

## **Optimización del acceso por materias a los recursos jurídicos: EuroVoc y VocBench.**

### **Ventajas de la utilización de vocabularios multilingües controlados y de una herramienta colaborativa de código abierto para su mantenimiento**

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

#### **Rudolf W. Strohmeier**

Oficina de Publicaciones de la Unión Europea, Luxemburgo

#### **Willem van Gemert**

Oficina de Publicaciones de la Unión Europea, Luxemburgo

#### **Christine Laaboudi**

Oficina de Publicaciones de la Unión Europea, Luxemburgo

#### **Anne Waniart**

Oficina de Publicaciones de la Unión Europea, Luxemburgo



This is a Spanish 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>

Las opiniones expresadas en el presente documento son las de sus autores y no representan la opinión de la Comisión Europea ni de ninguna otra organización.

---

## **RESUMEN:**

*La Oficina de Publicaciones de la Unión Europea es un organismo interinstitucional encargado de publicar la legislación y otros textos de la Unión Europea (UE) en las veinticuatro lenguas oficiales de la UE. Para afrontar el reto de la publicación multilingüe de datos jurídicos, la Oficina de Publicaciones utiliza vocabularios multilingües controlados, como tablas de autoridades y el tesoro EuroVoc.*

*EuroVoc es un tesoro multilingüe y multidisciplinario que abarca los ámbitos de actividad de las instituciones de la UE. Aplica los estándares web semánticos recomendados por la organización W3C. Estas tecnologías facilitan, además de la flexibilidad de modelado y las capacidades de razonamiento, el alineamiento semántico de EuroVoc con otros tesauros y vocabularios controlados, como los vocabularios nacionales y de ámbitos específicos.*

*La creación, el mantenimiento y la gestión del tesoro EuroVoc se efectúa a través de VocBench, una plataforma colaborativa de código abierto basada en la web para el desarrollo de tesauros, conforme con los estándares de SKOS y SKOS-XL. Gracias a la financiación del programa ISA<sup>2</sup> de la Comisión Europea, se ha desarrollado una nueva versión de VocBench que estará disponible en el tercer trimestre de 2017.*

**Palabras clave:** vocabularios controlados, tesoro, acceso por materias, estándares web semánticos

---

## **Acerca de la Oficina de Publicaciones de la Unión Europea**

La Oficina de Publicaciones de la Unión Europea es un organismo interinstitucional creado en 1969 por las instituciones de la UE y encargado de publicar la legislación y otros textos de la UE en las veinticuatro lenguas oficiales de la Unión Europea. Con ello, contribuye a una aplicación práctica del lema de la Unión Europea «Unidos en la diversidad», en referencia a la unión de los europeos en la lucha por la paz y la prosperidad beneficiándose al mismo tiempo de las múltiples culturas, tradiciones y lenguas de Europa.

En este contexto, como proveedor moderno de información multilingüe, la Oficina de Publicaciones ofrece servicios de gestión de la información de última generación a todas las instituciones, agencias y organismos de la UE, así como a los ciudadanos de la UE en general. Como actor reconocido de normalización, ha asumido también un papel destacado en la normalización de metadatos y formatos, incluidos datos jurídicos, para el intercambio de información. La Oficina de Publicaciones administra asimismo una serie de herramientas de gestión de la información utilizadas para ese fin por las instituciones, agencias y organismos de la UE, las administraciones nacionales y regionales en Europa y los gobiernos nacionales y usuarios privados de todo el mundo.

Una parte sustancial de la información que publica la Oficina de Publicaciones está disponible en las veinticuatro lenguas oficiales de la Unión Europea, por ejemplo todos los actos jurídicos publicados en EUR-Lex<sup>1</sup>, el portal del Derecho de la UE, y parte de las publicaciones de la UE, distribuidas en línea a través de EU Bookshop<sup>2</sup>. El aspecto multilingüe de las publicaciones de la Oficina de Publicaciones plantea una serie de retos en lo que respecta a la optimización del acceso por materias. Una de las soluciones que ofrece la Oficina de Publicaciones para afrontar esos retos es la utilización de vocabularios multilingües controlados con identificadores independientes de la lengua. Estos vocabularios están disponibles en el Registro de Metadatos de la Oficina de Publicaciones<sup>3</sup>. Todas estas herramientas se ofrecen como datos abiertos que puede reutilizar cualquier interesado en los Estados miembros de la Unión Europea e incluso fuera de ella. Al poner a disposición estas

---

<sup>1</sup> <http://eur-lex.europa.eu/>.

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

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

herramientas, la Oficina de Publicaciones facilita la comprensión mutua y la interoperabilidad entre sistemas jurídicos divergentes, contribuyendo así a la creación de un espacio común europeo para la colaboración y el aprendizaje mutuo en este ámbito.

## **EuroVoc**

EuroVoc es el vocabulario controlado más útil para optimizar el acceso a los contenidos de los datos jurídicos de la UE y nacionales, así como de otras fuentes de información. Es un tesoro multidisciplinario que contiene unos ocho mil conceptos en veintitrés lenguas oficiales de la UE y en tres lenguas más de países candidatos a la adhesión a la UE. Abarca todos los campos de actividad de las instituciones de la Unión Europea. Los conceptos están clasificados en función de las políticas y los campos de actividad de la UE en veintiún ámbitos y ciento veinte microtesoros o subámbitos. EuroVoc persigue dos objetivos de indización de contenidos: la legislación de la UE, en EUR-Lex; y las publicaciones de la UE, distribuidas por la Oficina de Publicaciones a través de EU Bookshop y OPac<sup>4</sup> (el Catálogo público online de la Oficina de Publicaciones). Puesto que los conceptos de EuroVoc son independientes de la lengua, los usuarios pueden buscar un concepto en su propio idioma y obtener resultados en otra lengua.

EuroVoc se compila de acuerdo con las normas de la Organización Internacional de Normalización ISO 25964 (*Tesoros e interoperabilidad con otros vocabularios*). Se adapta continuamente en función de tres paradigmas evolutivos: los ámbitos que afectan a las instituciones de la UE, los regímenes lingüísticos y las innovaciones tecnológicas. El equipo responsable de EuroVoc recopila y examina las propuestas procedentes de los parlamentos nacionales, de las instituciones o agencias de la UE y de los usuarios privados (ciudadanos o empresas). Se ha establecido un proceso de gobernanza para tratar la administración y el mantenimiento del tesoro, con la publicación de actualizaciones de EuroVoc, en principio, dos veces al año.

Se puede navegar y buscar en EuroVoc desde su sitio web (<http://eurovoc.europa.eu>). También se puede descargar desde el Portal de datos abiertos de la UE (<https://data.europa.eu/euodp/es/data/dataset/eurovoc>) en formatos legibles por máquina, como XML y SKOS/RDF. Asimismo, una serie de servicios web permiten a los usuarios hacer búsquedas en EuroVoc y visualizar o utilizar los resultados directamente en sus aplicaciones: <http://eurovoc.europa.eu/drupal/?q=es/node/1307&cl=en>. La Oficina de Publicaciones es pionera en la utilización de tecnologías semánticas en sus cadenas de producción y difusión. EuroVoc está disponible a través del punto final SPARQL de la Oficina de Publicaciones y puede utilizarse en combinación con otros recursos jurídicos que ofrece la Oficina de Publicaciones con fines de reutilización y enlace de datos.

EuroVoc, que se ha convertido en la herramienta de referencia para la indización de la información de la UE, es utilizado por instituciones, agencias y organismos de la UE, parlamentos nacionales y regionales y gobiernos en Europa, así como administraciones públicas, bibliotecas, profesionales de la información y usuarios privados en todo el mundo.

---

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

## Hacia la web semántica

Los datos enlazados abiertos (LOD, por sus siglas en inglés) se están convirtiendo rápidamente en una norma *de facto* y en un conjunto de prácticas en el mundo de la publicación de datos. Ello se debe principalmente a un amplio conjunto de estándares (RDF-S, OWL, etc.), ontologías ampliamente aceptadas (Dublin Core, DCAT, SKOS(-XL), schema.org, etc.) y una comunidad muy implicada.

EuroVoc es uno de los primeros vocabularios multilingües que utiliza tecnologías web semánticas que siguen las recomendaciones del W3C y las últimas tendencias en materia de normas sobre tesauros. Inicialmente EuroVoc se basaba en las ontologías SKOS (*Simple Knowledge Organization System*) y SKOS-XL (extensión de SKOS para etiquetas). El contenido de la información ha desbordado gradualmente los simples modelos SKOS y SKOS-XL y evolucionado hacia un nuevo modelo representativo de las necesidades derivadas de las actividades de la Oficina de Publicaciones. Para posibilitar la interoperabilidad y fomentar amplias posibilidades de uso de los datos publicados, la Oficina de Publicaciones ha desarrollado un perfil de aplicación (SKOS-AP-EU), para constituir tesauros y listas autorizadas controladas<sup>5</sup>. El perfil de aplicación SKOS-AP-EU integra en SKOS las propiedades de una serie de vocabularios bien conocidos. Este perfil de aplicación utiliza formalmente SHACL (*shape constraint language*). Permite la validación automática de vocabularios controlados y la generación simultánea de documentación.

## EuroVoc como datos enlazados

### *Mapeo de EuroVoc con vocabularios controlados más especializados*

El hecho de que EuroVoc esté disponible en formatos semánticos tales como SKOS ofrece una excelente oportunidad para mapearlo y alinearlos semánticamente con vocabularios más especializados.

EuroVoc abarca diversas disciplinas a un nivel genérico y, por lo tanto, no siempre puede reflejar la especificidad que requieren los expertos en un ámbito determinado. Para garantizar una cobertura más amplia, EuroVoc está alineado con tesauros más especializados, como el tesoro de la Organización para la Agricultura y la Alimentación (AGROVOC), los tesauros de las Naciones Unidas (UNBIS y UNESCO), el tesoro de economía STW, que mantiene el Leibniz Information Centre for Economics, etc. La Fundación Leibniz integrará las etiquetas de EuroVoc mapeadas en el tesoro STW para enriquecer su motor de búsqueda.

GEMET (GEneral Multilingual Environmental Thesaurus) se ha desarrollado como herramienta de indización para la Agencia Europea de Medio Ambiente (AEMA)<sup>6</sup>. El mapeo semántico de GEMET con EuroVoc supondrá un valor añadido facilitando el acceso a los documentos en el ámbito medioambiental, que constituye un campo clave de las actividades de la UE. Ambos tesauros son multilingües (veintitrés lenguas en el caso de EuroVoc y treinta y cinco en el caso de GEMET), con distintos niveles de abstracción y detalle. Por ejemplo: GEMET utiliza el concepto de «legislación ambiental»<sup>7</sup> y la correspondencia semántica en EuroVoc es «Derecho del medio ambiente»<sup>8</sup>. Gracias al mapeo de ambos

<sup>5</sup> Fuente: Costetchi E. y Van Gemert W. (2016).

<sup>6</sup> [https://europa.eu/european-union/about-eu/agencies/eea\\_es](https://europa.eu/european-union/about-eu/agencies/eea_es).

<sup>7</sup> <http://www.eionet.europa.eu/gemet/es/concept/2862>.

<sup>8</sup> <http://eurovoc.europa.eu/drupal/?q=es/request&uri=http://eurovoc.europa.eu/535>.

tesauros, los usuarios pueden combinar terminología que les resulte familiar de uno de los dos tesauros en sus lenguas maternas y hacer que el sistema ofrezca correspondencias con otras terminologías para obtener resultados equivalentes.

### *Mapeo de EuroVoc con vocabularios nacionales controlados para anotar la información jurídica nacional*

EuroVoc abarca distintas disciplinas a un nivel genérico y, por lo tanto, no puede reflejar de forma suficientemente detallada las diferencias entre los diversos sistemas jurídicos. No obstante, la vinculación de EuroVoc y los vocabularios controlados de los sistemas jurídicos puede ampliar los conceptos jurídicos genéricos, o los conceptos de la UE procedentes de EuroVoc, a los términos utilizados en la legislación nacional de un país determinado. En consecuencia, pueden utilizarse ambos vocabularios para optimizar el acceso a contenidos de la legislación nacional o la legislación de la UE (a través de EUR-Lex). Otra ventaja de la alineación semántica de tesauros es la ampliación de la estrategia de búsqueda a otras bases de datos jurídicos nacionales o europeas.

El beneficio que aporta la alineación semántica de un vocabulario nacional con EuroVoc puede resumirse de la manera siguiente: si un concepto de un sistema jurídico no tiene equivalente en EuroVoc, se traduce automáticamente; si el concepto nacional es demasiado específico, puede mapearse con un concepto de EuroVoc más genérico. Por ejemplo: la gestión del río Mosela corresponde a tres países (Alemania, Francia y Luxemburgo). En Legilux, el sitio web oficial de legislación en línea de Luxemburgo, el concepto que debe utilizarse es «canalisation de la Moselle»<sup>9</sup>. EuroVoc describe la geografía de los Estados miembros únicamente a nivel regional y no tiene equivalencia para este concepto tan específico. Mediante la alineación del concepto específico «canalisation de la Moselle» con el concepto de EuroVoc existente «[curso de agua interior](#)», podría ampliarse la estrategia de búsqueda para mejorar la comparación de los datos jurídicos nacionales y/o europeos.

Existen colaboraciones oficiales entre los Estados miembros y la Unión Europea en las que está previsto utilizar EuroVoc para indizar los datos gubernamentales y parlamentarios nacionales. Por ejemplo, el Consejo Nacional de la República Eslovaca<sup>10</sup> ha utilizado EuroVoc en el contexto de su sistema de información parlamentaria. La Asamblea Nacional de la República de Eslovenia<sup>11</sup> utiliza EuroVoc como base para la lista de códigos del portal UE de la Administración. El Parlamento y el Senado belgas<sup>12</sup> utilizan EuroVoc como vocabulario controlado para la búsqueda en sus bases de datos respectivas.

Otro ejemplo de este tipo de colaboración, en el contexto del [Identificador Europeo de Legislación](#) (ELI), es un proyecto dirigido por el Diario Oficial del Gran Ducado de Luxemburgo y la Oficina de Publicaciones. En esta iniciativa se han alineado los conceptos de EuroVoc con el vocabulario controlado para indizar el contenido de la legislación nacional del Gran Ducado de Luxemburgo publicada en <http://legilux.public.lu/>. La finalidad del proyecto era hacer interoperables la legislación de la UE y la de Luxemburgo y facilitar el acceso a los recursos jurídicos. Los resultados confirman que la utilización de EuroVoc como

---

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

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

<sup>11</sup> [http://www.mzz.gov.si/en/foreign\\_policy\\_and\\_international\\_law/eu\\_policies/language\\_issues/legal\\_and\\_linguistic\\_databases/eurovoc/](http://www.mzz.gov.si/en/foreign_policy_and_international_law/eu_policies/language_issues/legal_and_linguistic_databases/eurovoc/).

<sup>12</sup> <http://www.dekamer.be/kvvcr/showpage.cfm?section=/flwb&language=nl&cfm=ListKeyword.cfm?legislat=54> y [https://www.senate.be/www/?MIval=/index\\_senate&MENUID=12000&LANG=fr](https://www.senate.be/www/?MIval=/index_senate&MENUID=12000&LANG=fr).

puede tener diversas ventajas para optimizar el acceso a la legislación.

## VocBench

La Oficina de Publicaciones utiliza *VocBench* para el mantenimiento de EuroVoc. *VocBench* es una plataforma colaborativa de código abierto basada en la web para la gestión de vocabularios controlados multilingües que utiliza tecnologías semánticas y se ajusta a los estándares de SKOS y SKOS-XL. Es especialmente adecuada para la gestión de amplios tesauros en formato RDF.

La Oficina de Publicaciones es un miembro activo de la amplia comunidad de tesauros multilingües e impulsa el desarrollo de la próxima versión de *VocBench*. La última versión importante de *VocBench 3.0*, desarrollada gracias a la financiación del [programa ISA<sup>2</sup>](#), de la Comisión Europea, contiene una serie de funcionalidades nuevas, incluidos elementos de alineación semántica integrados. Esta versión, que estará disponible en el tercer trimestre de 2017, ofrecerá a los usuarios las herramientas necesarias para crear, mantener, enlazar y publicar sus vocabularios controlados, metadatos o léxicos como datos enlazados abiertos. *VocBench 3.0* ofrecerá un entorno colaborativo de propósito general para el desarrollo de cualquier tipo de series de datos en formato RDF, lo que supone una mejora de las capacidades de edición respecto a su antecesor, pero manteniendo las características peculiares que determinaron el éxito de este.

*VocBench* es una plataforma de código abierto que puede utilizar cualquier organización pública, empresa o usuario independiente para gestionar vocabularios controlados. Los archivos de instalación más recientes, así como las instrucciones de instalación de *VocBench 2.x/3.0*, están disponibles en el [sitio del proyecto](#).

## Conclusión

Garantizar un acceso optimizado por materias a los datos jurídicos desde una amplia gama de fuentes divergentes, como las instituciones de la UE y los editores de legislación nacional de los Estados miembros de la UE, constituye un elemento esencial del ecosistema de información jurídica multilingüe en la Unión Europea, ecosistema cuyo desarrollo favorece la promoción de la unidad entre sus diversos miembros. Gracias a la utilización de estándares abiertos y a la puesta a disposición de las herramientas pertinentes presentadas en este artículo para su reutilización en todo el mundo, la Oficina de Publicaciones va más allá y promueve la idea de que apostar por la solidaridad en la divergencia es una propuesta válida y valiosa que puede dar grandes beneficios a quienes estén dispuestos a ponerla en práctica.

## Referencias

Costetchi E. y Van Gemert W. (2016). *Towards executable application profile for European vocabularies* (Hacia un perfil de aplicación ejecutable para los vocabularios europeos). W3C. [https://www.w3.org/2016/11/sdsvoc/SDSVoc16\\_paper\\_23](https://www.w3.org/2016/11/sdsvoc/SDSVoc16_paper_23).

Oficina de Publicaciones de la Unión Europea (2010) Conferencia EuroVoc: *Mind the Lexical Gap* (Atención al vacío léxico). UE. Oficina de Publicaciones <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](http://link.springer.com/chapter/10.1007/978-3-319-25639-9_29).