

OntoVC: vocabulario compartido para la interoperabilidad semántica en repositorios de objetos de aprendizajes

Valeria C. Sandobal Verón¹, Mariel A. Ale², M. de los Milagros Gutiérrez²

¹Grupo de Investigación Educativa sobre Ingeniería (GIESIN), Facultad Regional Resistencia, Universidad Tecnológica Nacional, Resistencia, H3500CHJ, Argentina

²Centro de Investigación y Desarrollo de Ingeniería en Sistemas de Información (CIDISI), Facultad Regional Santa Fe, Universidad Tecnológica Nacional, Santa Fe, S3004EWB, Argentina

vsandobal@frre.utn.edu.ar, {male,mmgutierrez}@frsf.utn.edu.ar

Abstract. *At the last decades Learning Objects Repositories have had an explosive growth given the state policies that favored their development in the context of open access. However, this growth was not ordered, resulting in the use of different metadata standards, different policies for filling them, different interpretations, etc. As a consequence, the interoperability between Learning Object Repositories was weakened, mainly at the semantic level. In this article we propose the implementation of a shared vocabulary represented by an ontology, which is validated based on queries made to it.*

Resumen. *En las últimas décadas los Repositorios de Objetos de Aprendizaje han tenido un crecimiento explosivo dado las políticas de estado que favorecieron su desarrollo en el contexto del acceso abierto. Sin embargo, este crecimiento no fue ordenado, dando como resultado el uso de diferentes estándares de metadatos, diferentes políticas de llenado de los mismos, diferentes interpretaciones, etc. Como consecuencia se vio debilitada la interoperabilidad entre los Repositorios de Objetos de Aprendizaje, principalmente a nivel semántico. En este artículo se propone la implementación de un vocabulario compartido representado por una ontología, la cual es validada a partir de consultas realizadas a la misma.*

1. Introducción

La crisis del modelo tradicional de comunicación científica surge debido a los excesivos costos que los investigadores deben pagar para poder publicar los resultados de sus investigaciones en revistas científicas y el monopolio en pocas empresas editoriales, lo que conllevaba a la dificultad de los autores de acceder a los propios resultados de investigación. Tal como lo manifiesta Suber (2012): “el problema de fondo es que contribuimos con nuestro tiempo y dinero público para crear nuevo conocimiento y luego el control de los resultados lo ejercen empresas que creen, correcta o incorrectamente, que sus ingresos y su supervivencia dependen de limitar el acceso a ese conocimiento”. Ante esta paradoja en donde los investigadores publican sus resultados de investigación, rea-

lizan evaluaciones de sus pares sin recibir compensación alguna, y las editoriales cobran acceso, surgen las iniciativas de Acceso Abierto (AA).

Las iniciativas de Budapest (2001), Bethesda (2003), Berlín (2003) y la Declaración de Salvador (2005) pretenden instalar entre los investigadores las premisas de exponer públicamente sus publicaciones, poder compartir, reutilizar y generar servicios a partir de los recursos compartidos. Ante estas iniciativas surgen los repositorios digitales de AA. Un repositorio digital es “una colección de objetos digitales basada en la Web, de material académico producido por los miembros de una institución (o varias) con una política definida, cuyas características más importantes son: auto-archivo, interoperabilidad, libre acceso y preservación a largo plazo” [Alonso-Arévalo y colab. 2008]. Aquí se pueden identificar algunas de las características que el movimiento de AA persigue donde toma gran importancia el libre acceso y la preservación a largo plazo teniendo en cuenta que los objetos digitales en general se alojan en servidores que con el tiempo pueden perder sus referencias, como así también cambiar el formato en el que fueron alojados. Con este aspecto se busca que no se pierda a través del tiempo el objeto digital.

A las iniciativas de AA se van sumando las universidades e instituciones educativas, por lo que surgen los repositorios institucionales (RI), los cuales pueden definirse como “un conjunto de servicios que ofrece la universidad a los miembros de la comunidad para la gestión y disseminación de materiales digitales creados por la misma institución y sus miembros. Es esencialmente un compromiso organizacional a la custodia de estos materiales digitales, incluyendo su preservación a largo plazo, organización, acceso y distribución” [Lynch 2003]. En este caso se hace especial énfasis en la difusión de las producciones realizadas por los miembros de una misma institución, ya sea material académico o de investigación, buscando organizar este material como así también ofrecer servicios de gestión y preservación. Los RI cuentan con políticas institucionales definidas por cada una, donde se describen las formas de depósito, el manejo de los derechos de autor, las herramientas de preservación y el proceso de aceptación del material que se publica y difunde a través del repositorio. Los beneficios de la implementación de un RI son principalmente la posibilidad de acceder libre y gratuitamente a los trabajos de sus miembros, ampliar la visibilidad de estos trabajos a nivel nacional, regional e internacional, brindar al alumno fuentes de consulta confiable a través del acceso a tesis, proyectos finales, material de cátedra, entre otros.

De este material digital almacenado en los RI, surgen los que se denominan objetos de aprendizaje (OA). Los OA pueden definirse como “cualquier recurso digital que puede ser utilizado para la enseñanza” [Wiley 2003]. En este punto, se puede decir que es un tipo especial de objeto digital, donde el OA se considera una pieza de software que presenta tres componentes: un objetivo educativo, material instruccional que permita cumplir con el objetivo y una evaluación que permita verificar fehacientemente que el objetivo ha sido cumplido. [Barritt and Alderman 2004]. De estos OA se requiere que tengan como características la *accesibilidad*: desde el punto de vista que los OA deben ser etiquetados de manera correcta al ser almacenados y puedan ser recuperados de manera fácil en un repositorio; *reusabilidad*: con esta característica se busca que una vez creado el OA, este pueda ser utilizado independientemente del entorno (independiente de la plataforma) en el que se lo aplique; e *interoperabilidad*: se pretende que los OA puedan ser usados, reutilizados y obtenidos independientemente de la forma en que han sido etiquetados y no dependan de los sistemas de gestión de los repositorios en donde han

sido almacenados. [Polsani 2003]. Estos OA son almacenados en repositorios que se denominan repositorios de objetos de aprendizaje (ROA). Los ROA son los sistemas gestores que permiten almacenar, buscar, recuperar y preservar los OA.

En el contexto de los ROA de acceso abierto, la interoperabilidad es la característica más importante sobre la que descansa el principal objetivo del acceso abierto (AA): permitir el libre acceso a las publicaciones científicas en forma gratuita. El valor de estos repositorios se sustenta en la posibilidad de interconexión entre ellos, de manera de poder crear redes de repositorios y de esa forma aumentar la visibilidad, la reutilización de los recursos que en ellos se encuentran no solo por humanos sino también por máquinas [COAR 2012].

Se considera en este trabajo que la *interoperabilidad* es la habilidad de los sistemas para comunicarse con otros intercambiando información, metadatos, y objetos digitales entre ellos con un ida y vuelta en un formato utilizable [Rodrigues 2012].

Pueden definirse varios niveles de interoperabilidad, entre las que podemos mencionar la clasificación propuesta por Garrido Arenas y Lisowska (2014): (i) *infraestructura*, mediante la utilización de protocolos, tales como ISO-OSI y TCP/IP para llevar a cabo el intercambio de datos. (ii) *sintáctica*, entender datos de otros sistemas similares. (iii) *estructural*, existencia de modelos lógicos comunes. (iv) *semántica*: capacidad de los sistemas de información de tener un entendimiento común de los términos que se intercambiarán.

Por otro lado, en [IEEE 2009] se define *interoperabilidad sintáctica* como la capacidad de dos sistemas para que puedan establecer una comunicación e intercambiar información; y la *interoperabilidad semántica* que provee mecanismos necesarios para interpretar correctamente la información recibida de manera automática. En este tipo de interoperabilidad debe tenerse en cuenta los formatos de metadatos, los vocabularios controlados, ontologías y directrices de interoperabilidad. Éste último nivel de interoperabilidad es en el cual se enfoca este artículo.

En particular los metadatos utilizados para describir los OA presentan una gran dificultad, ya que existen actualmente gran variedad de estándares de metadatos, entre los que se puede mencionar: Dublin Core (DC): uno de los más utilizados para la descripción de cualquier tipo de objeto digital; Learning Object Model (LOM) específicamente utilizado para la descripción de OA; Metadata Object Description Schema (MODS), Common European Research Information Format (CERIF), Machine – Readable Cataloging (MARC). Ahora bien, para que los OA pueden ser encontrados y los ROA cumplan con las funciones para las cuales fueron creados es necesario según [García Aretio 2005] que entre los mismo haya lo que se denomina “complementariedad”, y para que esto sea posible los OA deben contar con metadatos que permitan identificarlos y además faciliten su indexación.

En la búsqueda de soluciones de la interoperabilidad semántica entre ROA ha llevado a proponer e implementar diferentes opciones a través de la utilización de ontologías para la definición de los metadatos. Desde ontologías de alto nivel, como las de Dublin Core (DC)¹, Learning Object Metadata (LOM)² y otros estándares de metadatos; donde si

¹ Ontología Dublin Core (DC): http://bloody-byte.net/rdf/dc_owl2dl/dcterms

² Ontología Learning Object Metadata(LOM): <http://slor.sourceforge.net/ontology/lom.owl>

bien se busca dar una mejor definición de los términos utilizados no resulta suficiente para la interoperabilidad entre ROA dado que se continúa con el problema de la heterogeneidad de metadatos cuando los ROA utilizan diferentes ontologías representando diferentes estándares. Es necesario avanzar un paso adelante que ataque este problema. En este trabajo se presenta un vocabulario compartido que puede ser utilizado para brindar interoperabilidad semántica a distintos ROA aun cuando éstos utilicen diferentes esquemas y ontologías de metadatos.

Este artículo se organiza como sigue. En la sección 2 se describen los trabajos relacionados. En la sección 3 se describe el vocabulario compartido propuesto en este trabajo y un caso de estudio que muestra su aplicabilidad. Finalmente, se presentan las conclusiones y trabajos futuros.

2. Trabajos Relacionados

En la búsqueda de la interoperabilidad entre los ROA cobra gran importancia la correcta selección del estándar de metadatos a implementar. Sin embargo, debe tenerse en cuenta que el protocolo para la recolección de metadatos (OAI-PMH) utiliza como base para su recolección el estándar DC. El protocolo OAI-PMH es un estándar que se considera como una primera iniciativa para la interoperabilidad a nivel sintáctico, y además permite que los OA sean visibles desde diferentes ROA. Este protocolo define dos clases de participantes: los *proveedores de datos* que son los ROA que exponen sus metadatos para que puedan ser cosechados y *proveedores de servicio* que son los encargados de cosechar los metadatos de los ROA y exponerlos en su propio repositorio.

Un avance más hacia la interoperabilidad es la definición de *directrices*, que son un conjunto de normas que se establecen para que los sistemas puedan comunicarse e intercambiar información. A nivel mundial, se encuentra la directriz Driver 2.0 que implementa el protocolo OAI-PMH para la interoperabilidad sintáctica, y para la semántica la utilización de vocabularios controlados [Driver 2008]. A partir de esta directriz surgieron otras, como Open Aire [OpenAire 2013] que toma los principales puntos de Driver, pretendiendo ser una guía para administradores de repositorios, tanto de artículos como de datos científicos. Open Aire obliga a que las publicaciones tengan criterios tales como: estar habilitadas para acceso abierto en texto completo, ser resultado de un proyecto de investigación y contar con un identificador para el proyecto. A nivel nacional, y a partir de la Ley de Creación de Repositorios Digitales [Ley 28699 2013], se definen las Directrices del Sistema Nacional de Repositorios Digitales (SNRD) [Directrices SNRD 2013] que tiene como objetivo principal establecer políticas, estándares y protocolos comunes para los integrantes del sistema. En particular los estándares y protocolos se basan en las directrices Driver 2.0, agregando lineamientos que se consideran pertinentes en busca de la interoperabilidad y la creación de redes de repositorios.

La creación de redes de repositorios constituye uno de los objetivos principales de la Confederación de Repositorios de Acceso Abierto (COAR por sus siglas en inglés), que pretende interconectar a los repositorios existentes obteniendo así una mayor visibilidad para los autores y sus instituciones. En Latinoamérica, surge con fuerza la creación de La Referencia [La Referencia 2015] que se constituye como una red federada de repositorios institucionales de publicaciones científicas. Esta red ha definido los estándares y las políticas que deben utilizar los repositorios que quieran formar parte de la misma. Se

basa principalmente en las directrices Driver 2.0, implementa el protocolo OAI-PMH para la recolección de metadatos, y recomienda la utilización de un conjunto básico de metadatos correspondientes al estándar DC. Como característica principal La Referencia cosecha desde los nodos nacionales, por lo cual los repositorios que no estén adheridos a los mismos no podrán ser tenidos en cuenta para La Referencia. El cosechador a nivel global, se denomina Open Aire [OpenAire 2007], y se diferencia de La Referencia sobre todo en que la cosecha de metadatos la realiza directamente de los repositorios institucionales registrados en OpenDoar; como así también en que no solo toma las publicaciones científicas sino también informes técnicos, material educativo, objetos de conferencia, entre otros.

Si bien las herramientas y propuestas anteriores pretenden subsanar la interoperabilidad entre ROA, resulta necesario buscar otro tipo de solución para lograr la interoperabilidad semántica. La implementación de soluciones ontológicas para subsanar la interoperabilidad semántica ha llevado a proponer diferentes enfoques, es así que se puede identificar tres diferentes direcciones para la implementación de una ontología, a saber: [Stuckenschmidt and Harmelen 2005]

- ontología única o global: en donde se utiliza una ontología global que provee de un vocabulario compartido para la especificación semántica
- múltiples ontologías locales: cada fuente de información es descripta por su propia ontología
- enfoque híbrido: cada una de las fuentes de información tiene su propia ontología y se crea en un nivel superior un vocabulario compartido con términos básicos del dominio

En la literatura se observa que frente a esta problemática se han utilizado con éxito los enfoques híbridos donde la heterogeneidad y la integración de datos se logra con la definición ontológica de un vocabulario compartido. El caso propuesto por [Wang and Ye 2009] establece la utilización de un enfoque híbrido para la integración de información estadística de China, teniendo en cuenta que existen múltiples fuentes de información y se pretende que la misma sea comprendida por la computadora y pueda realizar algún tipo de razonamiento lógico. En este artículo se establecen los pasos a seguir para definir la ontología local, el vocabulario compartido y el mapeo entre los mismos. La implementación de este enfoque tiene por objetivo integrar, compartir y reutilizar la información.

Por otro lado, [Zhang and Ma and Wang 2009], aborda la integración de datos que provee Internet, el cual presenta recursos en diferentes formas y formatos, tales como pueden ser la Bases de Datos relacionales y las orientadas a objetos; o documentos HTML y XML como datos semiestructurados. Esta heterogeneidad no permite utilizar estos datos de manera eficiente, para lo cual proponen implementar una solución ontológica con un enfoque híbrido. En este artículo, al igual que el anterior se describen las etapas de creación del vocabulario compartido, las ontologías locales y las reglas de mapeo entre ellas.

El enfoque híbrido de ontologías es utilizado también por [Zheng and Terpenney 2013] para la gestión de los sistemas de intercambio de obsolescencia. En este artículo se implementan las ontologías locales para describir las estructuras de múltiples fuentes de datos distribuidas en varios sistemas de gestión de obsolescencia, el vocabulario com-

partido u ontología global para la especificación de la semántica del conocimiento del dominio de obsolescencia.

En párrafos anteriores se mencionó la utilización de ontologías para los estándares de metadatos mayormente utilizados. La reutilización de estas ontologías en el marco de un enfoque de ontologías híbridas [Vian y colab. 2011] permite la identificación de términos comunes a estos estándares de metadatos, como así también integrar la información, facilitando la búsqueda y recuperación de objetos de aprendizaje con mayor precisión de lo que puede realizarse actualmente. Esta propuesta plantea un modelo donde los principales componentes son: los repositorios, el sistema multiagente que está compuesto por un Agente Indexador y un Agente Buscador; el mapeo de metadatos (utiliza de prueba dos repositorios uno que utiliza DC y el otro LOM); la ontología de dominio que es un set de ontologías de los tipos de objetos de aprendizaje de los repositorios (como por ejemplo una ontología de seguridad de la información) y un Servicio de Búsqueda que provee una interface.

Si bien, los casos mencionados en este apartado, pretenden solucionar la problemática de la interoperabilidad, nos enfocaremos en particular en la interoperabilidad semántica, en donde el desarrollo de ontologías vislumbra como el camino a seguir. En particular, la definición de enfoques híbridos con ontologías, resulta ser mejor opción para solucionar la interoperabilidad a nivel semántico entre repositorios.

3. Modelo propuesto

A partir de lo desarrollado en la sección anterior, se define una ontología de enfoque híbrido. Donde se describe un vocabulario compartido con los términos básicos y cada repositorio cuenta con una ontología local de acuerdo al estándar de metadatos que implementa como se muestra en la figura 1.

El vocabulario compartido (VC), que forma parte del más alto nivel del enfoque propuesto, tiene sus bases en la ontología DCOntoRep [Sandobal Verón y colab. 2015] que combina los metadatos del estándar DC y las directrices del (SNRD). Teniendo en cuenta que este estudio está enfocado en la interoperabilidad semántica de los ROA, es necesaria la incorporación de metadatos relacionados específicamente con Objetos de Aprendizaje (OA). Dado que DC es útil para describir cualquier tipo de objeto digital (OD) pero carece de descriptores específicos para OA, se incorpora al vocabulario compartido una selección de metadatos propuestos por el estándar LOM, que permiten la correcta y completa descripción de OA [Sandobal Verón y colab. 2016]. La ontología propuesta para el vocabulario compartido se clasifica según su generalidad y lo propuesto por [Guarino y colab. 1994] en una ontología de dominio, teniendo en cuenta que especializa los conceptos provistos por las ontologías de alto nivel de cada uno de los estándares de metadatos utilizados por los ROA. Es decir, modela los metadatos específicos para OA proporcionando mayor detalle a los mismos.

El objetivo del presente trabajo es presentar y validar el vocabulario compartido como una primera aproximación hacia la implementación completa del enfoque híbrido propuesto para lograr la interoperabilidad semántica entre ROA.

3.1 Definición del Vocabulario Compartido

La ontología llamada *DCOntoRep* [Sandobal Verón y colab. 2015] que conceptualiza el conjunto de metadatos del estándar DC, fue enriquecida con cuestiones propias del uso de metadatos en ROA, tal es el caso de la incorporación de los conceptos, reglas y restricciones contempladas en las directrices del SNRD.

Esta ontología fue tomada como base para la definición del vocabulario compartido, de aquí en adelante lo denominaremos *OntoVC*. El vocabulario compartido surge a partir de la implementación de un enfoque híbrido, donde se busca la interoperabilidad de fuentes de datos heterogéneas a partir de la definición de conceptos y términos comunes y básicos de cada una de las fuentes.

El enfoque híbrido propone la utilización de ontologías locales como fuentes de datos, para el caso propuesto será para cada uno de los estándares de metadatos utilizados por los ROA, y una ontología que permite unificar conceptos comunes a través de lo que se denomina vocabulario compartido. Entre las ventajas que presenta un enfoque híbrido, es posible mencionar la facilidad con que las nuevas fuentes de datos pueden incorporarse sin la necesidad de realizar cambios en lo que ya se tiene implementado, sino que sólo requiere la definición del mapeo entre la nueva ontología que se agrega y el vocabulario compartido; el esfuerzo de implementación es razonable, si bien se utilizan múltiples ontologías su uso es simple [Wache y colab. 2001].

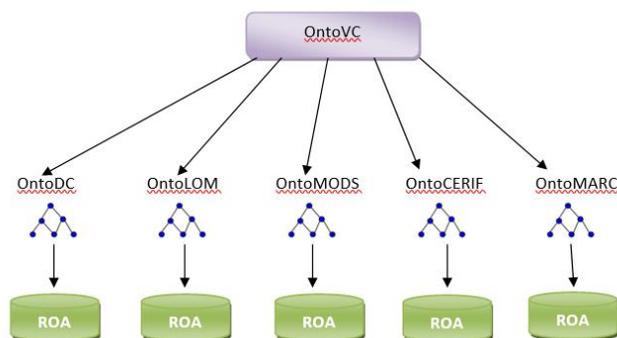


Fig. 1. Enfoque Híbrido

La figura 1 muestra el uso de la ontología *OntoVC* en los repositorios institucionales de AA utilizando el enfoque híbrido. Como se observa, cada repositorio utiliza su propio vocabulario, compartiendo todos ellos el vocabulario común, que se relaciona con cada una de estas ontologías a través de reglas de mapeo. Para la definición de *OntoVC* se siguió la metodología propuesta por [Sandobal Verón 2016], de tres etapas para el desarrollo del vocabulario compartido.

La primera etapa, *Definición del vocabulario compartido*, tiene a su vez tres subetapas: (i) analizar las fuentes de datos y comprender la información almacenada y las formas de expresión, (ii) seleccionar términos y conceptos que serán contemplados en el voca-

bulario compartido y (iii) definición de la ontología global o vocabulario compartido. Esta primera etapa es la que se desarrolla en el presente artículo.

La segunda etapa es la *Definición o creación de las ontologías locales*. Para nuestro caso las ontologías de los estándares de metadatos que utilizan los ROA ya han sido definidos y se toman los mismos como referencia.

La tercera etapa consiste en el *Mapeo* entre las ontologías locales y la ontología global o vocabulario compartido. Esta última etapa será desarrollada como trabajo futuro.

El desarrollo de la primera etapa consiste en:

- Analizar las fuentes de datos y comprender la información almacenada y las formas de expresión. Para poder llevar a cabo esta sub-etapa se consideraron cuestiones comunes en los diferentes estándares utilizados. Principalmente se tuvieron en cuenta los estándares LOM, DC, Metadata Object Description (MOD), Machine Readable Cataloging (MARC) entre otros.
- Seleccionar los términos y conceptos que serán contemplados en el vocabulario compartido. La definición de los conceptos se obtuvo de analizar la utilización de diferentes nombres para identificar una misma cosa, clasificaciones diferentes para codificaciones similares, mayor especificidad y detalles presentes en algunos estándares que resulta útil para la correcta descripción de los OA, entre otras cuestiones.
- Definición de la ontología global o vocabulario compartido: de acuerdo a los conceptos definidos en la etapa anterior se diseña OntoVC.

Estas últimas dos etapas se han llevado en conjunto, se muestra a continuación ejemplos, que permiten entender el trabajo realizado.

Para identificar al autor, se utilizó el metadato *creator* en la estructura propuesta por el vocabulario compartido, se considera como clase y tiene su equivalente para DC, LOM, MOD y MARC (Figura 2.a), como puede verse el nivel de profundidad en las estructuras es diferente como así también la manera de guardar el nombre y el apellido del autor, en algunos casos como DC, LOM y MOD se han puesto en una sola etiqueta de metadato (en algunos casos se utilizan dos instancias del mismo metadato como el caso de DC) y por otro lado en etiquetas de metadatos separados para MARC y DC en una misma instancia.

Para el caso de la etiqueta *Title* y *Subtitle*, se definieron en OntoVC clase y subclase respectivamente, donde se agrega además la relación *SubTitleOf* teniendo en cuenta las directrices del SNRD. Ambas clases están relacionadas con las etiquetas *Title* de DC y en caso de que el OA cuente con un subtítulo se crea otra instancia de la etiqueta *Title*. Para el caso de LOM se guarda este metadato en la estructura *General/Title*. El caso de MOD presenta una estructura similar a OntoVC, pero con la siguiente estructura *Title Information/Title* y *Title Information/Alternative Title*. Una aproximación a lo propuesto en OntoVC lo tiene MARC a través de la siguiente estructura *Title and Title-Related Fields/General Information/Key Title*, donde *Key Title* representa el título principal y *Title* y *Title-Related Fields/General Information/Abbreviated Title* pueden utilizarse como subtítulo (Figura 2.b).

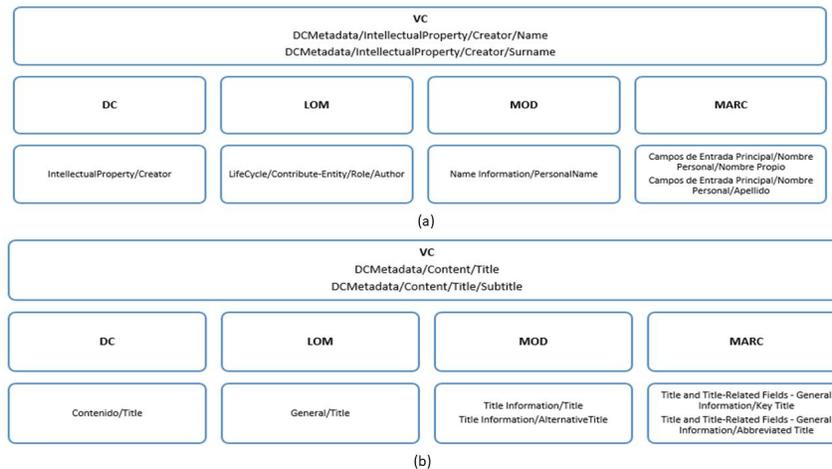


Fig. 2. (a) Metadatos *Creator*, (b) metadato título y subtítulo y nivel en la estructura para OntoVC y los estándares DC LOM, MOD y MARC

En el caso particular de los metadatos educativos que brindan información pedagógica y que por ende describen mejor a los OA (Figura 3), fueron obtenidos y seleccionados del estándar LOM. Si bien, los demás estándares no tienen una descripción específica para este tipo de metadatos, puede mencionarse algunas intersecciones como ser: la etiqueta *ageRange* de OntoVC se relaciona con *Audience Information/Audience* de MOD. Para el caso del estándar DC este metadato podría relacionarse con *Audience*. Para MARC el metadato utilizado es *AudienceCharacteristics*. Ninguno de los estándares analizados propone un rango para este metadato, por lo cual se considera oportuno utilizar el propuesto por LOM, tomando como referencia los siguientes valores: “3-5”, “6-8”, “9-12”, “13-15”, “16-17”, 18 up”. Otro de los metadatos incluidos para mejorar la descripción de un OA es *context* que es utilizado para definir el nivel educativo hacia el cual está dirigido el OA. Para el caso del DC el metadato utilizado es *educationLevel*; en el caso de MOD y MARC no hay una etiqueta que describa esta característica por lo cual agregaría valor a los OA utilizados en los ROA con estos estándares.

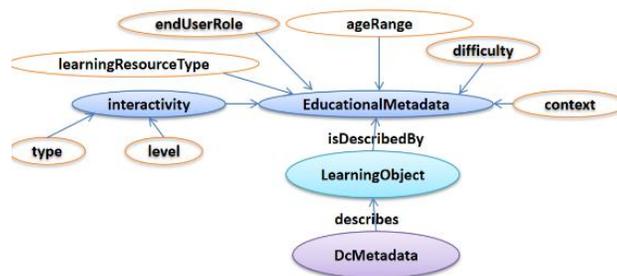


Fig. 3. Metadatos del estándar LOM agregados a OntoVC

A partir de estas consideraciones y siguiendo con las directrices del SNRD, como se ha venido trabajando desde la primera propuesta de ontología, se plantea la ontología que se muestra en la Figura 4 para representar el vocabulario compartido. Como puede observarse toma como base el estándar de metadato DC que es el más utilizado por los

ROA, y se incorporan metadatos de LOM que brindan mayor información sobre OA, permitiendo realizar búsquedas con un mejor nivel de detalle.

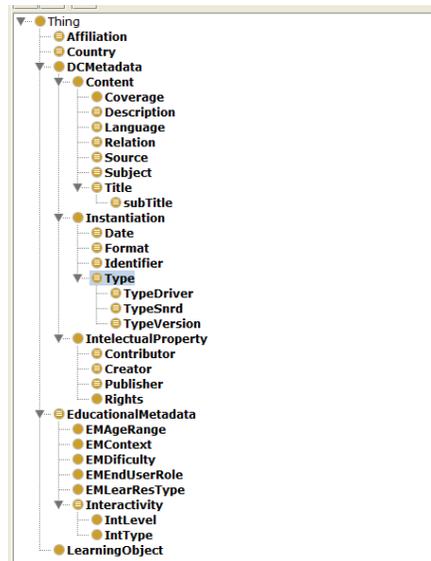


Fig. 4. Vista de las clases de OntoVC desde la herramienta Protege

Para poder representar las restricciones impuestas en las directrices del SNRD en cuanto a los valores que pueden tomar los metadatos y las relaciones entre los mismos, se definieron en la ontología reglas de integridad en SWRL (Semantic Web Rule Language), que limitan la forma en que los metadatos se relacionan. Así, si un OA es del tipo *Artículo* según la clasificación driver y *Article* para el SNRD, debe ser un tipo de Interacción *Expositive* y como tipo de Recurso Educativo *Lecture*, como muestran las reglas (1) y (2).

$$\text{LearningObject}(?lo) \wedge \text{TypeDriver} (?d, \text{"Article"}) \wedge \text{hasTypeDriver} (?lo, ?d) \wedge \text{TypeSnrd} (?s, \text{"Artículo"}) \wedge \text{hasTypeSnrd} (?lo, ?s) \rightarrow \text{IntType} (?t, \text{"Expositive"}) \quad (1)$$

$$\text{LearningObject} (?lo) \wedge \text{TypeDriver} (?d, \text{"Article"}) \wedge \text{hasTypeDriver} (?lo, ?d) \wedge \text{TypeSnrd} (?s, \text{"Artículo"}) \wedge \text{hasTypeSnrd} (?lo, ?s) \rightarrow \text{EMLearnResType} (?lo, \text{"Lecture"}) \quad (2)$$

Ahora bien, si un OA es del tipo driver *Conference Object* y SNRD *Documento de Conferencia*, también el tipo de interacción asociado debe ser *Expositive* y el tipo de Recurso Educativo *Slide*, como se define en las reglas (3) y (4).

$$\text{LearningObject} (?lo) \wedge \text{TypeDriver} (?d, \text{"Conference Object"}) \wedge \text{hasTypeDriver} (?lo, ?d) \wedge \text{TypeSnrd} (?s, \text{"Documento de Conferencia"}) \wedge \text{hasTypeSnrd} (?lo, ?s) \rightarrow \text{IntType} (?lo, \text{"Expositive"}) \quad (3)$$

$$\text{LearningObject} (?lo) \wedge \text{TypeDriver} (?d, \text{"Conference Object"}) \wedge \text{hasTypeDriver} (?lo, ?d) \wedge \text{TypeSnrd} (?s, \text{"Documento de Conferencia"}) \wedge \text{hasTypeSnrd} (?lo, ?s) \rightarrow \text{EMLearnResType} (?lo, \text{"Slide"}) \quad (4)$$

Teniendo en cuenta el nivel de interactividad que propone LOM, un OA puede tomar los siguientes valores: *very high*, *high*, *medium*, *low* y *very low*. Así, por ejemplo, si un OA es del tipo de interacción *Expositive* y el tipo de Recurso Educativo es *Lecture*, le co-

responde un nivel de interactividad *Very Low*, esto puede expresarse mediante la regla (5), teniendo en cuenta además las reglas (1) y (2).

$$\text{LearnigObject(?lo)} \wedge \text{IntType(?i, "Expositive")} \wedge \text{hasIntType(?lo,?i)} \wedge \text{EMLearnResType(?l, "Lecture")} \wedge \text{hasLRType(?lo,?l)} \rightarrow \text{IntLevel(?lo, Very Low)} \quad (5)$$

En el caso de los OA definidos a través de las reglas (3) y (4) le corresponde el nivel de interacción *Low*, ya que puede considerarse como una pequeña participación del usuario al pasar las *slides*, como lo expresa la regla (6).

$$\text{LearnigObject(?lo)} \wedge \text{IntType(?i, "Expositive")} \wedge \text{hasIntType(?lo,?i)} \wedge \text{EMLearnResType(?l, "Slide")} \wedge \text{hasLRType(?lo,?l)} \rightarrow \text{IntLevel(?lo, Low)} \quad (6)$$

También, existe una relación directa entre los contextos a los cuales van dirigidos los OA y la etiqueta rango de edad. Entonces, por ejemplo, para un contexto *Higher Education* se considera que el rango de edad es *18 up*, como muestra en la regla (7).

$$\text{LearnigObject(?lo)} \wedge \text{EMContext(?c, "Higher Education")} \wedge \text{hasEMContext(?lo,?c)} \rightarrow \text{EMAgeRange(?lo, "Up 18")} \quad (7)$$

3.2 Caso de Estudio

Desde sus orígenes, y con la primera versión de la ontología OntoDC y el enriquecimiento que se proponen en el presente artículo con la ontología OntoVC, se tiene como objetivo principal optimizar las búsquedas realizadas en los repositorios institucionales de acceso abierto independientemente del estándar de metadatos que utilicen. Para ello se han propuesto preguntas de competencias que permitirán validar la ontología, obteniendo como resultado una ontología contextualizada. Algunas preguntas de competencia (PC) sugeridas son:

PC 1: ¿Cuál es el título de los OA que tienen como tipo SNRD X?

PC 2: ¿Cuál es el tipo de recurso, usuario destinatario, rango de edad y tipo de interacción para el OA X?

PC 3: ¿Cuáles son los OA que tienen un nivel de interacción X?

PC 4: ¿Cuáles son los OA que tienen una dificultad X?

PC 5: ¿Cuáles son los OA que presentan un tipo de recurso educativo X?

Las preguntas de competencia sugeridas permiten en una primera instancia realizar búsquedas para armar una secuencia didáctica completa, teniendo en cuenta la edad de los destinatarios, la complejidad que se busca pueda tener el material presentado, como así también el nivel de interacción del destinatario con el OA.

Con el objetivo de validar la ontología OntoVC, buscando respuestas a las preguntas de competencias sugeridas con anterioridad, se pobló la misma con ejemplos de OA. Cabe aclarar, que se está trabajando en el desarrollo de una aplicación del tipo cosechador que permita extraer los metadatos de repositorios institucionales de accesos abierto con los cuales se poblará la ontología propuesta.

Las instancias creadas permiten realizar consultas SPARQL³(SPARQL Protocol and RDF Query Language), se muestra un ejemplo en la Figura 5, donde se crea el OA cuyo

³ SPARQL:<https://www.w3.org/TR/rdf-sparql-query/>

título es *SIIE2016*, su identificador es el número *123*, según la clasificación de driver es del tipo *Object Conference*, para el SNRD *Documento de Conferencia* y la versión que muestra es la de *Accepted*. Desde el punto de vista de los metadatos educativos y desde la interacción se considera como del tipo *Expositive* y nivel *Low*. Por el lado del tipo de recurso educativo es *Slide*, considerándose que el contexto para el que va dirigido es *Higher Education*, y la dificultad es *Easy*. Este OA tiene como usuario final a los estudiantes, en este caso se clasifica como *Learner*, y el rango de edad considerado es *18 Up*.

Las preguntas de competencias antes descritas permitirán realizar las consultas SPARQL que permiten validar OntoVC. Cabe aclarar que, si bien en la Figura 5 se muestra una instancia completa a modo de ejemplo, se han creado otras como los OA *Interoperabilidad Semántica* y *Web3.0*, a fin de que sirvan para corroborar si los filtros son realizados correctamente.

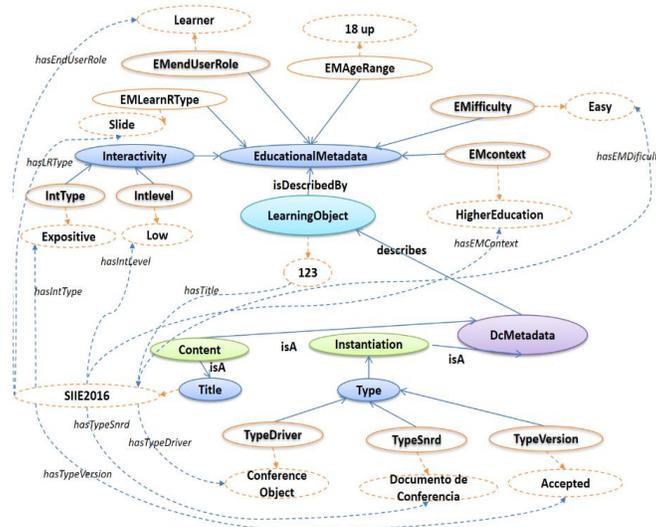


Fig. 5. Instancias creadas para el OA de título SIIE2016

Como respuesta a la PC 1:¿Cuál es el título de los OA que tienen como tipo SRND “X”?, donde el valor de X es “Objeto de Conferencia”. La consulta arrojó como resultado lo que se muestra en la Figura 6.

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX SV: <http://www.semanticweb.org/valeria/ontologies/2017/10/OntoVC#>
SELECT ?description
      WHERE { ?TypeSnrd SV:isTypeSnrdOf ?description.
              FILTER regex(str(?TypeSnrd),'DocumentoConferencia')}

```

| description |
|-------------|
| SIIE2016 |
| Web3.0 |

Fig. 6. Respuesta a la PC 1 con el valor de SNRD Documento de Conferencia

En el caso de la PC 2 se tomó el OA de título SIIE2016, la consulta SPARQL y el resultado se muestran en la Figura 7.

Active Ontology | Entities | Classes | Object Properties | Data Properties | Individuals | OWLViz | DL Query | SPARQL Query

SPARQL query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX SV: <http://www.semanticweb.org/valeria/ontologies/2017/10/OntoVC#>
SELECT ?learningResourceType ?endUserRole ?ageRange ?typeInt
      WHERE { ?Title SV:hasLRType ?learningResourceType.
              ?Title SV:hasEndUserRole ?endUserRole.
              ?Title SV:hasEMAgeRange ?ageRange.
              ?Title SV:hasIntType ?typeInt.
              FILTER regex(str(?Title),'SIIE2016')}

```

| learningResourceType | endUserRole | ageRange | typeInt |
|----------------------|-------------|----------|------------|
| slide | Learner | 18-up | Expositive |

Fig. 7. Respuesta a la PC 2, donde el título del OA que se busca es SIIE2016

Para la PC 3, se tomó como referencia el nivel de interacción Low. La figura 8 muestra el resultado de OA que tienen este tipo de interacción.

| SPARQL query: | |
|--|--|
| <pre> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX owl: <http://www.w3.org/2002/07/owl#> PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX SV: <http://www.semanticweb.org/valeria/ontologies/2017/10/OntoVC#> SELECT ?description WHERE { ?IntLevel SV:isIntLevelOf ?description. FILTER regex(str(?IntLevel), 'Low')} </pre> | |
| description | |
| SIIE2016 | |
| Web3.0 | |

Fig. 8. OA con nivel de interacción Low

Ahora bien, para la PC 4, y valor de dificultad Medium, el resultado se muestra en la Figura 9, se muestra también como resultado los usuarios destinatarios y el nivel de interacción.

| SPARQL query: | | | |
|---|-------------|-------|--|
| <pre> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX owl: <http://www.w3.org/2002/07/owl#> PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX SV: <http://www.semanticweb.org/valeria/ontologies/2017/10/OntoVC#> SELECT DISTINCT ?description ?endUserRole ?level WHERE { ?EMDifficulty SV:isDifficultyOf ?description. ?EMEndUserRole SV:hasEndUserRole ?endUserRole. ?InLevel SV:hasIntLevel ?level. FILTER regex(str(?EMDifficulty), 'Medium')} </pre> | | | |
| description | endUserRole | level | |
| InteroperabilidadSemantica | Learner | Low | |
| Web3.0 | Learner | Low | |

Fig. 9. OA con nivel de dificultad Medium,

Finalmente, para la PC 5 con el tipo de recurso educativo *Lecture*, tuvo como resultado lo que se puede apreciar en la Figura 10.

| SPARQL query: | |
|---|------------|
| <pre> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX owl: <http://www.w3.org/2002/07/owl#> PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX SV: <http://www.semanticweb.org/valeria/ontologies/2017/10/OntoVC#> SELECT DISTINCT ?description ?typeInt WHERE { ?EMLearnResType SV:isLRTypeOf ?description. ?IntType SV:hasIntType ?typeInt. FILTER regex(str(?EMLearnResType), 'Lecture')} </pre> | |
| description | typeInt |
| InteroperabilidadSemantica | Expositive |

Fig. 10. Respuesta a la PC 5, tomando como valor el tipo de recurso educativo Lecture

4. Conclusiones y Trabajos Futuros

Este trabajo presenta una primera versión de la ontología *OntoVC* utilizada para definir el vocabulario compartido. *OntoVC*, forma parte de la propuesta de enfoque híbrido realizada en este trabajo, para dar soporte a la interoperabilidad semántica entre repositorios de OA. Se presentó el vocabulario y se resaltó las relaciones que el mismo tiene con los distintos estándares de metadatos más ampliamente usados. También se definieron preguntas de competencia que ayudan a detectar los conceptos que son necesario incluir en dicho vocabulario. También a modo de validación, se pobló la ontología con algunos OA con el objetivo de ejecutar consultas a la misma que den respuesta a las principales preguntas de competencia. Esta ontología permite realizar búsqueda con mayor nivel de especificidad para los OA que estén alojados en los repositorios que implementen este nuevo enfoque. De esta manera es posible identificar problemas como por ejemplo la falta de algún concepto no tenido en cuenta. A partir de este trabajo, es posible continuar sobre una base sólida las siguientes etapas.

Como próximo paso de la propuesta realizada es implementar la ontología *OntoVC* en repositorios de prueba que sirvan de una primera aproximación a obtener los resultados esperados. Para esto resulta necesario avanzar en la tercera etapa mencionada más arriba, donde se debe formalizar el mapeo semiautomático y establecer reglas de transformación entre las ontologías locales (Ontologías de DC, LOM, MOD, MARC) y la ontología global o vocabulario compartido (*OntoVC*).

Referencias

- P. Suber (2012) “Open Access”, The MIT press Essential Knowledge Series, Cambridge, Massachusetts, ISBN 978-0-262-51763-8
- J. Alonso-Arévalo and I. Subirats-Coll and M. Martínez-Conde (2008) “Informe APEI sobre acceso abierto”, Edición APEI, España, ISBN: 978-84-691-7725-9
- C.A. Lynch (2003) “Institutional Repositories: Essential Infrastructure for Scholarship in The Digital Age Portal: for Scholarship in the Digital Age”, <https://www.cni.org/wp-content/uploads/2003/02/arl-br-226-Lynch-IRs-2003.pdf>

- D. Wiley (2011) "Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy", In D. A. Wiley (Ed.): *The Instructional Use of Learning Objects*: Online Version: <http://reusability.org/read/chapters/wiley.doc>
- C. Barritt and F. L. Alderman Jr. (2004) "Creating a Reusable Learning Objects Strategy: Leveraging Information and Learning in a Knowledge Economy", Ed. John Wiley & Sons, San Francisco, California.
- P. R. Polsani (2003) "Use and Abuse of Reusable Learning Objects" In *Journal of Digital Information*, 3(4), Online version: journals.tdl.org/jodi/article/viewArticle/89
- COAR: Confederation of Open Access Repositories (2012), "The Current State of Open Access Repository Interoperability", Working Group 2: Repository Interoperability. Octubre.
- Rodrigues, E. (2012) "Challenges and opportunities of interoperability for open access repositories", In *China Open Access Week*,
- H. Garrido Arenas and M. Lisowska Navarro (2014) "¿Estamos preparados para trabajar en red? Interoperabilidad: desafíos para la región latinoamericana" In *XX ISTECS (Ibero-American Science & Technology Education Consortium)*, General Assembly. Puebla, México.
- Institute of Electrical and Electronics Engineers (2009) IEEE Standard "Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries". New York.
- L. García Aretio (2005) "Objetos de aprendizaje. Características y repositorios" In *Boletín Electrónico de noticias de Educación a Distancia (BENED)*.
- H. Stuckenschmidt and F. V. Harmelen (2005) "Information Sharing on the semantic web", Springer – Verlag Berlín Heidelberg, ISBN: 978-3-540-26907-6
- H. Wang and Z. Ye (2009), "Building Multi-Level Data Warehouse Based on Hybrid-Ontology," *Computer Network and Multimedia Technology, CNMT 2009. International Symposium on, Wuhan*, pp. 1-4.
- L. Zhang, Y. Ma and G. Wang (2009), "An Extended Hybrid Ontology Approach to Data Integration," In *2nd International Conference on Biomedical Engineering and Informatics, Tianjin*, pp. 1-4.
- L. Zheng, J. Terpenny (2013), "A hybrid ontology approach for integration of obsolescence information", *Computer & Industrial Engineering*, ISSN: 0360-8352, Vol: 65, Issue:3, Page: 485-499
- J. Vian and R. L. R. Campos and C. E. G. Palomino and R. A. Silveira (2011) "A Multiagent Model for Searching Learning Objects in Heterogeneous Set of Repositories," In *Advanced Learning Technologies (ICALT), 11th IEEE International Conference on, Athens, GA*, pp. 48-52.
- V. C. Sandobal Verón and M. Ale and M. M. Gutiérrez (2015) "DCOntoRep: hacia la interoperabilidad semántica de Repositorios Institucionales de Acceso Abierto", In *1st Argentine Symposium on Ontologies and their Applications (SAOA)*, Rosario, Argentina, September 2-3, pp.91-100, vol.1449
- V. C. Sandobal Verón and M. Ale and M. M. Gutiérrez (2016) "An interoperability

- model based on ontologies for learning object repositories" In International Symposium on Computers in Education (SIIE), Salamanca, pp. 1-6.
- N. Guarino and M. Carrara and P. Giaretta (1994) "An ontology of meta-level categories", In Principles of Knowledge Representation and Reasoning (KR), San Francisco, CA, USA, pp. 270-280
- Directrices Driver 2.0. Directrices de proveedores de contenido, Noviembre, 2008
- OpenAire Guidelines for Literature Repositories v3.0. Available in: <https://guidelines.openaire.eu/en/latest/literature/index.html>, Abril, 2013
- Ley 26899: Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos. Noviembre, 2013, Available in: <http://repositorios.mincyt.gob.ar/recursos.php>.
- Directrices SNRD. Directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales. Ministerio de Ciencia, Tecnología e Innovación Productiva, Mayo 2013.
- La Referencia, Red Federada de Repositorios Institucionales de Publicaciones Científicas. Metadatos y Políticas de cosecha, Mayo 2015.
- OpenAire. Open Access Infraestructure for research in Europe. Framework Programme 7(2007-2013)
- H. Wache and T. Vögele and U. Visser, H. Stuckenschidt and G. Schuster and H. Neuman and S. Hübner (2001), "Ontology-Based Integration of Information – A survey of existing approaches", In Workshop on Ontologies and Information Sharing at the 17th International Joint Conference on Artificial Intelligence, Seattle, US, August 4-5, pp. 108-117