

TerésIA

Terminología + IA en español

INFORME DEL MODELO DE REPRESENTACIÓN (ONTOLOGÍA) DE DATOS EN TERESIA

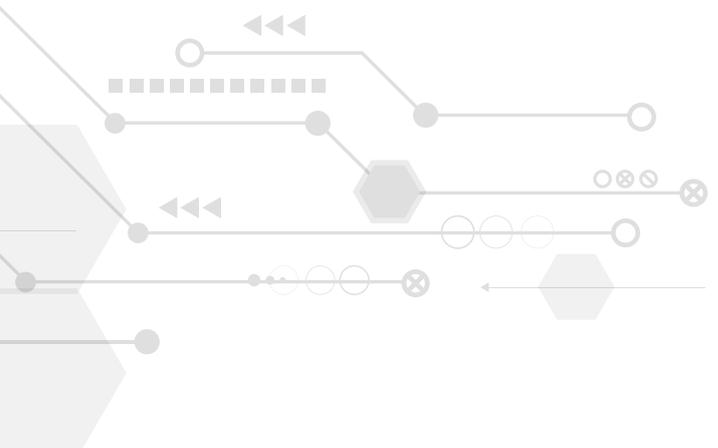
30 de septiembre de 2024

IMROD-1.0.0-20240930



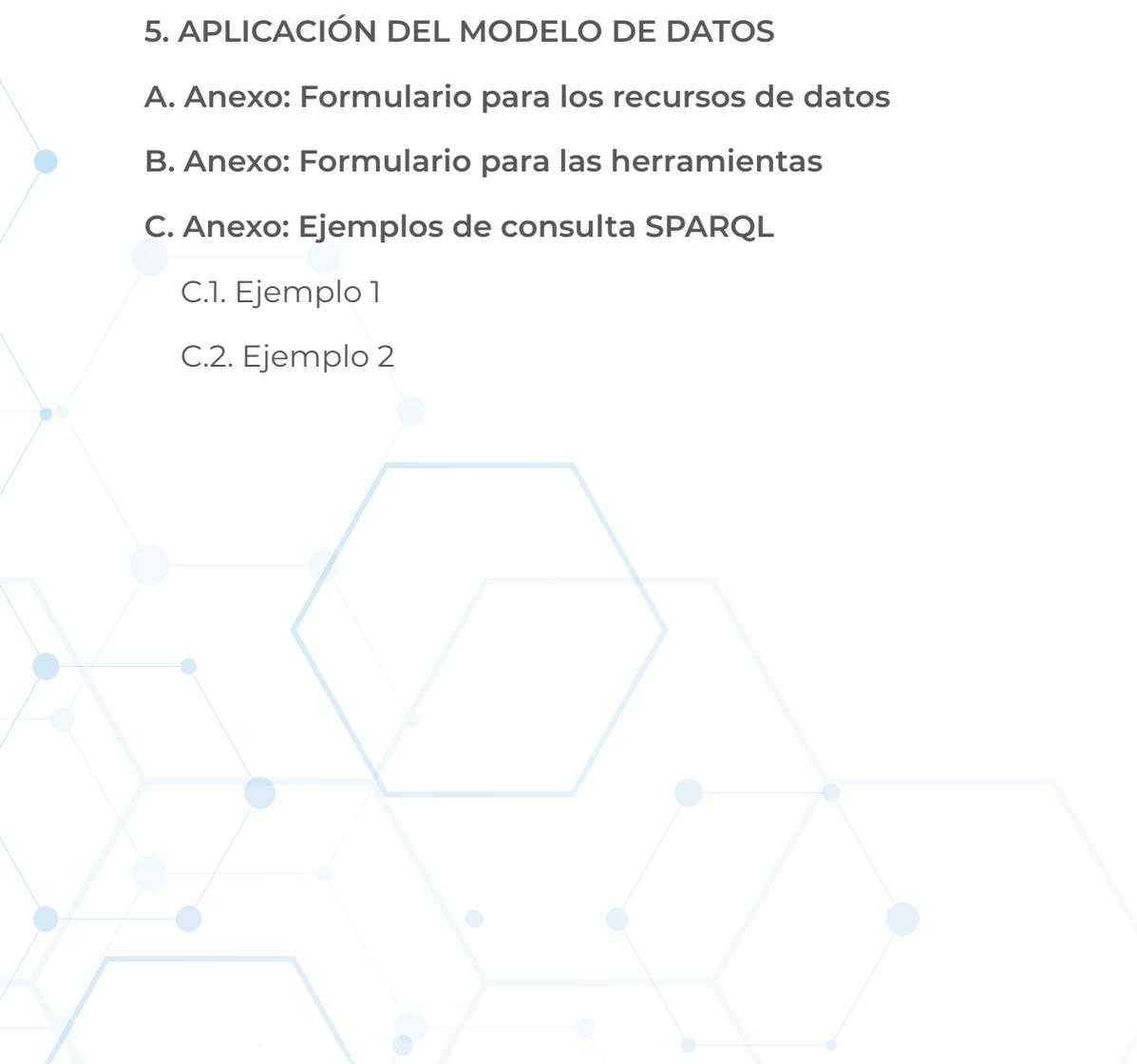
ASOCIACIÓN **aet=R**
ESPAÑOLA DE TERMINOLOGÍA





ÍNDICE

1. INTRODUCCIÓN	3
2. DATOS RECOPIRADOS	3
3. DATOS ENLAZADOS Y ONTOLOGÍAS	5
4. MODELADO PARA LOS RECURSOS LINGÜÍSTICOS EN TERESIA	7
4.1. Representación de los metadatos de recursos lingüísticos (datos)	7
4.2. Representación de los metadatos de las herramientas	10
5. APLICACIÓN DEL MODELO DE DATOS	11
A. Anexo: Formulario para los recursos de datos	15
B. Anexo: Formulario para las herramientas	17
C. Anexo: Ejemplos de consulta SPARQL	19
C.1. Ejemplo 1	19
C.2. Ejemplo 2	20



1. INTRODUCCIÓN

El presente documento corresponde al «Entregable 2.2 [Modelo de representación (ontología) de datos en TeresIA]», asociado al «Paquete de Trabajo 2 (Generación de repositorio de contenido e identificación de recursos existentes)» y al «Hito 3 (Modelos de representación de datos en TeresIA)». En esta versión final del informe, se describe un modelo de representación creado para los metadatos de los recursos que se expondrán a través del catálogo en el portal de TeresIA.

Para una mayor comprensión de esta tarea (el modelado de los metadatos de los recursos terminológicos), primero se presentará el proceso de recopilación de los datos, para la familiarización de contenido que será modelado. A continuación, se describirá qué son los datos enlazados y las ontologías, haciendo hincapié en las ontologías que se usarán en TeresIA. Una vez introducido el contexto, se detallará la representación de los datos obtenidos, aplicando el esquema establecido por la ontología de META-SHARE. Finalmente, se explicará brevemente el proceso completo de transformación y explotación de los datos mediante las consultas SPARQL y los grafos interactivos.

2. DATOS RECOPIRADOS

El portal de TeresIA cuenta con documentar y exponer los diferentes recursos terminológicos que ofrece cada institución. Cabe mencionar que no sólo se incluirán datos terminológicos en el catálogo (terminologías, tesauros, etc.); también se podrán exponer herramientas para la anotación y construcción de recursos terminológicos, así como corpus que se puedan utilizar para la extracción de términos y relaciones semánticas.

La información que se expondrá en el catálogo de TeresIA sobre los recursos terminológicos es recogida a través de un formulario interactivo¹. Dicho formulario sirve para documentar los recursos que se incluirán en el portal. El formulario cuenta con tres o cuatro páginas a rellenar, en base a la naturaleza del recurso que se está registrando (datos o herramientas). A continuación, se explica brevemente la información que recogen las diferentes páginas del formulario; asimismo, el **Anexo A** incluye fotos de las páginas del formulario para los recursos de datos, mientras que el **Anexo B** muestra fotos del formulario para herramientas.

1. Información básica: la primera página es común para los datos y las herramientas, y recoge los datos básicos del recurso, es decir, el nombre y tipo de recurso (por ejemplo, terminología, tesoro, corpus sin anotar, herramienta, etc.).

¹ Formulario: <https://form.jotform.com/proyectoTeresia/registro-recursos-terminolgicos>

2. Información sobre el contenido del recurso: esta sección es visible sólo para los recursos de datos y registra diversas características del contenido, como, por ejemplo, los idiomas, los dominios, relaciones entre términos, o número de términos o/y conceptos.
3. Aspectos técnicos del recurso: esta sección sólo se debe rellenar si se está registrando un recurso de datos y contiene preguntas sobre los aspectos técnicos, como, por ejemplo, los formatos en los que se puede disponer el recurso, el peso o la forma de acceso.
4. Información adicional de la herramienta: esta sección sólo aparecerá a aquellos que estén registrando una herramienta. Mediante esta sección se recoge la información relativa a la herramienta como el uso, información sobre la versión, o la forma de distribución.
5. Propiedad intelectual: esta última página es común para los datos y las herramientas, y recopila la información relativa a los autores, las entidades propietarias y las condiciones de uso.

Para la creación del formulario para el registro de los recursos terminológicos, se creó un primer borrador que fue testado y evaluado por el personal responsable del Centro de Traducción de los Órganos de la Unión Europea en Luxemburgo, que son los encargados del mantenimiento de IATE². IATE (por sus siglas en inglés, Interactive Terminology for Europe) es la base de datos terminológica de la Unión Europea (UE) y es utilizada en las instituciones y los organismos de la UE desde 2004 para recopilar, divulgar y gestionar la terminología específica de la UE. Tras el ejemplo de IATE, se analizó también la ontología META-SHARE³ y se adaptaron las preguntas en base a su esquema. La ontología META-SHARE es un modelo de datos que formaliza y estandariza la descripción de recursos lingüísticos. Se utiliza con la finalidad de facilitar la interoperabilidad, la reutilización y la accesibilidad de estos recursos a través de metadatos estructurados.

Finalmente, a fecha de 30 de septiembre de 2024, se han obtenido 20 respuestas, de las cuales 4 son herramientas, 14 terminologías y 2 diccionarios. El formulario de registro de recursos se elaboró a través de la plataforma **JotForm**, ya que permite una amplia selección de posibilidades para la elaboración y visualización de los campos (por ejemplo, selección múltiple con despleables). Asimismo, la plataforma permite descargar los datos recopilados en diversos formatos: PDF, CSV y Excel. Para la conversión y representación de estos datos en datos enlazados, los datos se descargan en CSV.

² <https://iate.europa.eu/home>

³ <http://w3id.org/meta-share/meta-share/2.0.0>

3. DATOS ENLAZADOS Y ONTOLOGÍAS

Los datos enlazados (*Linked Data* en inglés) son un enfoque basado en la conectividad e interoperabilidad de los datos. Para ello, se estructuran los datos conforme al estándar Resource Description Framework (RDF)⁴. Este estándar fue propuesto en 2004 por el World Wide Web Consortium (W3C), el organismo encargado de la estandarización de las tecnologías Web. (Montiel-Ponsoda and Gracia, 2019)

El RDF estructura los datos en triplas (sujeto-predicado-objeto) y el conjunto de todas las triplas genera un grafo. El primer elemento de una tripla es el sujeto, y representa el recurso del que se habla. Este elemento se representa mediante un identificador de recursos uniforme (URI, por sus siglas en inglés). El segundo elemento de una tripla es el predicado, y marca la relación o propiedad entre el sujeto y el objeto de la tripla. Los predicados o propiedades también son URIs. Finalmente, el tercer elemento de la tripla es el objeto, que indica el valor de la propiedad de un sujeto. Los objetos pueden ser literales (texto) o URIs. (Chaves-Fraga et al., 2022)

Para representar los datos de forma estándar, se recurre a las ontologías, es decir, a modelos o esquemas semánticos que constituyen una representación formal de un dominio. Las ontologías están formadas por varios elementos, como, por ejemplo, clases, propiedades e instancias. Las clases representan categorías o tipos de cosas en un dominio del conocimiento. Las propiedades son las relaciones o atributos que conectan clases o instancias. Finalmente, las instancias son elementos concretos o específicos que pertenecen a una clase.

Para la representación o modelado en RDF de los metadatos de los recursos registrados a través del formulario, se ha reusado el esquema establecido por la ontología META-SHARE⁵, la cual permite ser compatible con otras infraestructuras como CLARIN⁶ o European Language Grid⁷. Tal y como se muestra en la **Figura 1**, esta ontología se utiliza para describir recursos lingüísticos, ya sean corpus, recursos lexico-conceptuales (por ejemplo, diccionarios), herramientas y software, o descripciones de lenguaje (por ejemplo, normas gramaticales). Entre las características descriptivas que se pueden atribuir a un recurso, podemos destacar la autoría, así como las licencias de uso que tiene el recurso.

Figura 1 ►

⁴ <https://www.w3.org/RDF/>

⁵ <http://w3id.org/meta-share/meta-share/2.0.0>

⁶ <https://www.clarin.eu/>

⁷ <https://live.european-language-grid.eu/>

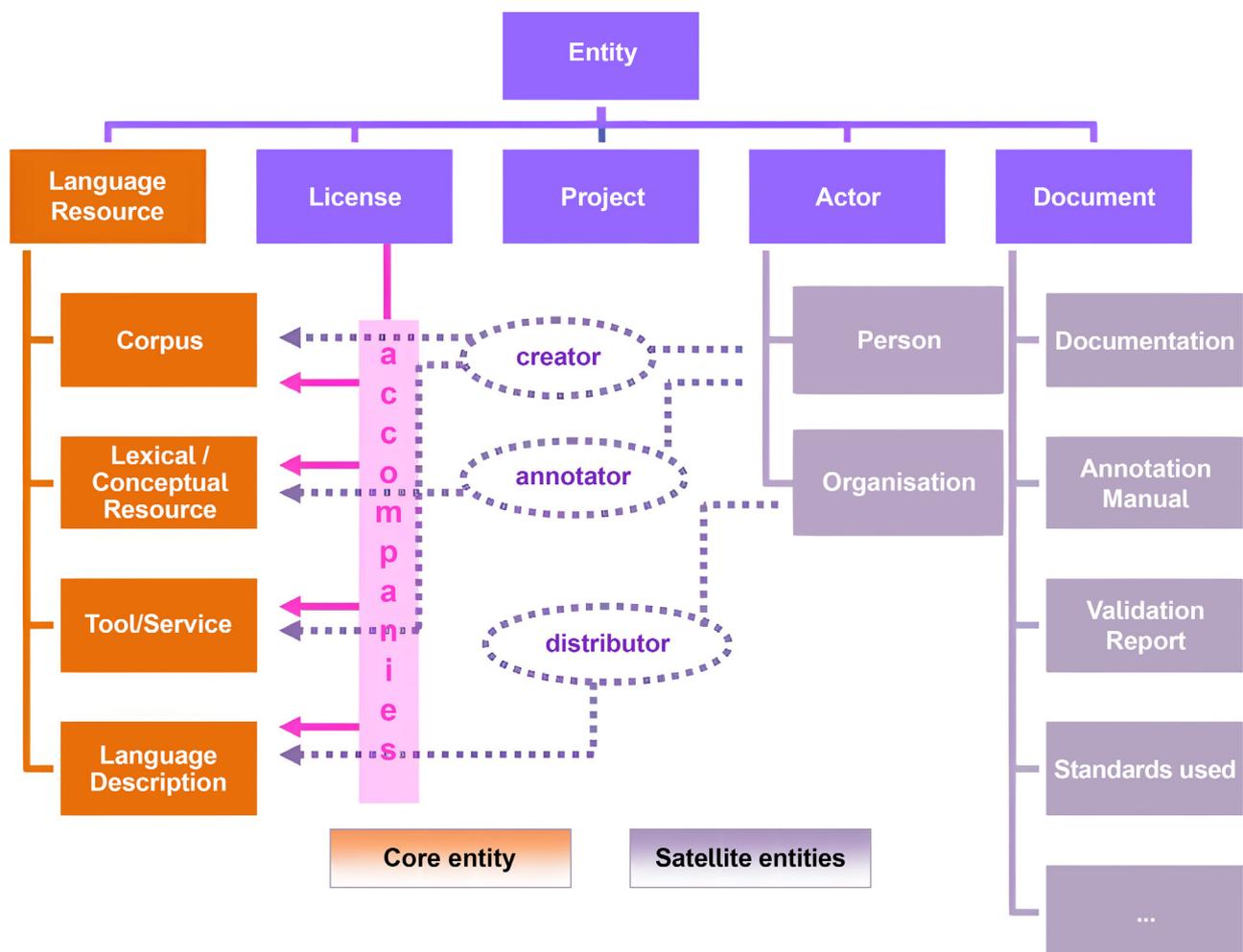


Figura 1: diagrama ontología META-SHARE, sacado de [Gavriliadou et al. \(2012\)](#)

Además de META-SHARE, se ha usado la ontología SKOS⁸, utilizada para la representación jerárquica de conceptos. Por ejemplo, para cubrir aquellas instancias que META-SHARE no considera, o la representación de los dominios de los recursos. En particular, se ha utilizado la versión SKOS de Eurovoc⁹ para la representación del dominio. Finalmente, también se han usado las ontologías DCTerms¹⁰ y RDFSchema¹¹ para etiquetar algunos elementos e indicar el formato del elemento (URI, string, integer...), respectivamente.

⁸ <http://www.lexinfo.net/ontology/3.0/lexinfo/>

⁹ <https://op.europa.eu/es/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/eurovoc>

¹⁰ <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

¹¹ <http://www.w3.org/2000/01/rdf-schema/>

4. MODELADO PARA LOS RECURSOS LINGÜÍSTICOS EN TERESIA

Como se ha mencionado anteriormente, el portal de TeresIA cuenta con visibilizar tanto recursos lingüísticos (datos) como herramientas. Tal y como se puede apreciar en la Figura 2, META-SHARE establece un esquema de modelado ligeramente diferente para los dos tipos de recursos, *Data Language Resource* ('Recurso Lingüístico de Datos') para los datos y *Tool Service* ('Herramienta Servicio') para las herramientas. Aunque ambos son subclases de *Language Resource* ('Recurso Lingüístico'). En las siguientes subsecciones, se presentarán el modelado establecido para cada tipo de recurso.

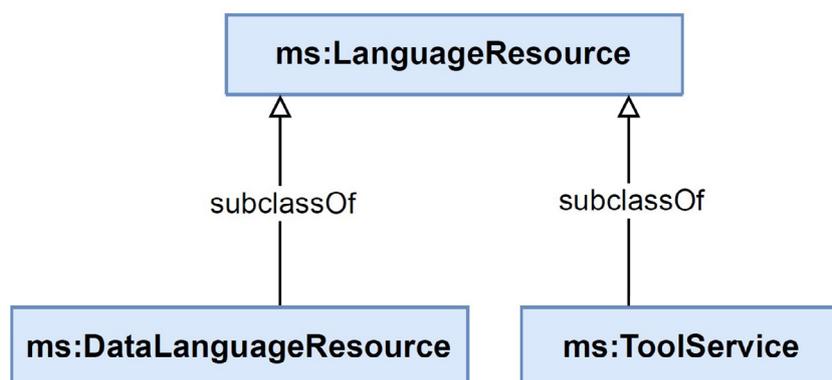


Figura 2: representación datos y herramientas en Meta-Share

4.1. Representación de los metadatos de recursos lingüísticos (datos)

Tal y como se puede apreciar en la **Figura 3**, todos los recursos lingüísticos de datos se pueden representar mediante la clase de META-SHARE (a partir de ahora indicado con el prefijo ms) `ms:DataLanguageResource`. Esta clase cuenta con dos subclases:

1. `ms:LexicoConceptualResource`: esta clase se utiliza para los recursos léxico-conceptuales como por ejemplo los tesauros o las terminologías. Para indicar el tipo concreto de recurso léxico-conceptual que es utilizamos las instancias recogidas en la clase `ms:LCRSubclass`. Estas instancias incluyen: diccionarios, glosarios, terminologías, tesauros, y wordnets.
2. `ms:Corpus`: esta clase se utiliza para representar corpus. Al igual que los recursos léxico-semánticos, podemos especificar el tipo de corpus mediante las instancias definidas en la clase `ms:CorpusSubclass`. Entre dichas instancias se incluyen: corpus anotado y corpus sin anotar.

Aquellos recursos declarados como 'otro' en el formulario, se modelarán sin dar más detalle que `ms:DataLanguageResource`.

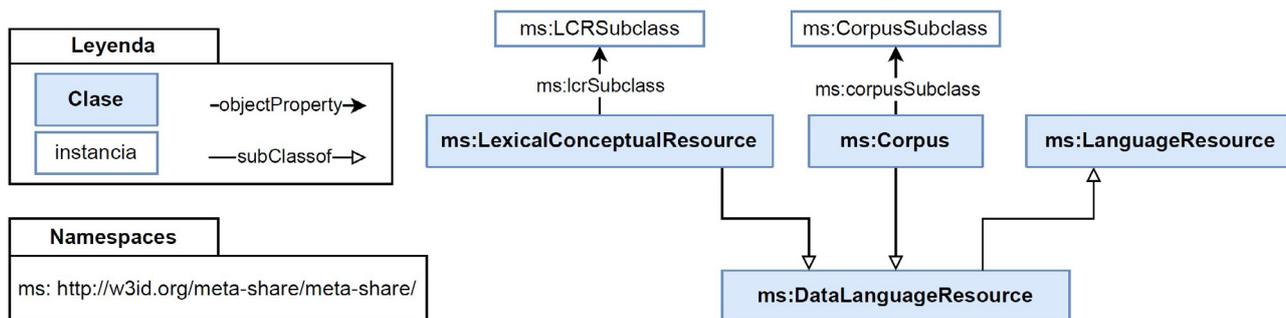


Figura 3: esquema subclases de recursos lingüísticos en Meta-Share

Podemos encontrar varias propiedades relacionadas con `ms:DataLanguageResource`, y, por herencia, con `ms:LexicalConceptualResource` y `ms:Corpus`. La **Figura 4** muestra el esquema completo para la representación de recursos lingüísticos de datos. A continuación, detallamos el modelado de los recursos de datos, explicando el uso de cada propiedad y clase:

1. `ms:resourceName`: se utiliza para indicar el nombre o título del recurso terminológico.
2. `ms:description`: esta propiedad deriva a un texto descriptivo sobre el recurso. Aquí se incluirán los datos obtenidos sobre la estructura de los datos, tipos de relaciones entre términos, la presencia de términos con valor normativo, o información sobre el ciclo de vida de los términos.
3. `ms:creationEndDate`: utilizado para indicar la fecha de creación del recurso.
4. `ms:versionDate`: usado para representar la fecha de la versión del recurso.
5. `ms:email`: mediante esta propiedad indicamos el correo electrónico de contacto provisto por los productores de los recursos.
6. `ms:domain`: esta propiedad se utiliza para la representación del dominio del recurso, y deriva a la clase `skos:Concept`. En concreto, se derivará a un `skos:Concept` de la versión SKOS de Eurovoc, accesible a través de una URI. Por ejemplo, para el dominio 'CIENCIA' se derivará a la URI correspondiente para ese concepto, es decir, se derivará a <http://eurovoc.europa.eu/100151>. Siguiendo el esquema SKOS, a partir de un concepto podemos acceder a otros más amplios (a través de `skos:broader`) o más detallados (mediante `skos:narrower`). Asimismo, podemos acceder al esquema de dominios de Eurovoc¹² a través de `skos:inScheme`, y podemos acceder a todos los dominios de primer nivel en el esquema de dominios a través de `skos:hasTopConcept`.
7. `ms:language`: empleado para indicar los idiomas incluidos en el recurso. Esta clase, a través de la propiedad `ms:usedInRegion`, puede derivar a `dct:Location`, con la cual se especifica la región de uso del idioma.

¹² Esquema de dominios de Eurovoc (SKOS): <http://eurovoc.europa.eu/domains>

8. `ms:mediaPart`: se utiliza para registrar la modalidad de los datos del recurso. Para registrar la modalidad se utilizarán las instancias o individuos definidos por META-SHARE en la clase `ms:MediaType`. Las modalidades incluidas son: audio, texto, imagen, y video.
9. `ms:isRelatedTo`: se utiliza para derivar a otra clase, en este caso `ms:Project`, para indicar el proyecto I+D al que está asociado el recurso.
10. `ms:resourceCreator`: se emplea para indicar el autor del recurso.
11. `ms:iprHolder`: se utiliza para modelar el propietario del recurso.
12. `ms:hasOriginalSource`: esta propiedad apunta a otro recurso, el cual se modelará como `ms:LanguageResource`, y se utiliza para representar las fuentes del recurso que está siendo detallado.
13. `ms:distribution`: esta propiedad deriva a la clase `ms:DatasetDistribution`, mediante la cual representamos los datos relacionados con la distribución del recurso. Esta clase tiene varias propiedades asociadas:
 - a) `ms:fee`: se usa para indicar la tarifa del recurso.
 - b) `ms:distributionLocation`: mediante esta propiedad indicamos el enlace al recurso.
 - c) `ms:size`: a través de esta propiedad, registramos el peso del archivo. Para ello, desde la clase `ms:Size`, modelamos el peso numérico con la propiedad `ms:sizeText` y la unidad a través de `ms:sizeUnit`. Para indicar la unidad, escogemos entre las instancias marcados por META-SHARE en la clase `ms:SizeUnit`, entre los cuales se incluyen: términos, conceptos, byte (B), kilobyte (KB), megabyte (MB), gigabyte (GB), y terabyte (TB).
 - d) `ms:format`: se utiliza para representar el formato del recurso (por ejemplo, CSV o TBX) y deriva a un concepto de SKOS (es decir, `skos:Concept`).
 - e) `ms:distributionForm`: se usa para indicar la forma de distribución del recurso (por ejemplo, descarga) y deriva a un concepto de SKOS (es decir, `skos:Concept`).
 - f) `ms:licenceTerms`: Se utiliza para representar la información relacionada con la licencia y términos de uso. Deriva a la clase `ms:LicenceTerms`, y a través de esta clase podemos representar el nombre de la licencia (`ms:LicenceTermsName`), el enlace a la licencia (`ms:LicenceTermsURI`), y una disponibilidad general mediante las instancias de `ms:LicenceCategory` (es decir, disponible, restringido o académico).

4.2. Representación de los metadatos de las herramientas

Las herramientas se modelan mediante la clase `ms:ToolService`. Varias de las propiedades asociadas a las herramientas son las mismas que se utilizan para los recursos de datos, como por ejemplo, el nombre del recurso (`ms:resourceName`), el correo electrónico de contacto (`ms:email`), el proyecto I+D asociado (`ms:isRelatedToProject`), el propietario de la herramienta (`ms:iprHolder`), el creador de la herramienta (`ms:resourceCreator`). En cuanto a las diferencias:

1. `ms:description`: al igual que con los datos, las herramientas cuentan con un campo de descripción, pero en este caso se incluirán datos como el uso de la herramienta.
2. `ms:version`: esta propiedad se utiliza para saber el nombre de la versión de la herramienta.

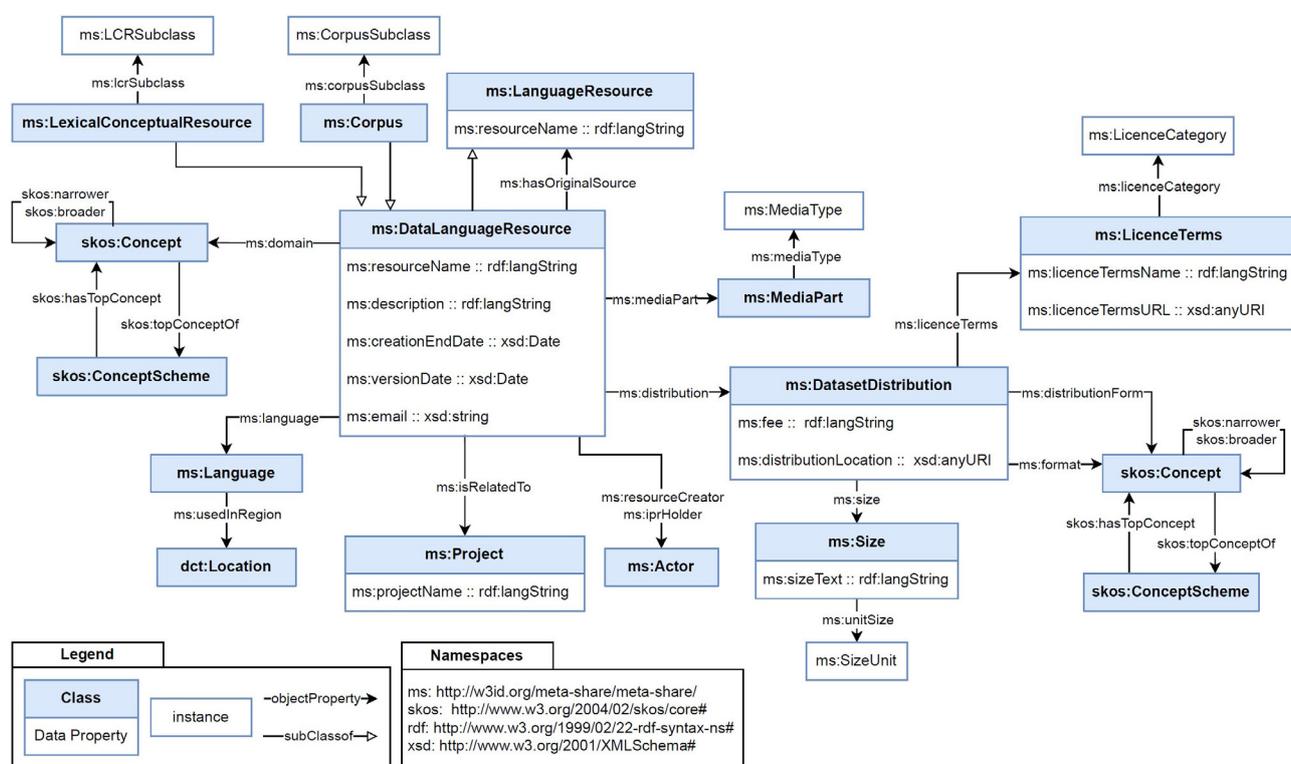


Figura 4: esquema metadato de recursos lingüísticos (datos)

3. `ms:distribution`: para las herramientas, esta propiedad apunta a la clase `ms:SoftwareDistribution`, con la cual se especifican los datos relacionados a la distribución de la herramienta. Las propiedades asociadas a esta clase son:
 - a) `ms:fee`: usada para especificar la tarifa para el uso de la herramienta.
 - b) `ms:distributionLocation`: se utiliza para indicar el enlace de acceso a la herramienta.
 - c) `ms:distributionForm`: mediante esta propiedad indicamos la forma de distribución de la herramienta. Para ello, utilizamos las instancias consideradas

en la clase `ms:SoftwareDistribution`, las cuales incluyen: servicio web, imagen Docker, código ejecutable, plugin, o librería.

- d) `ms:licenceTerms`: se utiliza para representar la información relacionada con la licencia y términos de uso. Al igual que con los datos, esta propiedad deriva a la clase `ms:LicenceTerms`, mediante la cual podemos representar: el nombre de la licencia (`ms:licenceTermsName`), el enlace a la licencia (`ms:licenceTermsURI`), y una disponibilidad general mediante las instancias de `ms:LicenceCategory` (disponible, restringido o académico).

5. APLICACIÓN DEL MODELO DE DATOS

El modelo de representación de datos creado en esta tarea, se utilizará para convertir en datos enlazados la información obtenida a través del formulario. Para dicha conversión, se contemplan varios pasos (ver **Figura 6**):

- 1. Limpieza de datos o preprocesado:** Los datos obtenidos a través de la plataforma Jot-Form están disponibles en un CSV en el que cada columna corresponde a una de las preguntas en el formulario. Estos datos, sin embargo, necesitan pasar por un proceso de limpieza para separar o juntar algunos datos. Por ejemplo, varios de los campos se unen para crear la descripción mientras que los datos de idioma y región de uso deben separarse. El archivo preprocesado o limpio que se genera es un XML.

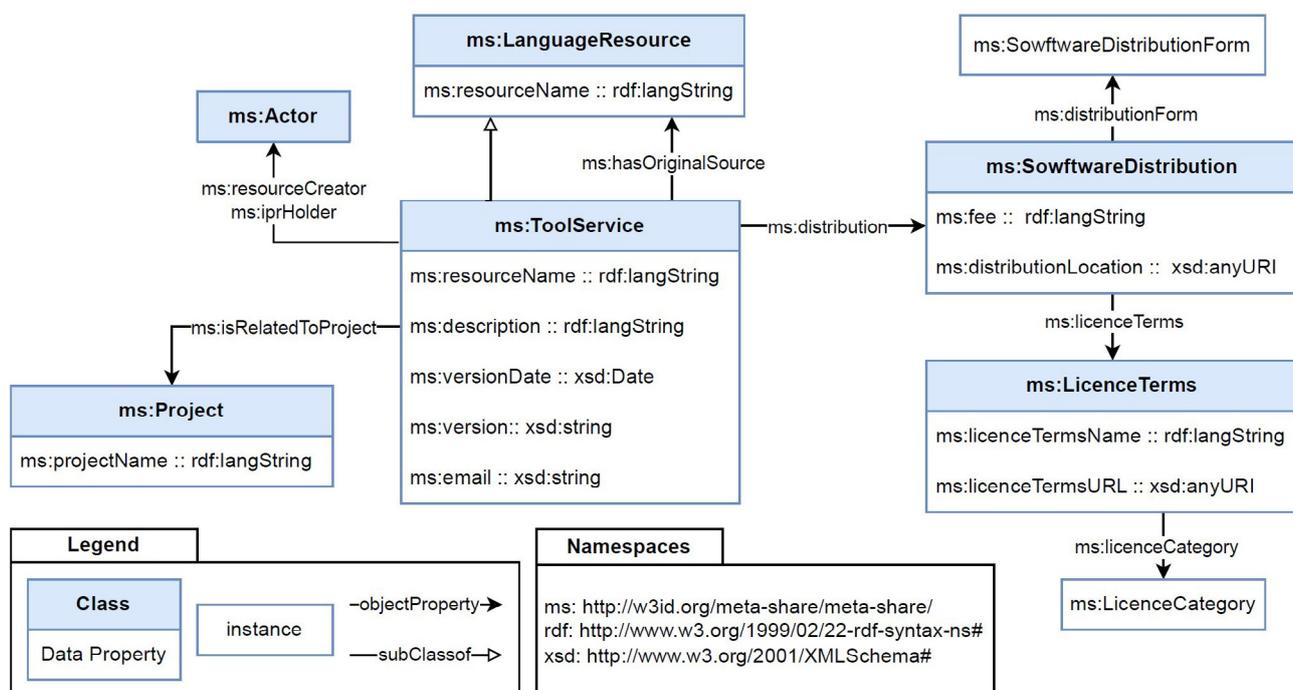


Figura 5: esquema metadato de herramientas

- 2. Mapeo de datos:** Para generar el archivo RDF a partir de los datos en XML, se necesita un fichero de mapeo, es decir, un fichero de equivalencias entre un formato y otro. Para la creación del fichero de mapeo, usamos la herramienta Mapeathor ([Iglesias-Molina et al., 2020](#)), la cual se compone de cinco hojas de Google Spreadsheet. En la primera página, se declaran las ontologías que se usarán. En este caso, siguiendo el modelo de representación creado, se usarán META-SHARE, SKOS, DCTerms y RDFSchema. En la segunda página, se señalan los sujetos de las tripletas, adjudicándoles su clase y URI. En la tercera página, se indica la fuente de datos de los sujetos (en nuestro caso, el fichero XML con los datos limpios) junto con el iterador del sujeto. En la cuarta página, se adjudican los predicados y objetos de las tripletas. Finalmente, en la quinta página se pueden crear funciones, aunque esta característica no ha sido empleada en la conversión de estos datos. Una vez rellenas las hojas del fichero Google Spreadsheet, se ejecuta el comando de Mapeathor para obtener el fichero de mapeo.
- 3. Conversión a RDF:** Finalmente, se utiliza la herramienta RMLMapper¹³ para generar los datos en RDF basándose en los datos limpios, las ontologías y el fichero de mapeo.

Una vez convertidos los datos a RDF, se cargarán los datos en una triple-store, para habilitar la opción de hacer consultas SPARQL. Mediante estas consultas se pueden filtrar los recursos según los campos de interés, como por ejemplo autor, idiomas, etc. En el Anexo C, se muestran algunos ejemplos de consultas SPARQL. Asimismo, los datos en RDF pueden cargarse en herramientas como GraphDB¹⁴ para su visualización en grafo. Por ejemplo, la [Figura 7](#) muestra el grafo que se genera para el Recurso 1 (nodo rojo). De este nodo obtenemos información como los idiomas (círculos verdes claros a la izquierda). También podemos expandir los nodos asociados al nodo del recurso, como por ejemplo, el nodo naranja (resource1_distribution) que contiene la información relativa a la distribución del recurso. Asimismo, en la ventana de la derecha, podemos ver la información de carácter literal (textual), como por ejemplo el nombre del recurso (Panace@) o su descripción.

Los grafos que se generan son interactivos y se pueden navegar expandiendo y ocultando los diferentes nodos, buscando la información que al usuario le interese. Por ejemplo, partiendo del Recurso 1, podemos enfocarnos en los idiomas y encontrar otros recursos que incluyan los mismos idiomas, tal y como se puede observar en la [Figura 8](#).

¹³ <https://github.com/RMLio/rmlmapper-java>

¹⁴ <https://graphdb.ontotext.com/>

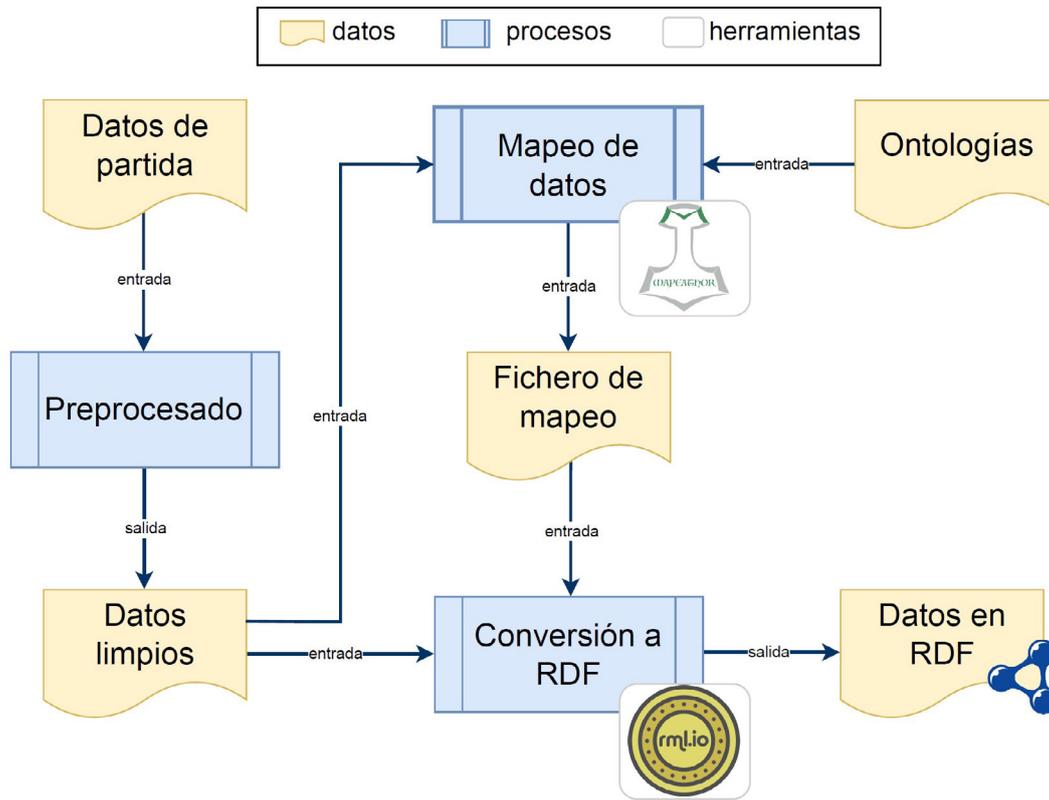
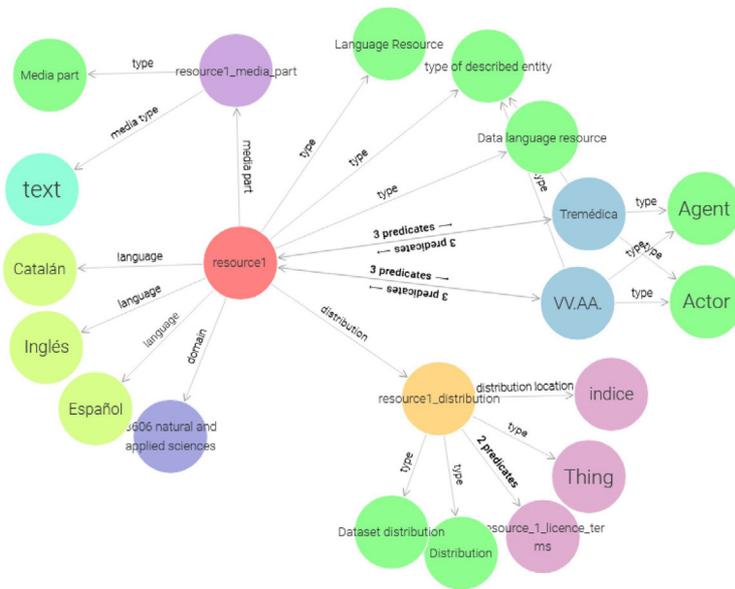


Figura 6: flujo del proceso de conversión a RDF

Visual graph



resource1

resource1

Types:
ms:DataLanguageResource

RDF Rank:
0

Search instance properties

vida de los términos: Sí

ms:resourceName
Panace@

dct:title
Panace@

dct:description
Estructura del recurso: Glosarios con equivalencias inglés > español y explicaciones sobre su uso. Relaciones entre términos: Traducción, sinónimos Presencia de términos con valor normativo: No Incluye información sobre el ciclo de vida de los términos: Sí

owl:topDataProperty
Estructura del recurso: Glosarios con equivalencias inglés > español y explicaciones sobre su uso. Relaciones entre términos: Traducción, sinónimos Presencia de términos con valor normativo: No Incluye información sobre el ciclo de vida de los términos: Sí

Figura 7: grafo del Recurso 1, características del recurso

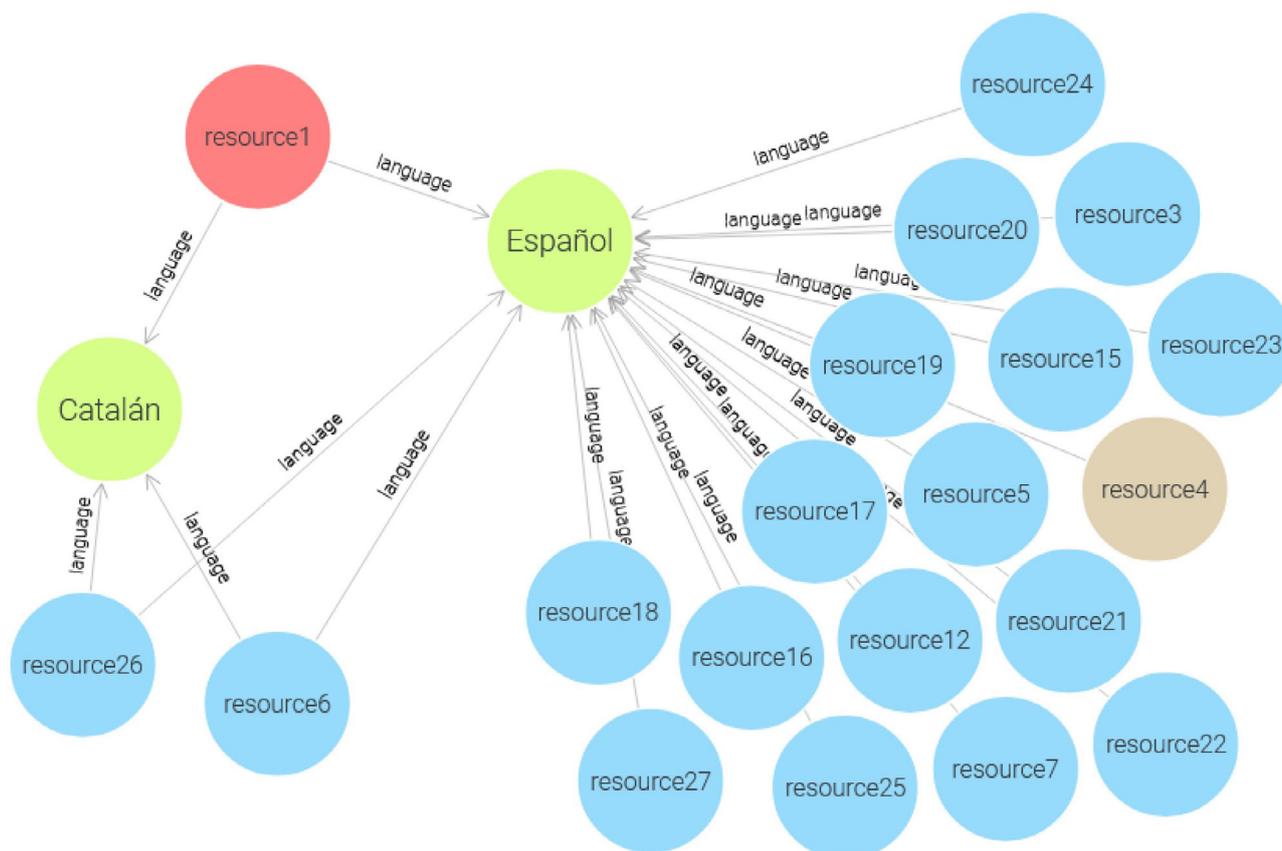


Figura 8: grafo del Recurso 1, enfoque en los idiomas

Referencias

- Chaves-Fraga, D., Óscar Corcho, and Ruckhaus, E. (2022). Guía práctica para la publicación de datos enlazados. <https://datos.gob.es/es/>
- Gavrilidou, M., Labropoulou, P., Desipri, E., Piperidis, S., Papageorgiou, H., Monachini, M., Frontini, F., Declerck, T., Francopoulo, G., Arranz, V., and Mapelli, V. (2012). The META-SHARE metadata schema for the description of language resources. In Calzolari, N., Choukri, K., Declerck, T., Doğan, M. U., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1090-1097, Istanbul, Turkey. European Language Resources Association (ELRA).
- Iglesias-Molina, A., Pozo-Gilo, L., Dona', D., Ruckhaus, E., Chaves-Fraga, D., and Corcho, Ó. (2020). Mapeathor: Simplifying the specification of declarative rules for knowledge graph construction. In *International Workshop on the Semantic Web*.
- Montiel-Ponsoda, E. and Gracia, J. (2019). Los datos lingüísticos enlazados y su impacto en la terminología. *Puntoycoma*, (161):19-23.

A. Anexo: Formulario para los recursos de datos

Información básica

Nombre del recurso *

Tipo de recurso *

- Corpus anotado
 Corpus sin anotar
 Diccionario
 Glosario
 Herramienta
 Terminología
 Tesaurus
 WordNet
 Otro

Save

Seguir

Figura 9: primera página del formulario de recursos de datos en TeresIA

Información sobre el contenido del recurso

¿Es un recurso multilingüe?

- Sí
 No

Cobertura lingüística *

Idioma

Seleccionar

Añadir idioma

Pais/región (variedad)

(Opcional)

Estructura de los datos

Ejemplo: Los conceptos/terminos contienen notas y definiciones, ejemplos de uso, etc.

Relaciones entre términos

Ejemplos: Hiperónimo/hipónimo, sinónimo, antónimo, variante terminológica, traducción, capital de, moneda de, etc.

Número de términos

Número de conceptos

Dominio(s) *

Seleccionar

Añadir dominio

¿Contiene términos con valor normativo?

- Sí
 No

¿Contiene información sobre el ciclo de vida de los términos? (términos en desuso, etc.)

- Sí
 No

Fuentes

P. Ej. Boletín Oficial del Estado

Añadir fuente

Atrás

Save

Seguir

Figura 10: segunda página del formulario para los recursos de datos en TeresIA

Aspectos técnicos del recurso

Fecha de creación

Fecha de última actualización

Modalidad de los datos

Peso del recurso

Valor numérico

Unidad

Formato del recurso *

- CSV
 DOCX (Word)
- HTML
 JSON
- ODF (Libre Office)
 PDF
- RDF
 TBX
- XLSX (Excel)
 XML
- Edición física
- Otro

Modalidad de consulta

- API
 Copia física
 Descarga
- Interfaz
 Widget
- Otro

Figura 11: tercera página del formulario para los recursos de datos en TeresIA

Propiedad intelectual

Entidad propietaria *

Autor

Disponibilidad *

Tarifa

Licencia de uso

Especifique la licencia de uso. Por ejemplo, "CC BY-ND 4.0", o añada un enlace.

Enlace al recurso

Contacto

Contacto visible al usuario, de forma que pueda preguntar por las licencias o informar de algún problema, por ejemplo

Proyecto I+D asociado

- Sí
 No

Figura 12: cuarta página del formulario para los recursos de datos en TeresIA

B. Anexo: Formulario para las herramientas

Información básica

Nombre del recurso *

Tipo de recurso *

- Corpus anotado
 Corpus sin anotar
 Diccionario
 Glosario
 Herramienta
 Terminología
 Tesauro
 WordNet
 Otro

Save

Seguir

Figura 13: primera página del formulario de herramientas en TeresIA

Información adicional de la herramienta

Uso de la herramienta

Por ejemplo: herramienta de etiquetado para Name Entity Recognition

Nombre de la versión

Por ejemplo: Mi Herramienta 3.1

Fecha de versión

MM-DD-AAAA



Forma de distribución

- Servicio web
 Plugin
 Imagen docker
 Librería
 Código ejecutable
 Otro

Atrás

Save

Seguir

Figura 14: segunda página del formulario para las herramientas en TeresIA

Propiedad intelectual

Entidad propietaria *

Autor

Disponibilidad *

Tarifa

Licencia de uso

Especifique la licencia de uso. Por ejemplo, "CC BY-ND 4.0", o añada un enlace.

Enlace al recurso

Contacto

Contacto visible al usuario, de forma que pueda preguntar por las licencias o informar de algún problema, por ejemplo

Proyecto I+D asociado

Sí

No

Atrás

Guardar

Enviar

Figura 15: tercera página del formulario para las herramientas en TeresIA

C. Anexo: Ejemplos de consulta SPARQL

C.1. Ejemplo 1

En este ejemplo, el interés de la búsqueda se centra en el dominio. En particular, el dominio de 'ciencia'. Para ello, desarrollamos la consulta SPARQL que se muestra en el Listing 1. Esta consulta muestra cómo se busca el identificador de los recursos (marcado con ?recursoURI) que tengan como dominio el código identificativo del dominio de 'ciencia' mediante su identificador de Eurovoc (100151). Una vez localizados los recursos del dominio de interés, extraemos el nombre del recurso (?nombre_recurso) y el identificador del propietario ?propietarioURI. Asimismo, a partir del identificador del propietario, obtenemos su nombre (?propietario). Finalmente, extraemos el título del dominio (?nombre_dominio), según la etiqueta preferente de Eurovoc. Tal y como muestra la Figura 16, mediante esta consulta sólo se obtienen dos recursos.

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ms: <http://w3id.org/meta-share/meta-share/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX eurovoc: <http://eurovoc.europa.eu/>

SELECT DISTINCT ?nombre_recurso ?propietario ?nombre_dominio {
  ?recursoURI ms:domain eurovoc:100151 ;
    ms:resourceName ?nombre_recurso ;
    ms:iprHolder ?propietarioURI .
  ?propietarioURI rdfs:label ?propietario .
  eurovoc:100151 skos:prefLabel ?nombre_dominio .
  FILTER (lang (?nombre_dominio) = 'es') }

```

Listing 1: consulta SPARQL para extraer los recursos del dominio 'Ciencia'

	nombre_recurso	propietario	nombre_dominio
1	"Cercatern"	"TERMCAT"	"36 CIENCIA"@es
2	"bUSCatermos"	"Universidade de Santiago de Compostela"	"36 CIENCIA"@es

Figura 16: resultados de la consulta SPARQL para extraer los recursos del dominio 'Ciencia' (Ejemplo 1)

C.2. Ejemplo 2

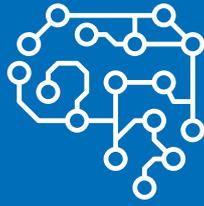
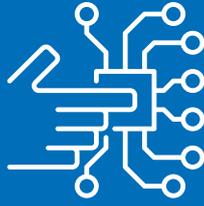
Suponiendo que los resultados obtenidos en el ejemplo anterior no fuesen suficientes, se podría crear una consulta más compleja para tener en cuenta los subdominios del dominio de la búsqueda ('ciencia'), como por ejemplo, la consulta SPARQL que se muestra en el Listing 2. Mediante esta consulta se buscan los identificadores de los subdominios (?subdominioURI) de 'ciencia' mediante su código de Eurovoc (100151). Con los identificadores obtenidos, buscaremos los identificadores de los recursos (?recursoURI) que tienen como dominio los identificadores de los subdominios (?subdominioURI), y obtendremos el nombre de dichos recursos (?nombre_recurso). Finalmente, sacamos el nombre de los subdominios (?dominio) en español. Con esta consulta, obtenemos más recursos, como se puede observar en la Figura 17.

```
PREFIX ms: <http://w3id.org/meta-share/meta-share/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX eurovoc: <http://eurovoc.europa.eu/>
PREFIX euvoc: <http://publications.europa.eu/ontology/euvoc#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT DISTINCT ?nombre_recurso ?propietario ?dominio {
  ?subdominioURI euvoc:domain eurovoc:100151 .
  ?recursoURI ms:domain ?subdominioURI ;
    ms:resourceName ?nombre_recurso ;
    ms:iprHolder ?propietarioURI .
  ?propietarioURI rdfs:label ?propietario .
  ?subdominioURI skos:prefLabel ?dominio .
  FILTER (lang (?dominio) = 'es') }
```

Listing 2: consulta SPARQL para extraer los recursos del subdominio 'Ciencia'

	nombre_recurso	propietario	dominio
1	"Panace@"	"Tremédica"	"3606 ciencias naturales y aplicadas"@es
2	"Formulación y nomenclatura química en español"	"Dialnet"	"3606 ciencias naturales y aplicadas"@es
3	"Diccionario panhispánico de términos médicos (DPTM)"	"Real Academia Nacional de Medicina de España"	"3606 ciencias naturales y aplicadas"@es
4	"Dicciomed, Diccionario médico biológico histórico y etimológico"	"Ediciones Universidad de Salamanca"	"3606 ciencias naturales y aplicadas"@es
5	"Catálogo de Normas UNE"	"Asociación Española de Normalización"	"3606 ciencias naturales y aplicadas"@es
6	"Dicciomed, Diccionario médico biológico histórico y etimológico"	"Ediciones Universidad de Salamanca"	"3611 humanidades"@es

Figura 17: resultados de la consulta SPARQL del Ejemplo 2



TerésIA

Terminología + IA en español