

Optimiser l'accès aux ressources juridiques: EuroVoc et VocBench. Avantages liés à l'utilisation de vocabulaires multilingues contrôlés ainsi que de l'outil de collaboration open source nécessaire à leur maintenance

French 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 French and differences from the original text may occur. This translation is provided for reference purposes only.

Rudolf W. Strohmeier

Office des publications de l'Union européenne, Luxembourg.

Willem van Gemert

Office des publications de l'Union européenne, Luxembourg.

Christine Laaboudi

Office des publications de l'Union européenne, Luxembourg.

Anne Waniart

Office des publications de l'Union européenne, Luxembourg.



This is a French 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>

Les opinions exprimées dans le présent document sont ceux de leurs auteurs et ne reflètent pas nécessairement les avis de la Commission européenne ni de tout autre organisme.

RESUME:

L'Office des publications de l'Union européenne est un service interinstitutionnel chargé de diffuser la législation et les publications de l'Union européenne dans les 24 langues officielles de l'UE. Pour relever le défi consistant à publier des données légales dans des versions multilingues, l'Office des publications utilise des vocabulaires contrôlés multilingues tels que des tables d'autorité et le thésaurus EuroVoc.

EuroVoc est un thésaurus multilingue et pluridisciplinaire couvrant les domaines d'activité des institutions européennes. Il est élaboré en utilisant les normes du web sémantique recommandées par

le W3C. Outre la flexibilité de modélisation et les capacités de raisonnement, ces technologies facilitent également l'alignement d'EuroVoc avec d'autres thésaurus et vocabulaires contrôlés, tels que les vocabulaires nationaux et thématiques.

L'élaboration, la maintenance et la gestion du thésaurus EuroVoc sont réalisés en utilisant VocBench, une plateforme web collaborative et open source pour le développement de thésaurus, respectant les normes SKOS et SKOS-XL. Grâce au financement par le programme ISA² de la Commission européenne, une nouvelle version majeure VocBench a été élaborée (elle sera disponible au 3^{ème} trimestre 2017).

Mots-clés: Vocabulaires contrôlés, thésaurus, accès au sujet, normes web sémantiques

À propos de l'Office des publications officielles de l'Union européenne

L'Office des publications de l'Union européenne est un service interinstitutionnel créé en 1969 par les institutions européennes et chargé de diffuser la législation et les publications de l'Union européenne dans les 24 langues officielles de l'UE. Ce faisant, l'Office contribue à la mise en œuvre de la devise de l'Union européenne «Unie dans la diversité», qui rappelle que les Européens se sont rassemblés pour œuvrer en faveur de la paix et de la prospérité, tout en jouissant de la diversité de cultures, de traditions et de langues qu'offre l'Europe.

Dans ce contexte, en tant que prestataire moderne d'informations multilingues, l'Office des publications propose des services de pointe en matière de gestion de l'information à l'ensemble des institutions, des agences et des organismes de l'UE, ainsi qu'aux citoyens au sens large. En tant qu'acteur reconnu de la normalisation, l'Office joue également un rôle prépondérant dans la normalisation des métadonnées et des formats d'échange d'informations, y compris des données légales. L'Office des publications gère également un certain nombre d'applications et d'outils web utilisés par les institutions, agences et organismes de l'UE, par les administrations nationales et régionales en Europe, ainsi que par des gouvernements nationaux et des utilisateurs privés à travers le monde.

Une part substantielle des informations publiées par l'Office des publications est disponible dans les 24 langues officielles de l'Union européenne, par exemple la totalité des actes législatifs publiés dans EUR-Lex¹, le portail du droit de l'UE, et une partie des publications de l'UE, diffusées en ligne via EU Bookshop². L'aspect multilingue de la production de l'Office des publications représente un certain nombre de défis en ce qui concerne l'optimisation de l'accès aux ressources. L'une des solutions mises en œuvre par l'Office pour régler ces problèmes consiste à utiliser des vocabulaires contrôlés multilingues avec des identificateurs non linguistiques. Ces vocabulaires peuvent être trouvés dans le répertoire de métadonnées de l'Office des publications³. L'ensemble de ces outils sont fournis en tant que données ouvertes, qui peuvent être réutilisées par toutes les parties intéressées dans les États membres de l'Union européenne et au-delà. En mettant à disposition de tels outils, l'Office des publications facilite la compréhension mutuelle et l'interopérabilité entre des systèmes

¹ <http://eur-lex.europa.eu/>.

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

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

législatifs divergents, contribuant ainsi à la création d'un espace européen commun pour la coopération et l'apprentissage mutuel dans ce domaine.

EuroVoc

EuroVoc est le vocabulaire contrôlé le plus utile pour optimiser l'accès aux données législatives européennes et nationales, ainsi que pour d'autres sources d'information. Il s'agit d'un thésaurus pluridisciplinaire qui propose quelque 8 000 concepts dans 23 langues officielles de l'UE, plus trois autres, correspondant aux pays candidats à l'adhésion. Il couvre les domaines d'activité des institutions de l'Union européenne. Les concepts sont classés par politique et domaine d'activité de l'UE (21 domaines et 120 micro-thésaurus/sous-domaines). EuroVoc vise deux finalités en matière d'indexation de contenus: la législation de l'UE, dans EUR-Lex, et les publications de l'UE, diffusées par l'Office des publications via «EU Bookshop» et OPac⁴ (le catalogue d'accès public en ligne de l'Office des publications). Les concepts EuroVoc étant indépendants de la langue, les utilisateurs peuvent rechercher un concept dans leur propre langue et trouver les résultats dans une autre langue.

EuroVoc est compilé conformément aux normes de l'Organisation internationale de normalisation (ISO 25964 – *Thésaurus et interopérabilité avec d'autres vocabulaires*). L'outil est adapté en continu afin de tenir compte de trois paradigmes en évolution: les domaines dans lesquels les institutions de l'Union sont concernées, le régime linguistique et les innovations technologiques. L'équipe de l'Office des publications, chargée de la maintenance d'EuroVoc, collecte et examine les propositions émanant des parlements nationaux, d'institutions ou d'agences européennes et d'utilisateurs privés (particuliers ou sociétés). Un processus de gouvernance a été mis en place pour l'administration et la maintenance du thésaurus, les mises à jour se faisant en principe deux fois par an.

EuroVoc est accessible pour la navigation et la recherche à travers son site web (<http://eurovoc.europa.eu>). Il peut également être téléchargé depuis le portail des données ouvertes de l'UE (<https://open-data.europa.eu/en/data/dataset/eurovoc>) dans des formats lisibles par machine, comme XML et SKOS/RDF. De plus, certains services en ligne permettent aux utilisateurs de consulter EuroVoc et d'afficher ou d'utiliser les résultats directement dans leurs applications: <http://eurovoc.europa.eu/drupal/?q=webservice&cl=en>. L'Office des publications est un pionnier en matière d'utilisation des technologies sémantiques dans ses chaînes de production et de diffusion. EuroVoc est disponible via le point d'entrée SPARQL de l'Office des publications et peut être utilisé conjointement avec d'autres ressources législatives accessibles via l'Office des publications à des fins de réutilisation et pour les «données liées» (ou «web des données»).

EuroVoc, qui est devenu l'outil de référence pour l'indexation des informations UE, est utilisé par les institutions, les agences et les organismes de l'UE, par des parlements nationaux et régionaux en Europe, par des administrations publiques, des bibliothèques, des professionnels de l'information et des utilisateurs privés à travers le monde entier.

Vers le web sémantique

Les données ouvertes et liées (DOL) sont en train de devenir une norme de fait et une référence en matière de publication de données. Cette évolution s'explique essentiellement grâce à un important jeu de standards (RDF-S, OWL, etc.), des ontologies largement

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

acceptées (Dublin Core, DCAT, SKOS(-XL), schema.org, etc.) et par une communauté fortement impliquée.

EuroVoc est l'un des premiers vocabulaires multilingues à utiliser des technologies du web sémantique qui reflètent les recommandations du W3C et les dernières tendances en matière de normes thésaurus. Initialement, EuroVoc a été établi sur la base des ontologies SKOS («Simple Knowledge Organization System») et SKOS-XL (extension SKOS pour labels). Graduellement, les informations ont évolué au-delà des simples modèles SKOS et SKOS-XL vers un nouveau modèle représentant les besoins spécifiques de l'Office des publications. Afin d'assurer l'interopérabilité et de promouvoir une large utilisation des données publiées, l'Office des publications a développé un profil d'application (SKOS-AP-EU), en vue de modéliser les thésaurus et les vocabulaires contrôlés⁵. Le profil d'application SKOS-AP-EU enrichit SKOS avec des propriétés provenant d'un grand nombre de vocabulaires reconnus. Ce profil d'application est exprimé formellement en utilisant SHACL («shape constraint language»). Il permet de valider automatiquement des vocabulaires contrôlés et de générer simultanément de la documentation.

EuroVoc en tant que données liées

Alignement d'EuroVoc sur des vocabulaires contrôlés plus spécialisés

Le fait qu'EuroVoc soit disponible dans des formats sémantiques tels que SKOS offre une excellente occasion de créer des liens sémantiques et d'aligner cet outil avec d'autres vocabulaires plus spécialisés.

EuroVoc couvre différentes disciplines au niveau générique et ne peut donc pas refléter la spécificité requise par les experts dans un domaine donné. Pour assurer une couverture plus large, EuroVoc est aligné sur des thésaurus plus spécialisés, tels que le thésaurus de l'Organisation pour l'alimentation et l'agriculture (AGROVOC), les thésaurus des Nations unies (UNESCO et UNBIS) et le Thésaurus de l'économie STW géré par le Leibniz Information Centre for Economics, etc. La fondation Leibniz va intégrer les labels EuroVoc alignés avec ceux du Thésaurus STW afin d'enrichir son moteur de recherche.

GEMET, le thésaurus multilingue général de l'environnement, a été élaboré en tant qu'outil d'indexation pour l'Agence européenne pour l'environnement (EEA)⁶. L'alignement sémantique de GEMET avec EuroVoc apporte une valeur ajoutée en facilitant l'accès aux documents dans le domaine de l'environnement, qui constitue un domaine clé des activités de l'UE. Les deux thésaurus sont multilingues (23 langues pour EuroVoc, 35 pour GEMET), avec différents niveaux d'abstraction et de détail. Par exemple: GEMET utilise le concept de «législation environnementale»⁷ et la correspondance sémantique dans EuroVoc est «droit environnemental»⁸. Grâce à l'alignement des deux thésaurus, les utilisateurs pourraient combiner la terminologie de l'un des thésaurus dans leur langue maternelle et laisser le système proposer des correspondances vers l'autre terminologie afin de récupérer des résultats équivalents.

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

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

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

⁸ <http://eurovoc.europa.eu/535>.

Alignement d'EuroVoc sur des vocabulaires nationaux contrôlés afin d'annoter les informations juridiques nationales

EuroVoc couvre différentes disciplines au niveau générique et ne peut donc pas refléter les différences entre des systèmes juridiques distincts à un niveau suffisamment détaillé. Toutefois, en liant EuroVoc et les vocabulaires contrôlés de systèmes juridiques, on peut relier les concepts juridiques génériques – ou les concepts UE d'EuroVoc – vers des termes utilisés dans la législation nationale d'un pays donné. Ainsi, les deux peuvent être utilisés pour optimiser l'accès aux ressources à la fois dans la législation nationale et dans celle de l'UE (via EUR-Lex). Un autre avantage des alignements de thésaurus est lié à l'extension de la stratégie de recherche vers d'autres bases de données nationales ou européennes.

L'avantage de l'alignement d'un vocabulaire national avec EuroVoc peut être résumé comme suit: si un concept provenant d'un système juridique a un équivalent dans EuroVoc, il est automatiquement traduit; si le concept national est trop spécifique, il peut être orienté vers un concept EuroVoc plus générique. Par exemple: la gestion de la Moselle concerne trois pays (Allemagne, France et Luxembourg). Dans Legilux, le site législatif officiel du Luxembourg, le concept à utiliser est celui de «canalisation de la Moselle»⁹. EuroVoc décrit la géographie des États membres uniquement au niveau régional et a une équivalence pour ce concept très spécifique. En alignant l'expression «canalisation de la Moselle» sur le concept EuroVoc existant de «[voies navigables intérieures](#)», la stratégie de recherche pourrait être étendue en vue d'améliorer la comparaison de données législatives nationales et/ou européennes.

Des coopérations officielles existent entre des États membres et l'Union européenne pour l'utilisation d'EuroVoc pour l'indexation de données administratives et parlementaires nationales. Par exemple, le Conseil national de la République slovaque¹⁰ a intégré EuroVoc dans le contexte de son système d'information parlementaire. L'Assemblée nationale de la République de Slovaquie¹¹ utilise EuroVoc comme base pour la liste de codes du portail UE de l'administration. Le parlement et le sénat belges¹² utilisent EuroVoc comme langage contrôlé pour effectuer des recherches dans leurs bases de données respectives.

Parmi les autres exemples de ce type de coopération, entrepris dans le cadre de l'[identificateur européen de la législation](#) (ELI), on peut citer le projet mené par le Journal officiel du Grand-Duché de Luxembourg et l'Office des publications. Cette initiative a aligné les concepts EuroVoc sur le vocabulaire contrôlé utilisé pour indexer le contenu de la législation nationale luxembourgeoise, publiée sur <http://legilux.public.lu/>. Ce projet vise à rendre interopérable la législation européenne et celle du Luxembourg et à faciliter l'accès aux ressources juridiques. Les résultats confirment que l'utilisation d'EuroVoc en tant que pont entre des vocabulaires contrôlés plus spécialisés des systèmes législatifs nationaux offre de nets avantages en matière d'optimisation de l'accès à la législation.

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> et https://www.senate.be/www/?MIval=/index_senate&MENUID=12000&LANG=fr.

VocBench

L'Office des publications utilise *VocBench* pour la maintenance d'EuroVoc. *VocBench* est une plateforme de collaboration «open source», destinée à la gestion de vocabulaires contrôlés multilingues, qui fait usage de technologies sémantiques et respecte les normes SKOS et SKOS-XL. *VocBench* est particulièrement adapté à la gestion de thésaurus en format RDF.

L'Office des publications est un membre actif de la grande communauté des thésaurus multilingues et contribue à la progression de *VocBench* vers sa prochaine version. La dernière version majeure de *VocBench 3.0*, développée grâce au financement du [programme ISA²](#) de la Commission européenne, contient plusieurs fonctionnalités, y compris des éléments de gestion. Disponible au 3^{ème} trimestre 2017, elle offrira aux utilisateurs la boîte à outils nécessaire pour créer, maintenir, lier et publier leurs vocabulaires contrôlés, leurs métadonnées ou leurs glossaires sous la forme de données ouvertes liées. *VocBench 3.0* proposera un environnement collaboratif pour le développement de tout type de données RDF; les capacités d'édition de la version précédente seront améliorées et les aspects spécifiques qui ont fait son succès seront maintenus.

VocBench est «open source» et peut être utilisé par toute organisation publique, par toute société et par tout utilisateur en vue de gérer ses vocabulaires contrôlés. Le dernier paquet d'installation ainsi que les instructions d'installation et de déploiement pour *VocBench 2.x/3.0* figurent directement sur le [site du projet](#).

Conclusion

L'une des composantes essentielles de l'écosystème de l'information juridique multilingue dans l'Union européenne, dont l'objectif est d'encourager l'unité dans la diversité, consiste à optimiser l'accès aux sujets et aux données juridiques depuis un grand nombre de sources différentes, notamment les institutions de l'UE et les instances chargées de publier la législation nationale des États membres de l'UE. Grâce à l'utilisation des normes ouvertes et à la mise à disposition – pour réutilisation dans le monde entier – des outils pertinents décrits dans le présent article, l'Office des publications va plus loin et défend l'idée selon laquelle la solidarité dans la diversité est une proposition valable et utile, qui présente des avantages substantiels pour ceux qui veulent bien la mettre en pratique.

Références

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

Office des publications de l'Union européenne, 2010, Conférence Eurovoc: Mind the Lexical Gap. EU. Office des publications. <http://eurovoc.europa.eu/drupal/?q=node/936>.

Stellato A. et al. (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, vol 9088. Springer, Cham. http://link.springer.com/chapter/10.1007/978-3-319-25639-9_29.