

Ein Vergleich ausgewählter Normdatendienste

Moritz Horn <moritz.horn@gbv.de>

Jakob Voß <jakob.voss@gbv.de>

Verbundzentrale des GBV (VZG)

2017-05-31 CC-BY-SA

DOI [10.5281/zenodo.800577](https://doi.org/10.5281/zenodo.800577)

Dieser Report fasst Ergebnisse eines Vergleichs von APIs zum Zugriff auf ausgewählte Wissensorganisationssysteme zusammen.

1 Einführung

Im Rahmen des Projekt [coli-conc](#) wurden in einem Workshop der DINI AG KIM allgemeine Anforderungen an Normdatendienste zum Zugriff auf verschiedene Wissensorganisationssysteme ermittelt (Voß 2016a). Der vorliegende Report untersucht die Eigenschaften einiger vorhandener Normdatendienste. Die Untersuchung basiert auf den Ergebnissen eines umfangreicheren Vergleichs von APIs zum Zugriff auf kontrollierte Vokabulare (Horn 2016).

2 Untersuchte Normdatendienste

Untersucht wurden vier Schnittstellen zum Zugriff auf die folgenden Vokabulare:

Vokabular	Typ	API-Betreiber
Regensburger Verbundklassifikation (RVK)	Klassifikation	Uni Regensburg ¹
Gemeinsame Normdatei (GND)	Normdatei	hbz (lobid.org) ²
Standardthesaurus Wirtschaft (STW)	Thesaurus	ZBW ³
Library of Congress Subject Headings (LCSH)	Thesaurus	LoC ⁴

¹http://rvk.uni-regensburg.de/Portal_API/ (nicht dokumentierter Endpunkt)

²<http://lobid.org/api>

³<http://zbw.eu/labs/de/project/econ-ws> (Sammlung mehrerer APIs)

⁴<http://id.loc.gov/authorities/subjects.html>

3 Rückgabeformate

Abgesehen von der RVK-API, die ein eigenes JSON- bzw. XML-Format verwendet, basieren die Rückgabeformate aller APIs auf RDF, das in unterschiedlichen Serialisierungen (RDF/XML, Turtle...) ausgewählt werden kann. Die RDF-Formate basieren bei STW und LCSH auf SKOS (Miles und Bechhofer 2009), während für die GND eine eigene Ontologie definiert wurde (Haffner 2016). Das Format der STW beschränkt sich weitgehend auf SKOS. Die Library of Congress bietet mit MADS/RDF zusätzlich eine Erweiterung an.¹ Noch umfangreicher ist die GND-Ontologie mit 228 Eigenschaften.

¹<http://www.loc.gov/standards/mads/rdf/>

4 Umfang der zurückgelieferten Daten

Trotz unterschiedlicher Rückgabeformate lassen sich Gemeinsamkeiten in den über die verschiedenen APIs abfragbaren Datenfeldern ausmachen. Die folgende Übersicht beschränkt sich auf einen Vergleich der abfragbaren Verknüpfungen: Dies sind hierarchische (broader, narrower, ancestors) und assoziative Verknüpfungen (related) zwischen Begriffen eines Vokabulars sowie mappings auf andere Vokabulare.

Feld	RVK	GND	STW	LCSH
broader	X	X	X	X
narrower	X	-	X	X
ancestors	X	-	-	-
related	-	X	X	X
mappings	-	X	X	X

Die RVK-API bietet als einzige der untersuchten Dienste eine Möglichkeit alle transitiv übergeordneten Klassen (ancestors) einer Klasse abzufragen, was allerdings bei monohierarchischen Klassifikationen auch einfacher umzusetzen ist. Allerdings wird bei Auswahl einer Klasse nur mitgeteilt, ob untergeordneten Klassen vorhanden sind, die in diesem Fall mit einer weiteren Abfrage ermittelt werden können. In der GND fehlen untergeordnete Begriffe ganz.

Die Einträge von STW zu GND und LCSH zu GND sind jeweils über Mappings miteinander verknüpft. Zudem enthält die LCSH Mappings auf die Normdatei RAMEAU der französischen Nationalbibliothek. In der GND sind diese Mappings zwar auch enthalten, werden aber nicht über die API mitgeliefert. Die GND enthält dafür Verknüpfungen zu den eigenen GND Subject Categories und der DDC. Sowohl GND als auch STW sind weiterhin (über de.wikipedia bzw. DBPedia) mit Wikipedia verknüpft¹). In den STW-Daten sind zusätzlich Verlinkungen auf Einträge im Katalog des ZBW enthalten.

¹Inzwischen wäre ein Mapping mit Wikidata sinnvoller.

5 Abfragemöglichkeiten für typische Anwendungsfälle

Die folgende Tabelle liefert einen Vergleich der APIs mit den von Voß (2016a) genannten Anwendungsfällen und Anforderungen mit Ausnahme von Schreib-Zugriffen.

Anforderung	RVK	GND	STW	LCSH
Abfrage von Synonymen	X	-	X	-
(Fuzzy-)Suche in Normdateien	X	X	(X)	-
Metasuche in mehreren Normdateien	-	-	-	X
Multilinguale Suche	-	-	X	-
Autovervollständigung/Typeahead	X	X	X	-
Entity-Kurzinfo	(X)	X	(X)	X
Abfrage der kanonischen Form	-	X	(X)	(X)
Abfrage von Konkordanzen und Mappings	-	-	X	-
Abfrage von Änderungen	-	-	X	(X)
Nutzungsstatistik	-	-	-	-
Ermittlung von indexierten Ressourcen	-	-	(X)	-

Einige Anwendungsfälle können mit den vorhandenen APIs nicht oder nur eingeschränkt (z.B. indirekt über andere Abfragen) umgesetzt werden. Die APIs der STW bieten den weitesten Funktionsumfang, was allerdings auch mit der Komplexität des Vokabulars zusammenhängt. Die API der LCSH kann mit ihrer Beschränkung auf einfachem Linked Open Data, d.h. Abfrage per bekannter URI, den wenigsten Anforderungen genügen. Die Typeahead-Funktionen sind bei allen APIs mangelhaft, da sie anscheinend nicht auf Ranking-Verfahren basieren.

6 Zusammenfassung und Ausblick

Zusammenfassend lässt sich feststellen dass die verschiedenen untersuchten Normdatendienste trotz grundsätzlicher Vergleichbarkeit sehr unterschiedlich ausfallen. Selbst die RDF-basierten APIs sind uneinheitlich umgesetzt. Generell lassen sich mit SKOS zwar Ansätze einer Vereinheitlichung erkennen, solange aber SKOS-Erweiterungen, Umfang der zurückgelieferten Daten und Abfragemöglichkeiten variieren, ist eine einheitliche Abfrage verschiedener Normdatendienste aber nicht möglich.

Im Rahmen des Projekt coli-conc wird deshalb neben der Vereinheitlichung von SKOS-Varianten mit dem Format JSKOS (Voß 2016d) eine darauf basierende API zur Abfrage von Wissensorganisationssystemen entwickelt. Eine Teilmenge dieser JSKOS-API wurde bereits unter dem Namen ELMA (Voß 2016c) festgelegt und umgesetzt.

Zur weiteren Entwicklung und Evaluation von Normdatendiensten sollen im Verlauf des Projektes neben Normdatendiensten für einzelne Vokabulare die APIs verschiedener Webanwendungen zum Zugriff auf Wissensorganisationssysteme verglichen werden (Voß 2016b).

Nachweise

Haffner, Alexander. 2016. „GND Ontology“. DNB. <http://d-nb.info/standards/elementset/gnd>.

Horn, Moritz. 2016. „Vergleich von webbasierten Programmierschnittstellen zum Zugriff auf kontrollierte Vokabulare“. Bachelorarbeit. Hochschule Hannover. <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bsz:960-opus4-10538>.

Miles, Alistair, und Sean Bechhofer, Hrsg. 2009. „SKOS Simple Knowledge Organization System Reference“. W3C Recommendation. <http://www.w3.org/TR/skos-reference>.

Voß, Jakob. 2016a. „Anforderungen an Normdatendienste“. coli-conc technical report 4. doi:[10.5281/zenodo.50180](https://doi.org/10.5281/zenodo.50180).

———. 2016b. „Open Source web applications for Knowledge Organization Systems“. coli-conc technical report 7. doi:[10.5281/zenodo.61262](https://doi.org/10.5281/zenodo.61262).

———. 2016c. „Entity Lookup Microservice API (ELMA)“. Version 0.0.4. <http://gbv.github.io/elma/>.

———. 2016d. „JSKOS data format for Knowledge Organization Systems“. Version 0.1.2. <http://gbv.github.io/jskos/>.