB, Steffen ; STUDER, Rudi: Handbook on Ontologies. In: STAAB, Steffen (Hrsg.) ; STUDER, Ruder (Hrsg.): *Ontology Engineering Methodology*. 2. Ausgabe. Springer Verlage Berlin Heidelberg, 2009
- [41] TUMMARELLO, Giovanni ; CYGANIAK, Richard ; CATASTA, Michele ; DANIELCZYK, Szymon ; DELBRU, Renaud ; DECKER, Stefan: Sig.ma: Live views on the Web of Data. In: *Web Semantics: Science, Services and Agents on the World Wide Web 8* (2010), Nr. 4, S. 355 – 364
- [42] TUOMINEN, Jouni ; FROSTERUS, Matias ; VILJANEN, Kim ; HYVÖNEN, Eero: ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services. In: *Proceedings der 6. European Semantic Web Conference (ESWC)*, 2009

- [43] ZAPILKO, Benjamin ; SURE, York: Converting TheSoz to SKOS. 2009 (GESIS-Technical Report 2009/07). – Forschungsbericht