

**Optimierung des thematischen Zugangs zu rechtlichen Ressourcen:
EuroVoc und VocBench.
Vorteile des Einsatzes von mehrsprachigen kontrollierten Vokabularen
und eines quelloffenen kollaborativen Tools zu ihrer Pflege**

German translation of the original paper: “Optimizing subject access to legal resources: EuroVoc and VocBench. Benefits of using multilingual controlled vocabularies and an open source collaborative tool for their maintenance”.

Translated by Directorate-General for Translation of the European Commission.

The text of this document has been translated into German and differences from the original text may occur. This translation is provided for reference purposes only.

Rudolf W. Strohmeier

Amt für Veröffentlichungen der Europäischen Union, Luxemburg

Willem van Gemert

Amt für Veröffentlichungen der Europäischen Union, Luxemburg

Christine Laaboudi

Amt für Veröffentlichungen der Europäischen Union, Luxemburg

Anne Waniart

Amt für Veröffentlichungen der Europäischen Union, Luxemburg



This is a German translation of “Optimizing subject access to legal resources: EuroVoc and VocBench. Benefits of using multilingual controlled vocabularies and an open source collaborative tool for their maintenance” copyright © 2017 by the European Union. This work is made available under the terms of the Creative Commons Attribution 4.0 Unported License: <http://creativecommons.org/licenses/by/4.0>

Die in diesem Dokument vertretenen Ansichten sind die des Autors und spiegeln nicht notwendigerweise den Standpunkt der Europäischen Kommission oder einer anderen Organisation wider.

ZUSAMMENFASSUNG:

Das Amt für Veröffentlichungen der Europäischen Union ist ein interinstitutionelles Amt mit der Aufgabe, Rechtsvorschriften und Publikationen der Europäischen Union in den 24 Amtssprachen der EU zu veröffentlichen. Um der Herausforderung der mehrsprachigen Veröffentlichung rechtlicher Daten gerecht zu werden, greift das Amt für Veröffentlichungen auf mehrsprachige kontrollierte Vokabulare wie Referenztabellen und den EuroVoc-Thesaurus zurück.

EuroVoc ist ein mehrsprachiger multidisziplinärer Thesaurus zu den Tätigkeitsbereichen der EU-Organe. Er beruht auf den von der Organisation W3C (World Wide Web Consortium) empfohlenen

Standards für das semantische Web. Die genannten Technologien ermöglichen nicht nur die Modellierung von Flexibilität und der Fähigkeit zu schlussfolgern, sondern erleichtern auch die Anpassung von EuroVoc an andere, etwa nationale oder themenbereichsspezifische Thesauri und kontrollierte Vokabulare.

Die Erstellung, Pflege und Verwaltung des Thesaurus EuroVoc erfolgt mithilfe von VocBench, einer quelloffenen kollaborativen Webplattform zur Entwicklung von Thesauri, die den Standards SKOS und SKOS-XL entspricht. Dank der Finanzierung durch das Programm ISA² der Europäischen Kommission wurde eine neue Hauptversion von VocBench entwickelt, die im dritten Quartal 2017 zur Verfügung stehen wird.

Schlüsselwörter: kontrollierte Vokabulare, Thesaurus, thematischer Zugang, Standards des semantischen Webs

Über das Amt für Veröffentlichungen der Europäischen Union

Das Amt für Veröffentlichungen der Europäischen Union ist ein 1969 von den Organen der EU geschaffenes interinstitutionelles Amt mit der Aufgabe, Rechtsvorschriften und Publikationen der Europäischen Union in allen 24 Amtssprachen der EU zu veröffentlichen. Es trägt dadurch zur praktischen Umsetzung des Mottos der Europäischen Union „In Vielfalt geeint“ bei, welches zum Ausdruck bringt, dass sich die Europäer in der EU zusammengeschlossen haben, um sich gemeinsam für Frieden und Wohlstand einzusetzen, und dass gleichzeitig die vielen verschiedenen europäischen Kulturen, Traditionen und Sprachen den gesamten Kontinent bereichern.

In diesem Zusammenhang bietet das Amt für Veröffentlichungen als moderner mehrsprachiger Informationsvermittler allen Organen, Einrichtungen und sonstigen Stellen der EU sowie allgemein ihren Bürgern Informationsmanagementdienste auf dem neuesten Stand an. Als etablierter und anerkannter Akteur im Bereich der Standardisierung hat das Amt für Veröffentlichungen auch eine führende Rolle bei der Standardisierung von Metadaten und Formaten zum Austausch von Informationen, darunter auch Daten im Bereich der juristischen Dokumentation, übernommen. Das Amt für Veröffentlichungen verwaltet zudem eine Reihe von Informationsmanagement-Instrumenten, die von den Organen, Einrichtungen und sonstigen Stellen der EU, von Verwaltungen auf nationaler und regionaler Ebene in Europa sowie von nationalen Regierungen und privaten Nutzern weltweit verwendet werden.

Ein wesentlicher Teil der vom Amt für Veröffentlichungen publizierten Informationen steht in den 24 Amtssprachen der Europäischen Union zur Verfügung, dies gilt etwa für alle auf EUR-Lex¹, dem Portal für den Zugang zum EU-Recht, veröffentlichten Rechtsakte sowie für Teile der online über den EU Bookshop² verbreiteten EU-Veröffentlichungen. Der Aspekt der Mehrsprachigkeit birgt eine Reihe von Herausforderungen für die Produktion des Amtes für Veröffentlichungen im Hinblick auf die Optimierung des thematischen Zugangs. Einer der vom Amt für Veröffentlichungen umgesetzten Lösungsansätze zur Bewältigung dieser Herausforderungen ist der Einsatz von mehrsprachigen kontrollierten Vokabularen mit sprachunabhängigen Identifikatoren. Diese Vokabulare finden sich im Metadatenregister des

¹ <http://eur-lex.europa.eu/>

² <http://publications.europa.eu/bookshop>

Amtes für Veröffentlichungen³. Alle diese Instrumente werden als offene Daten bereitgestellt, die von allen interessierten Parteien in den Mitgliedstaaten in der Europäischen Union und außerhalb der EU wiederverwendet werden können. Indem das Amt für Veröffentlichungen solche Instrumente bereitstellt, fördert es das gegenseitige Verständnis und die Interoperabilität voneinander abweichender Rechtssysteme und trägt so auf diesem Gebiet zur Schaffung eines gemeinsamen europäischen Raumes der Zusammenarbeit und des voneinander Lernens bei.

EuroVoc

EuroVoc ist das nützlichste kontrollierte Vokabular für die Optimierung des thematischen Zugangs zu Daten im Bereich der juristischen Dokumentation auf EU- und nationaler Ebene sowie für sonstige Informationsquellen. Es handelt sich um einen multidisziplinären Thesaurus, der rund 8000 Konzepte in 23 Amtssprachen der EU sowie in drei weiteren Sprachen von Beitrittskandidaten bietet. Er deckt alle Tätigkeitsbereiche der Organe der Europäischen Union ab. Die Begriffe sind nach den Politik- und Tätigkeitsbereichen in 21 Bereiche und 120 Mikrothesauri (Unterbereiche) eingeteilt. EuroVoc wird auf zwei Gebieten zur Indexierung von Inhalten eingesetzt: bei EU-Rechtsvorschriften in EUR-Lex und bei den vom Amt für Veröffentlichungen über den EU Bookshop und OPac⁴ (den öffentlich zugänglichen Online-Katalog des Amtes für Veröffentlichungen) verbreiteten Publikationen. Da die Konzepte in EuroVoc sprachunabhängig sind, können die Nutzer ein Konzept in ihrer eigenen Sprache suchen und die Ergebnisse in einer anderen Sprache erhalten.

EuroVoc wird nach der Norm ISO 25964 *Thesauri und die Interoperabilität mit anderen Vokabularen* der Internationalen Normungsorganisation erstellt. Es wird kontinuierlich angepasst, damit es mit der Entwicklung von drei Paradigmen Schritt hält: der Bereiche, in denen die Organe der EU betroffen sind, der Sprachenregelungen und der technischen Innovationen. Das Wartungsteam des Amtes für Veröffentlichungen nimmt die Vorschläge der nationalen Parlamente, der Organe und Einrichtungen der EU sowie von privaten Nutzern (Bürgern oder Unternehmen) entgegen und prüft sie. Für die Verwaltung und Pflege des Thesaurus wurde ein Governance-Verfahren eingerichtet, wobei die Aktualisierungen von EuroVoc im Prinzip zweimal jährlich veröffentlicht werden.

Die Website von EuroVoc bietet eine Browsing- und Suchfunktion (<http://eurovoc.europa.eu>). Sie kann außerdem vom Offenen Datenportal der EU (<https://open-data.europa.eu/en/data/dataset/eurovoc>) in maschinenlesbaren Formaten wie XML und SKOS/RDF heruntergeladen werden. Darüber hinaus können die Nutzer mithilfe einer Reihe von Webdiensten Anfragen an EuroVoc stellen, deren Ergebnisse in ihren Anwendungen direkt angezeigt oder verwendet werden können. <http://eurovoc.europa.eu/drupal/?q=webservice&cl=de>. Das Amt für Veröffentlichungen ist ein Vorreiter bei der Verwendung semantischer Technologien in seinen Produktions- und Verbreitungsketten. EuroVoc ist über den SPARQL-Endpunkt des Amtes für Veröffentlichungen erhältlich und kann zusammen mit sonstigen vom Amt für Veröffentlichungen zur Verfügung gestellten rechtlichen Ressourcen zur Wiederverwendung und Verknüpfung von Daten eingesetzt werden.

EuroVoc, das zum Standardtool zur Indexierung von EU-Informationen geworden ist, wird von den Organen, Einrichtungen und sonstigen Stellen der EU, nationalen und regionalen

³ <http://publications.europa.eu/mdr/>

⁴ <http://opac.publications.europa.eu/search/query?theme=system>

Parlamenten und Regierungen in Europa, öffentlichen Verwaltungen, Bibliotheken, Informationsvermittlern und Privatnutzern auf der ganzen Welt verwendet.

Auf dem Weg zum semantischen Web

Linked Open Data (LOD) wird in der Welt der Datenveröffentlichung zunehmend zu einem de-facto-Standard und gibt Verfahren vor. Dies liegt vor allem an einem umfangreichen Satz von Standards (RDF-S, OWL usw.), weithin anerkannten Ontologien (Dublin Core, DCAT, SKOS (-XL), schema.org usw.) sowie an einer starken Einbeziehung der Anwendergemeinschaft.

EuroVoc ist eines der ersten mehrsprachigen Vokabulare, das Technologien des semantischen Webs einsetzt, in denen die W3C-Empfehlungen sowie die neuesten Entwicklungen bei Thesaurusnormen Berücksichtigung finden. Ursprünglich basierte die EuroVoc-Ontologie auf SKOS (Simple Knowledge Organization System) und SKOS-XL (SKOS extension for Labels). Nach und nach entwickelte sich die Inhaltsinformation über die einfachen SKOS- und SKOS-XL-Modelle hinaus zu einem neuen Modell, das die Erfordernisse des Amtes für Veröffentlichungen widerspiegelte. Um Interoperabilität zu ermöglichen und die breite Verwendbarkeit der veröffentlichten Daten zu fördern, entwickelte das Amt für Veröffentlichungen ein Anwendungsprofil (SKOS-AP-EU) zur Gestaltung von Thesauri und kontrollierten Schlagwortlisten.⁵ Das SKOS-AP-EU-Anwendungsprofil erweitert SKOS um Eigenschaften einer Reihe bekannter Vokabulare. Dieses Anwendungsprofil findet seinen formalen Ausdruck in der Verwendung von SHACL (shape constraint language). Es ermöglicht die automatische Validierung von kontrollierten Vokabularen und die zeitgleiche Erzeugung von Dokumentation.

EuroVoc als verknüpfte Daten

Mapping zwischen EuroVoc und stärker spezialisierten kontrollierten Vokabularen

Die Tatsache, dass EuroVoc in semantischen Formaten wie SKOS vorliegt, bietet eine hervorragende Gelegenheit zum Mapping und Abgleich mit stärker spezialisierten Vokabularen.

EuroVoc erfasst eine Reihe unterschiedlicher Disziplinen auf allgemeiner Ebene und kann daher nicht immer so spezifisch sein, wie es von Experten auf einem bestimmten Gebiet benötigt wird. Um eine breitere Abdeckung zu gewährleisten, wird EuroVoc mit stärker spezialisierten Thesauri wie dem Thesaurus der Ernährungs- und Landwirtschaftsorganisation (AGROVOC), den Thesauri der Vereinten Nationen (UNBIS und UNESCO), dem Standard-Thesaurus Wirtschaft (STW) des Leibniz-Informationszentrums Wirtschaft usw. abgeglichen. Die Leibniz-Stiftung wird die mit dem STW gemappten EuroVoc-Label integrieren, um ihre Suchmaschine zu bereichern.

GEMET, der allgemeine mehrsprachige Umwelt-Thesaurus (General Multilingual Environmental Thesaurus), wurde als Indexierungstool für die Europäische Umweltagentur (EUA)⁶ entwickelt. Der Mehrwert des semantischen Mappings von GEMET mit EuroVoc wird in dem erleichterten Zugang zu Dokumenten im Bereich Umwelt, einem Haupttätigkeitsfeld

⁵ Quelle: Costetchi E., Van Gemert W. (2016).

⁶ https://europa.eu/european-union/about-eu/agencies/eea_de

der EU, bestehen. Beide Thesauri sind mehrsprachig (23 Sprachen für EuroVoc, 35 für GEMET) und unterscheiden sich im Hinblick auf Abstraktionsgrad und Detailliertheit. Beispiel: GEMET verwendet das Konzept „environmental legislation“⁷, die semantische Entsprechung in EuroVoc ist „environmental law“⁸. Durch das Mapping der zwei Thesauri könnten die Nutzer bekannte Terminologie aus einem der beiden Thesauri in ihren Muttersprachen kombinieren und sich vom System die Entsprechungen mit anderen Terminologien aufzeigen lassen, um äquivalente Ergebnisse aufzufinden.

Mapping von EuroVoc mit nationalen kontrollierten Vokabularen, um nationale rechtliche Informationen zu annotieren

EuroVoc erfasst eine Reihe unterschiedlicher Disziplinen auf allgemeiner Ebene und kann daher die Unterschiede zwischen verschiedenen Rechtssystemen nicht ausreichend detailliert widerspiegeln. Durch die Verknüpfung von EuroVoc mit den kontrollierten Vokabularen der einzelnen Rechtssysteme können jedoch die allgemeinen rechtlichen Konzepte oder die EU-Konzepte von EuroVoc auf die in der nationalen Gesetzgebung eines gegebenen Landes verwendeten Termini ausgedehnt werden. Im Ergebnis können beide Vokabulare verwendet werden, um den thematischen Zugang entweder zum nationalen oder zum EU-Recht (über EUR-Lex) zu optimieren. Ein weiterer Vorteil des Thesauriabgleichs ist die Ausweitung der Suchstrategie auf andere nationale oder europäische Rechtsdatenbanken.

Die Vorteile des Abgleichs des nationalen Vokabulars mit EuroVoc können folgendermaßen zusammengefasst werden: Hat ein Konzept eines Rechtssystems ein Äquivalent in EuroVoc, wird es automatisch übersetzt. Ist das nationale Konzept zu spezifisch, kann ein Mapping mit einem allgemeineren EuroVoc-Konzept erfolgen. Beispiel: Die Verwaltung des Flusses Mosel betrifft drei Länder (Deutschland, Frankreich und Luxemburg). Auf Legilux, der offiziellen Website zum luxemburgischen Recht, wird das Konzept „canalisation de la Moselle“⁹ verwendet. EuroVoc beschreibt die Geografie der Mitgliedstaaten nur auf regionaler Ebene und hat ein Äquivalent für dieses sehr spezifische Konzept. Durch Abgleich des spezifischen Terminus „Kanalisation der Mosel“ mit dem bestehenden EuroVoc-Konzept „[Binnenwasserstraße](#)“ könnte die Suchstrategie ausgeweitet werden, um nationale und/oder europäische Rechtsdaten besser vergleichen zu können.

Die Mitgliedstaaten arbeiten offiziell untereinander und mit der Europäischen Union zur Indexierung von Daten der nationalen Regierungen und Parlamente mithilfe von EuroVoc zusammen. Beispielsweise hat der Nationalrat der Slowakischen Republik¹⁰ Eurovoc im Rahmen seines parlamentarischen Informationssystems eingeführt. Die Nationalversammlung der Republik Slowenien¹¹ verwendet EuroVoc als Grundlage für die Codeliste des EU-Portals der Regierung. Das Parlament und der Senat in Belgien¹² setzen EuroVoc als kontrolliertes Vokabular zur Suche in ihren jeweiligen Datenbanken ein.

⁷ <http://www.eionet.europa.eu/gemet/concept?ns=1&cp=2862>

⁸ <http://eurovoc.europa.eu/535>

⁹ <http://legilux.public.lu/search/A/?fulltext=&thematique=canalisation+de+la+moselle>

¹⁰ <http://kniznica.nrsr.sk/index.php/en/eurovoc>

¹¹ http://www.mzz.gov.si/en/foreign_policy_and_international_law/eu_policies/language_issues/legal_and_linguistic_databases/eurovoc/

¹² <http://www.dekamer.be/kvvcr/showpage.cfm?section=/flwb&language=nl&cfm=ListKeyword.cfm?legislat=54> und https://www.senate.be/www/?MIval=/index_senate&MENUID=12000&LANG=fr

Ein weiteres Beispiel einer Zusammenarbeit dieser Art ist ein Projekt im Zusammenhang mit dem [European Legislation Identifier](#) (ELI) unter Führung des Amtsblattes des Großherzogtums Luxemburg und des Amtes für Veröffentlichungen. Im Rahmen dieser Initiative wurden die Konzepte in EuroVoc mit dem kontrollierten Vokabular zur Indexierung des Inhalts der auf <http://legilux.public.lu/> veröffentlichten nationalen Gesetzgebung des Großherzogtums Luxemburg abgeglichen. Ziel des Projekts war es, die europäischen und die nationalen Rechtsvorschriften Luxemburgs interoperabel zu machen und den Zugang zu rechtlichen Ressourcen zu erleichtern. Die Ergebnisse bestätigen, dass der Einsatz von EuroVoc als Brücke zwischen den stärker spezialisierten kontrollierten Vokabularen der nationalen Rechtsordnungen deutliche Vorteile bei der Verbesserung des Zugangs zur Gesetzgebung bringt.

VocBench

Das Amt für Veröffentlichungen setzt zur Pflege von EuroVoc *VocBench* ein. *VocBench* ist eine webgestützte, quelloffene kollaborative Plattform für das Management von mehrsprachigen kontrollierten Vokabularen; sie setzt semantische Technologien ein und entspricht den SKOS und SKOS-XL-Standards. Sie ist besonders geeignet für die Verwaltung großer Thesauri im RDF-Format.

Das Amt für Veröffentlichungen ist aktives Mitglied der großen, mehrsprachigen Thesauri-Gemeinschaft und treibt die Entwicklung einer neuen Version von *VocBench* voran. Die neueste Hauptversion *VocBench 3.0*, entwickelt dank der Finanzierung durch das [Programm ISA²](#) der Europäischen Kommission, enthält eine Reihe neuer Funktionen, darunter integrierte Abgleichsfunktionen. *VocBench 3.0*, das im dritten Quartal 2017 zur Verfügung stehen wird, wird die notwendigen Instrumentarien enthalten, damit Nutzer ihre kontrollierten Vokabulare, Metadaten oder Glossare als Linked Open Data erstellen, pflegen, verknüpfen und veröffentlichen können. Es wird eine kollaborative Mehrzweckumgebung für die Entwicklung von RDF-Datensätzen aller Art bieten und die Bearbeitungsmöglichkeiten gegenüber der Vorgängerversion verbessern, deren besondere, für ihren Erfolg ausschlaggebende Eigenschaften jedoch beibehalten.

VocBench ist quelloffen und kann von allen öffentlichen Organisationen, allen Unternehmen und allen unabhängigen Nutzern zur Verwaltung ihrer kontrollierten Vokabulare eingesetzt werden. Das neueste Installationspaket sowie die Aufbau und Einsatzanweisungen für *VocBench 2.x/3.0* finden sich direkt auf der [Website des Projekts](#).

Fazit

Die Gewährleistung eines optimierten thematischen Zugangs zu rechtlichen Daten aus den verschiedensten Quellen, etwa den Organen der EU und den Stellen, die die Rechtsvorschriften der EU-Mitgliedstaaten veröffentlichen, ist ein wesentlicher Bestandteil des mehrsprachigen „Ökosystems“ für juristische Informationen in der Europäischen Union, das erfolgreich die Einheit unter seinen verschiedenen Mitgliedern fördert. Das Amt für Veröffentlichungen geht noch weiter, indem es offene Standards verwendet und die in diesem Artikel vorgestellten einschlägigen Instrumente zur weltweiten Weiterverwendung zur Verfügung stellt. Es setzt sich damit für die Idee ein, dass Solidarität in Verschiedenheit eine echte, wertvolle Option ist, die denjenigen, die bereit sind, sie umzusetzen, beträchtliche Vorteile bringen kann.

Literaturangaben

Costetchi E., Van Gemert W. (2016). Towards executable application profile for European vocabularies. W3C. https://www.w3.org/2016/11/sdsvoc/SDSVoc16_paper_23.

Amt für Veröffentlichungen der Europäischen Union, 2010. EuroVoc-Konferenz: Mind the Lexical Gap. EU. Amt für Veröffentlichungen. <http://eurovoc.europa.eu/drupal/?q=node/936>.

Stellato A. u. a. (2015). VocBench: A Web Application for Collaborative Development of Multilingual Thesauri. The Semantic Web. Latest Advances and New Domains. ESWC 2015. Lecture Notes in Computer Science, Nr. 9088. Springer, Cham. http://link.springer.com/chapter/10.1007/978-3-319-25639-9_29.