

**Beitrag M: Fabio Ricci, Dietmar Wikarski**

## **SKOS Shuttle – ein Service zur Taxonomy Governance am Beispiel von Umweltinformationssystemen**

Fabio Ricci<sup>1</sup>, Dietmar Wikarski<sup>2</sup>

<sup>1</sup>*Semweb GmbH, [fabio.ricci@semweb.ch](mailto:fabio.ricci@semweb.ch),*

<sup>2</sup>*Technische Hochschule Brandenburg, [dietmar.wikarski@th-brandenburg.de](mailto:dietmar.wikarski@th-brandenburg.de)*

### **Abstract**

Nowadays, thesauri are unavoidable bridges for semantic meshups between several domains. So they can be considered also as knowledge kernels of intelligent (environmental) information systems. Since information „lives“ (changes quickly), the quality of involved thesauri and relaying information systems depends on their up-to-dateness. In this contribution we present a web service and sketch a method how thesaurus modifications can be propagated to any number of connected semantic stores, allowing information to be meshed up using current items. This allows for an effective „information governance“ of thesauri-connected information sources and thereby enhances the efficiency of information retrieval.

### **Zusammenfassung**

Thesauri sind heute unverzichtbare Brücken zur semantischen Verknüpfung unterschiedlicher Domains und bilden damit auch ein Kernstück intelligenter Umweltinformationssysteme (UIS). Da Informationen „leben“, d.h. sich sehr schnell verändern, wird die Qualität der involvierten Thesauri und der sich darauf abstützenden UIS direkt daran messbar, wie aktuell deren Daten gehalten sind. In diesem Beitrag werden ein Service und eine Methode skizziert, wie Thesaurus-Modifikationen an (grundsätzlich) beliebigen „Semantic Stores“ nachgezogen und abgefragt werden können. Dies gestattet eine effektive „Information Governance“ via Thesauri angebundener (Umwelt-)Informationsquellen und erhöht damit die Qualität und die Effizienz der Informationsgewinnung.

## 1 Thesauri als RDF-Wissensbasen für Umweltinformationen

Seit Jahrhunderten verwenden Naturwissenschaftler formale Klassifikationen, insbesondere Taxonomien, um die von ihnen untersuchten „Gegenstände“ eines Wissensgebiets zu ordnen und zueinander in Beziehung zu setzen.

Laut aktueller Wikipedia ist eine Taxonomie ([griechisch](#) τάξις bzw. táxis ‚Ordnung‘ und νόμος bzw. nómos ‚Gesetz‘) ein einheitliches Verfahren oder ein Modell (Klassifikationsschema), mit dem Objekte nach bestimmten Kriterien [klassifiziert](#), das heißt in [Kategorien oder Klassen](#) (auch [Taxa](#) genannt) eingeordnet werden.

Naturwissenschaftliche Disziplinen verwenden den Begriff der Taxonomie für eine in der Regel hierarchische Klassifikation (Klassen, Unterklassen usw.). Eine Taxonomie kann somit auch als hierarchisch organisiertes kontrolliertes Vokabular aufgefasst werden.

Ein Thesaurus ist eine Taxonomie, in der neben der hierarchischen Klassifikation („Unterbegriff“/„Oberbegriff“) weitere wohldefinierte Relationen zwischen den Begriffen definiert sind. Seine Verwendung ist sinnvoll (und oft notwendig), wenn sich Experten über gemeinsames Wissen verständigen und dieses erweitern wollen. Ein Thesaurus beschreibt die wesentlichen Begriffe („Concepts“) eines Wissensgebietes und wesentliche Relationen dieser Begriffe zueinander. Als wesentliche Relationen haben sich dabei neben den Ober- bzw. Unterbegriffsrelationen die Synonymrelation („ist synonym zu“) einschließlich eines zu verwendenden Spitzen- bzw. Referenz-Begriffs („top term“) sowie die Verwandtschaftsrelation („ist verwandt mit“) etabliert.

Durch die im Vergleich zu Taxonomien zusätzlichen Relationen (Synonym und Verwandtschaft) sind Thesauri adäquater für die Kommunikation zwischen Menschen (und Menschen helfenden Maschinen), die i.a. eine durch Synonyme und Unschärfe geprägte Kommunikation pflegen.

### RDF zur Darstellung von Semantic Stores

Zur systematischen Speicherung, insbesondere aber zum effizienten Wiederfinden elektronisch gespeicherter Aussagen haben sich im Unterschied zu relationalen Datenbanken (die zum schnellen Wiederfinden von DATEN sehr gut geeignet sind) so genannte Triple Stores – im folgenden als Semantic Stores bezeichnet – bewährt. Das entsprechende, seit 2001 standardisierte und inzwischen sehr weit verbreitete

Rahmenkonzept heißt RDF (Ressource Description Framework) [W3C, 2001] und bildet heutzutage die Grundlage für sog. *Linked Data*, die im allgemeinen eine höhere Dateninteroperabilität als relational basierte Datenmodelle gewährleisten.

### **SKOS (Simple Knowledge Organisation System) zur Formalisierung von Thesauri und Taxonomien**

SKOS ist eine vom [W3C](#) als Empfehlung veröffentlichte, auf RDF und [RDF-Schema](#) (RDFS) basierende formale Sprache zur Kodierung von Thesauri, Klassifikations-schemata, Taxonomien oder anderen kontrollierten Vokabularen [W3C, 2009].

Mit SKOS wird ein konzeptionelles Modell bereitgestellt, das die einfache Veröffentlichung und Kombination strukturierter und maschinenlesbarer kontrollierter Vokabulare für das [Semantische Web](#) ermöglicht. Es beinhaltet eine Zusammenstellung von [Standards](#) und [Spezifikationen](#) zur Unterstützung von Wissensorganisationssystemen („Knowledge Organisation Systems“), so dass taxonomische Information systematisch zu einer standardisierten Begriffsnavigation genutzt werden kann.

Mit SKOS wurden inzwischen zahlreiche Thesauri auch für den Bereich von Umweltinformationen entwickelt bzw. existierende in SKOS umgewandelt und veröffentlicht, z.B. AGROVOC [Rajbhandari, 2012], EARTH [Albertoni, 2010], GEMET [Eionet, 2009], REEGLE [REEEP, 2005] und UMTHESES [UMTHES, 2017].

### **Zur Aktualität semantischer Technologien**

Semantische Technologien (im weiteren Sinne) werden eingesetzt, seit es Computer gibt. Dabei geht es ganz allgemein darum, die Bedeutung von Daten durch ihre Metadaten darzustellen. Ein großen Aufschwung erlangte die Forschung zu semantischen Technologien mit der immer stärkeren Nutzung des Internet zum Ende der 1990er Jahre und mit der Etablierung des Begriffs „Semantic Web“ durch Tim Berners-Lee im Jahr 2001 [Berners-Lee, 2001].

Parallel zu den weiterhin sichtbaren, aber weniger spektakulären Forschungsaktivitäten wurden durch die enorme wirtschaftliche Bedeutung des Internet und die Konsolidierung der GAFAM-Unternehmen (Google, Apple, Facebook, Amazon, Microsoft) sowie IBM semantische Technologien verstärkt in deren Suchsystemen eingesetzt [Cardinal, 2012], [Everhart, 2012] und durch die Anwender (wenn auch

unbemerkt, d.h. implizit) genutzt. Zur Sicherung und Verbesserung der Qualität der sich immer weiter vergrößernden und konföderierenden (Zusammenschließen verschiedener) Wissensbestände ist es aus Sicht der Autoren unabdingbar, insbesondere die explizite Arbeit mit Thesauri durch geeignete Technologien und Services zu unterstützen. Dies trifft insbesondere auch für den Bereich der Umweltinformationssysteme zu, wo einschlägige Thesauri bereits seit fast 20 Jahren existieren, heute aber weiterhin gepflegt und noch mehr genutzt werden sollten.

SKOS ist seit 2009 der von W3C empfohlene Standard für die Modellierung von Thesauri. Durch seine niedrige Verständnisschwelle und gleichzeitige Universalität bietet er eine geeignete Brücke zwischen unterschiedlichen Thesauri und vereinfacht dadurch die Begriffsnavigation.

## **2 SKOS-Thesauri zur Taxonomy Governance von Umweltinformationssystemen**

SKOS-Thesauri können aus älteren Thesauri gewonnen oder ganz neu entwickelt werden.

Laut [Abecker 2011] „versprechen semantische Technologien, in Kombination mit bewährten Daten-Management-Ansätzen, einen einfachen begriffsbasierten Informationszugang, ähnlich dem, den man von Google kennt“. In [Abecker 2012] wird unter anderem gezeigt, wie SKOS-Thesauri dank des LusTRE-Prototyps im eENVplus-Framework als Interlinking-Brücken für Daten aus Umweltinformationssystemen gewinnbringend eingesetzt werden können. Dabei wurde eine effiziente semantische explorative Suche realisiert, bei der eine gezielte Datenexploration via „interlinked“ (verknüpften) Begriffe ermöglicht wird.

Beide genannte Arbeiten demonstrieren, wie Thesauri als semantische Technologien bei der Nutzung von Umweltinformationssystemen gewinnbringend eingesetzt werden können. Thesauri, aber auch Taxonomien können somit als eine wesentliche Brücke zwischen Daten und Nutzer im Bereich UIS angesehen werden.

Die Qualität von Taxonomien, bzw. Thesauri hängt u.a. davon ab, wie aktuell die darin verlinkten Begriffe gehalten werden. Hinsichtlich des Automatisierungsgrades ihrer Aktualisierung können dabei unterschiedliche Vorgehensweisen eingesetzt werden, wobei wir in automatische, semi-automatische und manuelle Aktualisierung

unterscheiden, welche die Aktualisierung von Namen, Begriffen und Relationen, sowie die Löschung, Archivierung und Schaffung neuer Konzeptschemata („Concept Schemes“, vgl. [W3C, 2009]) betrifft.

Obwohl eine automatische Anpassung auf den ersten Blick sehr erwünscht zu sein scheint („die Maschine macht alles“), können dabei eine Reihe von Anomalien (veraltetete, unvollständige oder teilwidersprüchliche Inhalte) entstehen, die, wenn nicht entdeckt, „offiziell“ geteilt und damit fehlerhaft „konsolidiert“ werden. Dies ist oft dann der Fall, wenn das anpassende System keine automatische logische Überprüfung gewährleistet.

Da Thesauri grundsätzlich von Menschen für Menschen erstellt werden, sollten auch Menschen ggf. vorgeschlagene Modifikationen freigeben, bevor diese offiziell geteilt werden. Letzteres fällt unter „semiautomatische“ Anpassung und setzt voraus, dass das Thesaurus-Pflege-System eine entsprechende Workflow-Infrastruktur zur Verfügung stellt.

Aufgrund der starken Semantikbezogenheit von Taxonomien und Thesauri ist aber die aktuell noch recht gut verwendete und verbreitete Methodik die manuelle. Hier werden im Thesaurus-Pflegesystem Namen und Relationen manuell modifiziert. Einige Systeme unterstützen eine zweistufige Methode, bei der auch nach menschlichen Modifikationen in einem Workflow in einer zweiten Phase diese Modifikationen angenommen oder abgelehnt werden können.

### **Taxonomy Governance und ein Anwendungsszenario**

Unter „Taxonomy Governance“ sollen in diesem Beitrag alle Methoden verstanden werden, die eingesetzt werden, um taxonomische Information in einem Unternehmen zu behandeln und aktuell zu halten. Diese Definition ist stark an die für „Information Governance“ [Smallwood 2014] angelehnt. Das in diesem Beitrag vorgestellte Szenario wird durch den folgenden Aspekt charakterisiert:

Unterschiedliche (SKOS-) Thesauri sind an ein Umweltinformationssystem (UIS) angebunden. Dieses UIS verwendet die offiziellen RDF-Teile dieser Thesauri („Semantic Stores“) und setzt deren Aktualität voraus (vgl. Abbildung 1).

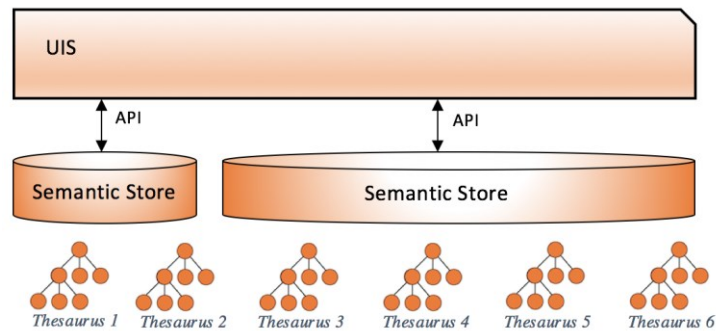


Abbildung 1: Szenario bei mehreren angebotenen RDF-Thesauri in einem UIS

Wir möchten nun zeigen, wie mit Hilfe von SKOS Shuttle [Ricci, 2016], s. Abschnitt 3, eine gut wartbare Thesaurus-Einkapselung im Sinne einer effizienten Aktualisierbarkeit und Anwendbarkeit erreicht werden kann.

Dazu werden zunächst die Thesauri (Abbildung 1, unterste Schicht) in SKOS Shuttle eingekapselt (siehe Abbildung 2) und jeder Thesaurus wird weiterhin auf seinem Semantic Store gehostet. Benutzer können dadurch die RDF-Daten viel leichter verändern und pflegen. In der untersten Schicht in der Abbildung nutzt das UIS weiterhin die RDF-Daten für die Datenverknüpfung.

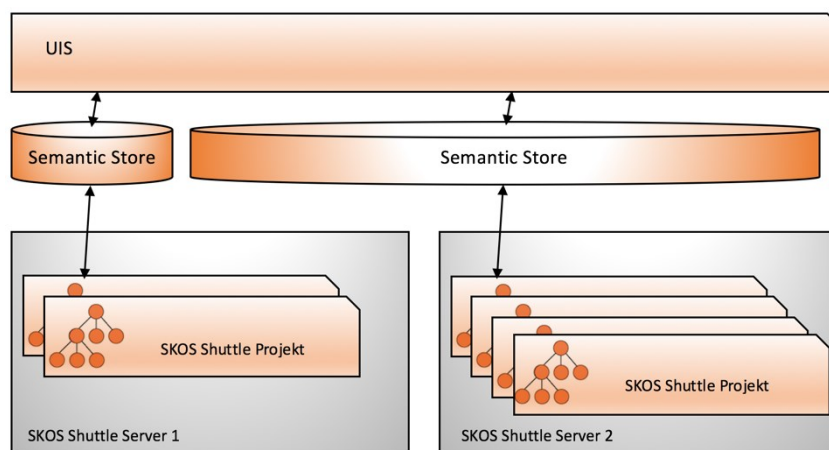


Abbildung 2: Gleiche Konstellation wie in Abbildung 1, aber unter Einsatz von SKOS Shuttle

Bei den Abbildungen 1 und 2 ist die Pflege der Thesauri (d.h. kontinuierliche Aktualisierung) noch nicht berücksichtigt.

In Abbildung 3 wird exemplarisch ein Projekt dargestellt, bei dem ein Thesaurus in SKOS Shuttle verwaltet wird. Systematiker (englisch: „Taxonomists“, wir verwenden im Folgenden das neue deutsche Wort „Taxonom“) aktualisieren ihre Projekteinhalte

(und damit ihre Thesauri) auf einem „Referenz“ Semantic-Store, z.B. via SKOS Shuttle regelmäßig. Jede Änderung wird direkt oder in vordefinierten Zeitabständen in den Semantic Stores nachgezogen, die an das UIS angeschlossen sind.

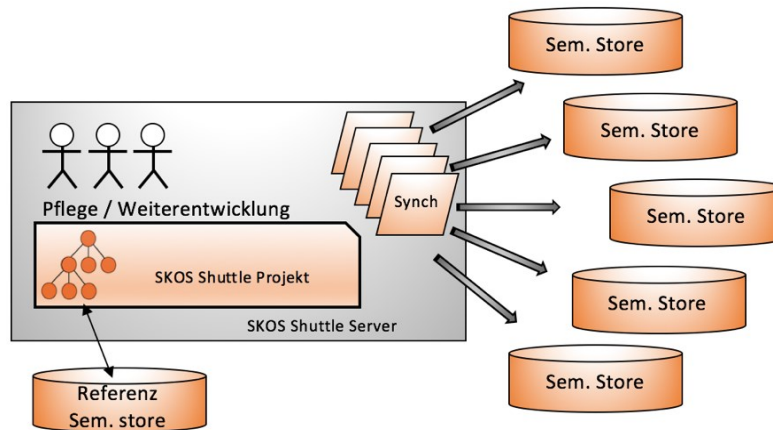


Abbildung 3: Synchronisation beliebiger Semantic Stores via Thesauripflege

Angeschlossene synchronisierte Semantic Stores spiegeln jeweils eine Replik des aktualisierten Thesaurus im Referenz-Semantic-Store wider.

Diese Methode garantiert, dass jede Thesaurusänderung im Referenz-Semantic-Store in den angeschlossenen Stores wieder zu finden ist. Als Konsequenz davon greift das UIS dann auf die aktualisierten Thesauri zu.

### Qualitätsaspekte der Synchronisation

Im dargestellten Szenario werden drei Orte des Thesauruseinsatzes angenommen: Ein Pflegeort – hier zweckmäßig auch *Veränderungsort* genannt, an dem Taxonomen Anpassungen am Thesaurus anbringen und damit RDF-Veränderungen produzieren, und zwei eher passivere Orte (z.B. eine Agentur und ein Ministerium) – zweckmäßig hier *Empfangsorte* genannt – an denen die Veränderungen empfangen und umgesetzt werden. Sofern jeweils nur eine Veränderungsquelle in einem Szenario vorhanden ist, bietet das dargestellte Synchronisationsverfahren genug Sicherheit, keine Anomalien zu erzeugen.

Will man dieses Szenario verallgemeinern und mehreren Orten die Fähigkeit der Thesaurusveränderung zuschreiben, so können unter Umständen Anomalien auftreten, die wir im Folgenden kurz skizzieren. Im Anschluss davon werden wir noch einige hinreichende qualitätserhaltende Bedingungen vorschlagen.

## Einige notwendige Definitionen

Abkürzend sprechen wir im folgenden von „**Graphen**“ (eigentlich **RDF-Graphen**) und verstehen hier darunter eine Menge von **RDF-Statements** (auch RDF-Triples genannt). Die Knoten eines Graphen sind Subjekte oder Literale, seine Kanten sind Aussagen (Statements) über jeweils ein Subjekt. Ein **Subgraph** eines Graphen ist eine Teilmenge der RDF-Statements eines Graphen.

Ein „**RDF-Delta**“ (formal: **RDF- $\delta$** ) ist ein Subgraph, der aus einer Menge zu löschender und hinzuzufügender RDF-Statements eines gegebenen Graphen besteht.

RDF-Delta ist formal wie folgt definiert :  $RDF-\delta = \lambda \cup \alpha$ , wobei

- $\lambda$ : RDF-Statements im Graph, die entfernt werden sollen
- $\alpha$ : RDF-Statements im Graph, die hinzugefügt werden sollen.

Wenn ein RDF-Delta auf einen Graphen angewendet wird, setzen wir vereinfachend voraus, dass die  $\lambda$  und  $\alpha$  zum selben Zeitpunkt gelöscht bzw. hinzugefügt werden, d.h. wir betrachten die Anwendung eines RDF-Delta als „atomar“.

Zwei RDF-Mengen  $RDF-\delta_1$  und  $RDF-\delta_2$  nennen wir **disjunkt** (formal:  $RDF-\delta_1 \cap RDF-\delta_2 = \emptyset$ ), wenn  $\lambda_1$  und  $\lambda_2$  ebenfalls disjunkt sind ( $\lambda_1 \cap \lambda_2 = \emptyset$ ), d.h.  $\lambda_1$  und  $\lambda_2$  haben kein Subjekt gemeinsam und betreffen somit diskunkte Untergraphen im selben RDF-Graphen.

Wir nennen ein RDF-Delta **auf einen Graphen anwendbar**, wenn sein  $\lambda$  zu diesem Graphen gehört. D.h. ein RDF-Delta ist auf einem Graphen genau dann anwendbar, wenn die zu ersetzenden RDF-Statements im Graph vorhanden sind. (Weitere Abschwächungen der Anwendbarkeit eines RDF-Delta sind denkbar, z.B. es könnte einfach gefordert werden, dass jedes Subjekt im  $\lambda$  bereits im Graphen sein muss. Dies kann als Konfigurationsoption angeboten werden.)



## Anwendung der Definitionen auf das vorgestellte Szenario

Im dargestellten Szenario mit einer Veränderungsquelle und zwei Empfangsorten ist ein RDF-Delta immer anwendbar. Werden keine zu ersetzenden Statements im Graphen gefunden, so ist das RDF-Delta nicht anwendbar.

Sollte AGENTUR zusätzlich eine Änderungsbefugnis erhalten, besteht jetzt die Möglichkeit, dass der Empfangsort synchrone, zeitlich überlappende oder zeitlich versetzte Veränderungen aus beiden Veränderungsquellen erhält.

Sieht man in der Praxis von einer gleichzeitigen Anwendung ab, so ergeben sich Situationen, die sich aus zeitlich überlappenden oder zeitlich versetzten Veränderungen ergeben:

Sind beide RDF- $\delta_1$  und RDF- $\delta_2$  disjunkt, so besteht keine Gefahr von Anomalien im veränderten RDF-Graphen. Im anderen Fall (beide RDF- $\delta$  beziehen sich zum Teil auf dasselbe Subjekt), wird die Anwendung des ersten RDF- $\delta_1$  die Anwendbarkeit des zeitlich versetzt bearbeiteten RDF- $\delta_2$  verhindern. Somit wird RDF- $\delta_2$  unanwendbar für den jeweiligen Empfangsort. Als Folge davon wird RDF- $\delta_2$  nicht angewendet und der jeweilige Empfangsort weist dann einen älteren Stand auf als die anderen Orte.

Somit ergibt sich folgendes Fazit: Ist mehr als ein Veränderungs-Ort im Spiel, ergeben sich an einem Empfangsort potentiell Anomalien im Zielgraphen, wenn nicht disjunkte RDF-Delta Mengen verarbeitet werden. Da in RDF-Graphen grundsätzlich widersprüchliche Statements darstellbar sind, sollte die Lösung dieser Widersprüche mittels eines geeigneten Werkzeuges – ähnlich wie bei der SVN Repository Synchronisation bei [Eclipse, 2016] unter Einbezug jedes Veränderungs-Orts durchgeführt werden können. Abbildung 4 soll Besagtes schematisch verdeutlichen.

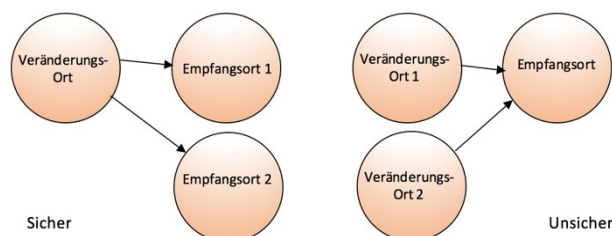


Abbildung 4: Sichere gegen unsichere Synchronisation

Anmerkungen:

- 1) Falls jeder Thesaurus in SKOS Shuttle eingekapselt ist, können Veränderungen direkt an die angeschlossenen Empfangsorte mitgeteilt werden. Dies sorgt dafür, dass die RDF-Datenbestände in den Empfangsorten zeitnah angepasst werden, was die Anwendung weiterer RDF-Deltas ermöglicht.
- 2) Das hier dargestellte hypothetische Anwendungsszenario ist bewusst stark vereinfacht. Die angestellten Überlegungen gelten auch für den realistischeren bzw. interessanteren Fall, dass mehrere Thesauri gleichzeitig verändert werden bzw. dass ein unterschiedlicher Thesaurusbestand an Veränderungsorten und Empfangsorten vorliegt.

### **3 SKOS Shuttle - Ein praktisches Beispiel für Umweltinformationssysteme (Live Demo)**

Zur Veranschaulichung der im vorangegangenen Abschnitt präsentierten Methode wird im Folgenden ein Beispiel (im Vortrag als Live-Demo) in SKOS Shuttle konfiguriert und erläutert. (Live können URLs der eingesetzten Server abweichen.)

SKOS Shuttle ist ein mandantenfähiger Web Service zum Importieren, Modifizieren, Entwickeln, Pflegen und Betreiben von SKOS-Thesauri [Ricci 2016]. Er gestattet die Definition und Verwaltung mehrerer Thesaurus-Projekte, den Datenaustausch zwischen diesen Projekten sowie unterschiedliche Arten von Migrationen in einer rollengeteilten kollaborativen Multi-User-Umgebung. Thesaurus-Projekte können in SKOS Shuttle auf folgende Best-in-Class Semantic Stores andocken (alphabetische Reihenfolge): AllegroGraph™, Blazegraph™, GraphDB™, Jena-Fuseki™, MarkLogic™, Stardog™ und Virtuoso™, sowie auf weitere SPARQL 1.1 konforme Semantic Stores [W3C, 2013]. SKOS Shuttle wurde von Semweb GmbH entwickelt, seit 2016 von verschiedenen Institutionen getestet und wird von diesen als Grundlage für ein modernes Thesaurusmanagement geschätzt.

Ein SKOS-Shuttle-Projekt kapselt einen Thesaurus in einem anzuschließenden sicheren Semantic Store (Abbildung 3) ein. Für all die Fälle, in denen kein externer Semantic Store verfügbar ist, bietet die Semweb GmbH einen eigenen Semantic Store an, dessen Namespace dann nur über einen (eindeutigen) Projektnamen – statt einer URL – eindeutig identifiziert wird (z.B. „UMWELTHESAURUS“).

In Abbildung 5 sind drei Hauptbereiche eines exemplarischen Anwendungsfalls sichtbar. Im oberen Bereich wird das (hypothetische) Projekt „UMWELTKLIMA“, seine Default-Sprache (Englisch) und die dazugehörige Semantic-Store-(SPARQL Endpoint)-URL (passwortgeschützt) definiert. In diesem Anwendungsfall wird angenommen, dass es erwünscht sei, jede Änderung im Thesaurus des Projekts „UMWELTKLIMA“ direkt mit zwei weiteren Semantic Stores (z.B. einer „UMWELTAGENTUR“, und einem „UMWELTMINISTERIUM“, zu synchronisieren. In der Live-Demo wird das Projekt „UMWELTKLIMA“ mit dem Inhalt des bekannten Umwelt-Thesaurus EARTH gefüllt und die nachgezogenen Aktualisierungen 1:1 in den angeschlossenen Semantic Stores festgelegt.

The screenshot displays the Semantic Desktop interface for configuring a project. At the top, there is a navigation bar with a search field containing 'en' and 'label or :descriptor', and a 'Welcome Demo!' button. Below this is a 'Projects' section with a filter set to 'UMWELT'. The main area shows the configuration for the 'UMWELTKLIMA' project, including its name, base URI, default language (en), and login details. Below the project configuration, there are two notification configuration sections for 'UMWELTAGENTUR' and 'UMWELTMINISTERIUM'. Each section includes a table with columns for 'Name of this swrdfpush configuration', 'Graph URI in target system', 'URL of the notification receiving service', 'User for URL', and 'Password for user'. The 'UMWELTKLIMA' project configuration includes a 'Recompute language track' button. The notification configurations include 'Notification parameters' and 'Notification output format' options.

Abbildung 5: Definition eines Thesaurus Projekts mit zwei Synchronisationen an Semantic Stores

Wenn beide „Target“ Projekte (UMWELTAGENTUR und UMWELTMINISTERIUM) ebenso jeweils als SKOS-Shuttle-Projekte definiert würden, könnte – falls erwünscht – auch die Bidirektionalität der Synchronisation gewährleistet werden. Dazu können ein einziger oder mehrere SKOS-Shuttle-Server eingerichtet werden.

Abbildung 6 zeigt eine Detailansicht des mit dem EARTH Thesaurus befüllten Projekts UMWELTKLIMA in SKOS Shuttle und Abbildung 6 die Anwendung des Autocomplete zur Findung geeigneter Begriffe startend von einem Namenpräfix (hier: Risk).

The screenshot displays the SKOS Shuttle interface for the project 'UMWELTKLIMA'. At the top, there is a navigation bar with a search field containing 'en CONDITIONS', a home icon, and a user profile 'en Welcome Fabioricciusi!'. Below this, the project information is shown: 'UMWELTKLIMA skos (en) http://linkeddata.ge.imati.cnr.it/resource/EARTH/' with a homestore and sync information. A 'DYNAMIC ASPECTS' section is visible, showing 'CONDITIONS' with a document icon. The main interface features a navigation menu with 'Details', 'Statements (edit)', 'Documents', 'SPARQL', and 'History'. A 'Language filter' is set to 'en'. The 'Preferred Labels' section shows 'CONDITIONS' in English and 'CONDIZIONI' in Italian. The 'Alternative Labels' section shows 'Situations'. The 'Broader Concept' section shows 'DYNAMIC ASPECTS'. The 'Narrower Concepts' section lists 57 concepts, including 'conditions of complex systems\*', 'acoustic comfort', 'aerobic conditions', 'aerobiosis', and 'alert'. The interface also includes options for 'Full View', 'Add language(s)', 'Remove language(s)', 'Export RDF', 'Import/Merge RDF', and 'Import/Replace RDF'. A 'Speed Cache' and 'Use default systematics' checkbox are present. An 'Auto logout in 1775 secs' message is visible in the bottom right corner.

Abbildung 6: Detailansicht eines Konzepts in SKOS Shuttle

Im oberen linken („Breadcrumb“) Teil von SKOS Shuttle werden Projektinformationen eingeblendet. Unter anderem werden konfigurierte Synchronisationen (unter „Synching to“) mit angegeben.

Die Top-Ebene eines SKOS-Shuttle-Projekts zeigt ein Profil des Thesaurus anhand seiner RDF-Daten an. Dabei werden Konzepte und Namen in einer Statistik numerisch bzw. graphisch dargestellt. Aus der Top-Ebene können Konzeptschemas im Thesaurus gepflegt und darin navigiert werden. Auf Konzeptschema-Ebene – hier nicht dargestellt – sind ein bis n Top-Konzepte zu finden, auf Top-Konzept-Ebene befinden sich reguläre Konzepte, die mittels der SKOS-Beziehungen *broader*, *narrower*, *related* die Polyhierarchie des Thesaurus bilden.

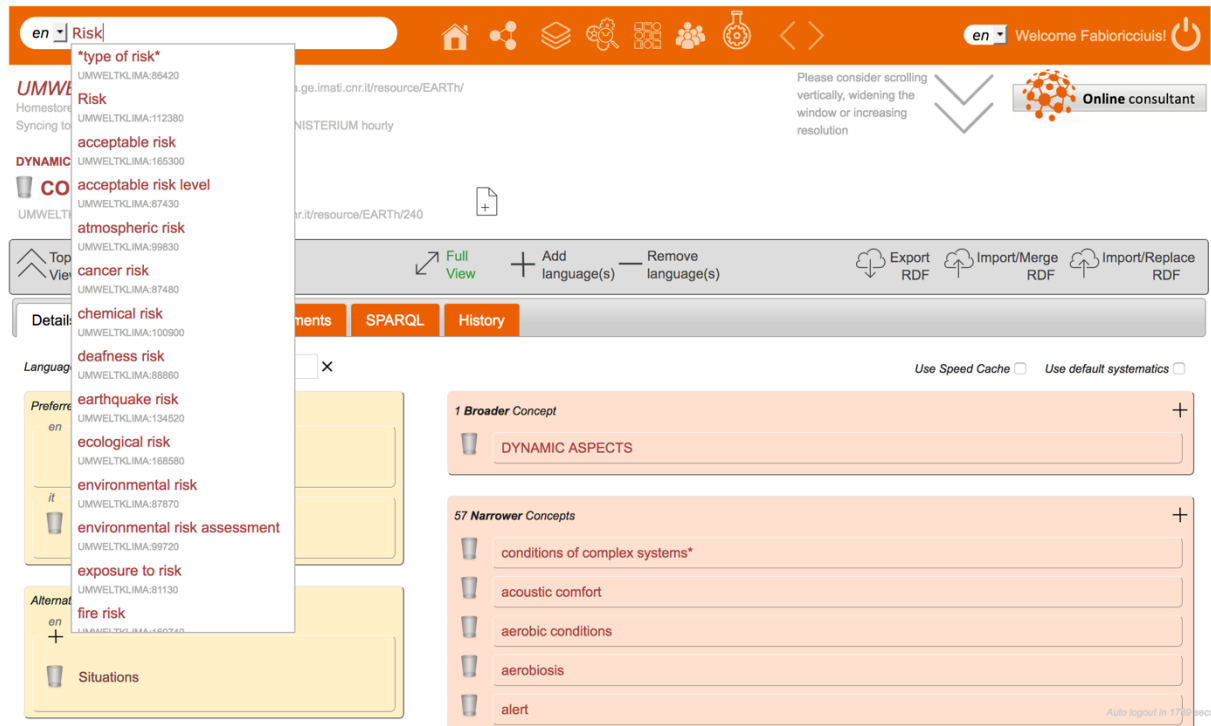


Abbildung 7: SKOS Shuttle mit Autocomplete angesetzt auf „Risk“

Nachdem Modifikationen vorgenommen wurden, speichert SKOS Shuttle sogenannte „RDF-Push-Notifikationen“ (im folgenden kurz als „RPN“ bezeichnet) in seiner lokalen Datenbank und sendet diese an die konfigurierten Semantic Stores. Jede erzeugte Notifikation kann u.a. vor bzw. nach ihrem Senden eingesehen, neu gesendet werden, wie dies auf Abbildung 7 ersichtlich ist.

Der Inhalt einer RPN besteht hauptsächlich aus einer kontextualisierten SPARQL-UPDATE-Modifikationsanfrage, die auf dem Target Semantic Store ausgeführt wird.

Durch die Möglichkeit der Angabe einer Kontextgraph-URL wird die Auswirkung der Modifikationsanfrage (z.B. Hinzufügen von RDF-Statements) an den Kontext des Target Semantic Store angepasst.

The screenshot displays the SKOS Shuttle web interface. At the top, there is a navigation bar with a language selector set to 'en', a search bar containing 'label or :descriptor', and a 'Welcome Demo!' message. Below this, a 'Projects' section shows a table of project configurations. The selected project is 'UMWELTKLIMA', with a 'Thesaurus base URI' of 'http://semweb.ch/umwelts/'. To the right of the table are buttons for 'Restore / Backup', 'Validated', and 'Users' (showing 'demouser').

Below the project table, the 'Notification configurations for UMWELTKLIMA' section is visible. It contains two configuration cards: 'UMWELTAGENTUR' and 'UMWELTMINISTERIUM'. Each card includes a form for 'Name of this swRDFPush configuration', 'Graph URI in target system', 'URL of the notification receiving service', 'User for URL', and 'Password for user'. There are also buttons for 'Test notification access' and 'Flush waiting notifications'.

On the right side of the notification configuration area, there are two log windows showing recent activity. The top log shows a 'DELETE DATA' request sent to 'http://umweltministerium.bs' at 2017-03-08 21:28:50.0, with a successful 'response: ok'. The bottom log shows a similar request sent to 'http://umweltagentur.bsp.rdf' at the same time, also with a successful 'response: ok'.

Abbildung 8: Synchronisations-Schlangen in SKOS-Shuttle zur Synchronisationsverwaltung

RPNs werden in einer relationalen Datenbank mit einem Timestamp geordnet gespeichert. Damit werden Synchronisationsschlangen gebildet. SKOS Shuttle ermöglicht die Ansicht und Verwaltung einzelner Notifikationen aus diesen Schlangen. Dabei zeigen die Synchronisationsschlangen jeweils die letzten n generierten Synchronisationsnotifikationen an. SKOS Shuttle erlaubt weiterhin das Sammeln (zeitliches Gruppieren) und Versenden von RPNs stündlich, täglich, wöchentlich oder monatlich. In dem dargestellten Beispiel wurden RPNs als „direkt“ versendbar konfiguriert. In diesem Fall geschieht die Versendung von RPNs unmittelbar nach jeder Thesaurus Modifikation seitens der ThesauruspflegerIn.

## 4 Zusammenfassung und Ausblick

Nach intensiven Forschungen und entsprechenden Ergebnissen zur Nutzung semantischer Technologien zu Beginn des dritten Jahrtausends, insbesondere zur Erstellung und Nutzung von Thesauri für Umweltinformationssysteme (UIS), scheinen diese Bemühungen in den letzten Jahren etwas erlahmt zu sein. Die Autoren dieser Arbeit sehen aber den Einsatz semantischer Technologien nach wie vor als unabdingbar an, um Metadaten in Umweltinformationssystemen unter Einbeziehung ihrer Aktualität zu verknüpfen und um linguistische Brücken zwischen Begriffen und Metadaten zu errichten.

Im vorliegenden Beitrag wurden daher zunächst wesentliche Aspekte der Taxonomy Governance und des Thesaurusmanagement erläutert und ein praktischer Fall für UIS skizziert.

Zur effektiven Nutzung von Thesauri ist es insbesondere notwendig, dass die Thesauri verschiedener verknüpfter UIS immer auf einem aktuellen Stand sind. Dazu bietet der Thesaurus Manager SKOS Shuttle neben seinen übrigen wichtigen Funktionen den zusätzlichen Service, eine oder mehrere Taxonomien in allen angeschlossenen Thesauri aktuell zu halten und stellt damit einen „Taxonomy Governance Service“ zur Verfügung, den existierende Thesaurusmanagement-systeme bisher nicht bieten.

In Anbetracht des wachsenden Bedarfs, unterschiedliche Semantic Stores gleichzeitig verwenden zu können, sieht die Roadmap von SKOS Shuttle u.a. vor, eine geeignete Grafikkomponente sowie ein Korpusmanagement zu integrieren.

## 5 Literaturverzeichnis

- Abecker A., Albertoni, R., Cipolloni C., De Martino M., Moradiafkan Y., Wössner R. (2014): Software Services exploiting the eENVplus framework of interlinked thesauri for metadata management, INSPIRE Conference Aalborg (Dänemark) [http://inspire.ec.europa.eu/events/conferences/inspire\\_2014/pdfs/19.06\\_5\\_11.00\\_Andreas\\_Abecker.pdf](http://inspire.ec.europa.eu/events/conferences/inspire_2014/pdfs/19.06_5_11.00_Andreas_Abecker.pdf) (gekürzt: <https://goo.gl/T3o4wm>) (zuletzt am 27.08.2017).
- Abecker A., Kazokos W. (2011): Semantischer Zugang zu Umwelt und Geodaten: Das Hippolytos-Projekt, INNOVATION - [http://www.disy.net/fileadmin/common/dokumente/aktuell/presse/pressespiegel/hippolytos-projekt\\_2011-05.pdf](http://www.disy.net/fileadmin/common/dokumente/aktuell/presse/pressespiegel/hippolytos-projekt_2011-05.pdf) (gekürzt: <https://goo.gl/n5YCak>) (zuletzt am 27.08.2017).
- Albertoni, R.; De Martino, M.; Di Franco, S.; De Santis, V.; Plini, P. (2010): EARTH: an Environmental Application Reference Thesaurus in the Linked Open Data Cloud. In: *Semantic Web Journal*, [http://www.semantic-web-journal.net/system/files/swj288\\_2.pdf](http://www.semantic-web-journal.net/system/files/swj288_2.pdf) (zuletzt am 27.08.2017).



- Tim Berners-Lee, James Hendler, Ora Lassila (2001): The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In: *Scientific American*, 284 (5), S. 34–43, May 2001 .
- Cardinal, David (2012): The semantic web: Can it make Google as smart as IBM Watson? <https://www.extremetech.com/computing/129243-the-semantic-web-can-it-make-google-as-smart-as-ibm-watson> (gekürzt: <https://goo.gl/13B0rN>) (zuletzt am 27.08.2017).
- Eclipse foundation 2016: [http://www.eclipse.org/org/foundation/reports/annual\\_report.php](http://www.eclipse.org/org/foundation/reports/annual_report.php) (zuletzt am 27.08.2017).
- Eionet - European Environment Information and Observation Network (2009): GEMET - GEneral Multilingual Environmental Thesaurus, <http://www.eionet.europa.eu/gemet> (zuletzt am 27.08.2017).
- Everhart, Erin (2012): How Google's Semantic Search Will Change SEO (2012) <http://mashable.com/2012/03/22/google-semantic-search-seo/#1kXVDBt2YGqM> (gekürzt: <https://goo.gl/MXUluP>) (zuletzt am 27.08.2017).
- REEEP - Renewable Energy and Energy Efficiency Partnership (2005) REEGLE - Thesaurus renewable energy and energy efficiency, <http://www.reegle.info> (zuletzt am 27.08.2017).
- Rajbhandari S., Keizer J. (2012): The AGROVOC Concept Scheme, In: *Journal of Integrative Agriculture* 2012, 11(5): S.694-699 - Food and Agriculture Organization of the United Nations, Rom, <http://www.fao.org/docrep/015/an910e/an910e00.pdf> (zuletzt am 27.08.2017).
- Ricci, F. – SKOS Shuttle – Thesaurus Management as a Service (2016), <https://skosshuttle.ch> (zuletzt am 27.08.2017).
- Smallwood R.F. (2014): Information Governance: Concepts, Strategies, and Best Practices. John Wiley & Sons, 2014.
- UMTHES (2017): <http://www.umweltbundesamt.de/tags/thesaurus> <https://sns.uba.de/umthes/de.html> (zuletzt am 27.08.2017).
- The World Wide Web Consortium (W3C) – (2001-2017): Resource Description Framework, [https://de.wikipedia.org/wiki/Resource\\_Description\\_Framework](https://de.wikipedia.org/wiki/Resource_Description_Framework) (zuletzt am 27.08.2017).
- The World Wide Web Consortium (W3C) (2009): SKOS Simple Knowledge Organization System Reference, <https://www.w3.org/TR/2009/REC-skos-reference-20090818/> (zuletzt am 27.08.2017).
- The World Wide Web Consortium (W3C) (2013): SPARQL 1.1 Query Language Recommendation. <https://www.w3.org/TR/sparql11-query/> (zuletzt am 27.08.2017).