

Una aproximación para la descripción semántica de requisitos para categorización docentes de investigadores Ecuatorianos

Elizabeth Cadme, Nelson Piedra

UTPL, Universidad Técnica Particular de Loja, San Cayetano Algo (S/N), Loja, Ecuador, 11 01 608.

Autores para correspondencia: iecadme@utpl.edu.ec, nopiedra@utpl.edu.ec

Fecha de recepción: 21 de septiembre de 2014 - Fecha de aceptación: 17 de octubre de 2014

RESUMEN

La nueva generación de la Web, la Web Semántica, plantea potenciales oportunidades para dotar de significado a los contenidos Web. Las ontologías constituyen una de las principales herramientas para especificar explícitamente los conceptos de un dominio concreto, sus propiedades y sus relaciones; de manera que la información se publique en formatos que sean inteligibles automáticamente por agentes máquinas que pueden localizar y gestionar de forma precisa la información. En este trabajo se presenta un marco de trabajo para una red de ontologías para representar conceptos, atributos, operaciones y restricciones, en relación a los ítems curriculares que se usan en procesos nacionales de categorización de docentes universitarios ecuatorianos. En una primera parte se muestra el contexto del dominio, trabajos relacionados, luego se describe el proceso seguido, la abstracción del modelo ontológico y finalmente se presenta una ontología. Es una ontología de dominio debido a que proporciona el significado de los conceptos y sus relaciones dentro del dominio de ítems curriculares producidos por docentes universitarios, que son requisitos de los procesos de categorización docente universitaria en Ecuador.

Palabras clave: Web semántica, datos enlazados, ontología, metodología, datos universitarios.

ABSTRACT

The new generation of the Web, the Semantic Web, poses potential opportunities to give meaning to web content. In this paper a framework is presented for ontology network to represent concepts, attributes, operations and restrictions related to curricular items used in national categorization processes of Ecuadorian university teachers. The first part of the paper describes the domain context and related work, the following the abstraction of the model is described, and finally ontology is presented. The domain ontology provides the concepts meaning and their relationships within the domain of curricular items produced by academics, which are requirements of university teaching categorization process in Ecuador.

Keywords: Semantic web, linked data, ontology, methodology, universities data.

1. INTRODUCCIÓN

Berners-Lee *et al.* (2001), pone de manifiesto la Web Semántica para que la Web tradicional pueda ser entendida por la máquina y el hombre a través de técnicas de datos enlazados (*linked data*), y permite que tanto las máquinas como las personas, interpreten el significado de los datos y su relación con otros. Se asocian los principales planteamientos orientados a la apertura de los datos para que puedan ser tratados desde su significado.

Para la publicación de datos enlazado, Berners-Lee *et al.* (2006) propone cuatro reglas: 1) Utilización de URIs para identificar las cosas, 2) Usar URIs HTTP para que las personas puedan

buscar recursos, 3) Utilizar estándares RDF para describir recursos y SPARQL para realizar consultas, 4) Incluir enlaces a otras URIs para descubrir más cosas.

Además, se plantean cinco principios de apertura (denotados por estrellas), que indican el nivel de apertura que adquieren los datos que están en Web, desde ponerlos disponibles en Web utilizando licencias abiertas, pasando por estructurarlos y publicarlos en formatos específicos hasta enlazarlos con otros datos.

En el “informe de tendencias e iniciativas actuales de datos abiertos”¹ de la unión europea, publicado en septiembre de 2014, se han recogido las últimas tendencias de apertura de datos, y en su sección de publicación de datos los conjuntos prioritarios se refieren a Estadísticas nacionales, Mapas nacionales, Datos electorales a nivel nacional y Presupuestos nacionales; y, entre las áreas temáticas con mayor potencial se encuentran las publicaciones científicas. Varias organizaciones han trabajado sus silos de información con técnicas de datos enlazados y han pasado a formar parte de Linked Data Cloud² la (nube de datos abiertos y enlazados) en donde se reflejan datasets en la Web de datos, compartidos y enriquecidos con otros. Entre los ejemplos más visibles de publicación y consumo de datos enlazados están la Biblioteca Nacional Española (BNE)³, BBC World Cup Website⁴, fue reconocido entre los mejores ejemplos de Web Semántica en el año 2010 y utilizó web semántica dinámica. También se han unido varias iniciativas de Open Data de algunos gobiernos de países tanto de Europa como de América, que fomentan la apertura de datos hacia todos sus ciudadanos brindándoles varios tipos de información y utilizando técnicas de datos enlazados, abanderando esta iniciativa los gobiernos de Reino Unido con Data.gov.uk⁵ y Estados Unidos con Data.gov⁶. En Ecuador los primeros set de datos disponibles en la nube de datos abiertos y enlazados son: datos enlazados de la Universidad Técnica Particular de Loja (UTPL) ver: <http://datahub.io/dataset/utpl-lod>; datos enlazados de la Universidad de Cuenca, ver: <http://datahub.io/dataset/universidad-de-cuenca-linkeddata>; Datos de producción científica ecuatoriana publicados como Linked Data, ver: <http://datahub.io/dataset/opendataec>; Datos de recursos educativos abiertos, publicados por la iniciativa Serendipity⁷, ver <http://datahub.io/dataset/serendipity> y Datos Geográficos de Ecuador, ver <http://datahub.io/dataset/geoecuador>.

La iniciativa de Datos Abiertos y Enlazados ha incentivado a organizaciones e individuos para que publiquen y abran sus datos en formatos abiertos, así como al uso de aquellos datos que están disponibles en la Web. En el medio académico y de investigación, las Universidades y otras organizaciones educacionales manejan dentro de su estructura varias áreas, siendo una de las más importantes su producción intelectual conformada en gran parte por las publicaciones de cuerpo docente, lo cual genera una gran cantidad de información almacenada en diferentes plataformas y formatos. Por otra parte, en Ecuador se están ejecutando reformas a varios niveles (legislativo, organizacional, económico, etc.) para mejorar la calidad de la Educación Superior del país.

En este trabajo se expone el caso específico de descripción y modelamiento de una red de ontologías para el dominio de producción científica-académica universitaria como parte del proceso de inclusión dentro de la Web de Datos. A través del uso de estas tecnologías se podrá interoperar y enriquecer los datos universitarios ya sea para visualización, análisis y toma de decisiones a nivel de autoridades, instituciones evaluadoras de gobierno o para uso de investigadores y cualquier otro tipo de usuario, pues los campos que pueden interesarse por datos enlazados universitarios son diversos. A continuación, se describe el contexto del dominio, trabajos relacionados, luego se detalla el proceso seguido, la abstracción del modelo ontológico y finalmente se presenta la ontología resultante, que tiene como propósito representar semánticamente conceptos, atributos, operaciones y restricciones, en

¹ Informe de tendencias e iniciativas actuales de datos abiertos: http://datos.gob.es/sites/default/files/bestpractices_opendata_sep2014_1_1.pdf

² Linked Open Data Cloud Diagram: <http://data.dws.informatik.uni-mannheim.de/lodcloud/2014/>

³ Biblioteca Nacional Española: <http://www.bne.es/es/Inicio/Perfiles/Bibliotecarios/DatosEnlazados/index.html>

⁴ BBC WorldCup Website: http://news.bbc.co.uk/sport2/hi/football/world_cup_2010/default.stm

⁵ DATA.GOV.UK Opening up government: <http://data.gov.uk/>

⁶ Data.gov (United States): <https://www.data.gov/>

⁷ Serendipity: <http://serendipity.utpl.edu.ec>

relación a los ítems curriculares que se usan en procesos nacionales de categorización de docentes universitarios ecuatorianos.

2. IMPORTANCIA DE LA DESCRIPCIÓN DE FUENTES DE DATOS

2.1. Ciclos de vida para publicación de linked data

Considerando que el proceso datos enlazados o *linked data* es un conjunto de principios y buenas prácticas que permiten publicar, interconectar y enriquecer información en la Web de datos, se han publicado datos de diversos dominios como los citados anteriormente. Entre las principales propuestas LOD2, de Auer (2011), contiene ocho etapas que cumplen procesos de almacenamiento, autoría, interconexión, clasificación, calidad, evolución-reparación, y búsqueda-navegación-exploración; etapas que no se deben considerar de manera separada pues al constituir un ciclo, se alimentan de una a otra y se deja la posibilidad de mejorarlas utilizando nuevas estrategias, algoritmos, vocabularios que mejoren la publicación, mapeo, unión, etc., y recursos que enriquezcan estas fases planteadas con la finalidad de lograr mejorar la base de conocimiento que se irá generando en beneficio la Web de Datos. Otro modelo es propuesto en el proyecto DERI por Hausenblas (2011), consta de seis fases que involucran “conciencia de datos”, modelamiento, publicación, descubrimiento, integración y casos de uso; en este ciclo de vida se resalta la diversidad y especial tratamiento de datos en la primera fase, las siguientes fases sintetizan lo descrito para LOD2, son semejantes a las que se describen en Villazón-Terrazas *et al.* (2011) ciclo compuesto de cinco fases desde la especificación de datos, modelamiento, generación, publicación y explotación. Así mismo en el dominio datos relacionados con educación tenemos el ciclo descrito en Piedra *et al.* (2014) que propone un marco de trabajo para la publicación de datos enlazados, que inicia con una fase de visionamiento del modelo de negocio, identificación y selección e las fuentes de datos, modelamiento, generación de datos RDF, publicación, consumo y visualización.

Piedra *et al.* (2014) hace referencia a la existencia de dominios conformados por silos de datos estructurados por cada comunidad, que pueden ser homogéneos y heterogéneos. Y se considera la necesidad de identificar en primera instancia cuál es el conjunto de datos con el cual se va a trabajar y si los datos son estructurados o no, delimitando claramente el alcance. Por otra parte, Villazón-Terrazas *et al.* (2011), destaca la necesidad de determinar con qué datos contamos y qué datos son los que se van a publicar, considerando aquellos que pueden ser públicos, y los que tienen carácter reservado y no participarían en el proceso de datos enlazados, así mismo resalta la importancia de revisar el tipo de licencia que poseen los datasets.

Se han seleccionado los ciclos de vida según se han ido adaptando a los recursos que hemos obtenido para este dominio de trabajo (producción científico-académica universitaria). En esta primera fase se ha llegado hasta el modelamiento, ha sido crucial llevar un proceso exhaustivo del análisis de fuentes de datos como se describe más adelante. El punto de partida es el análisis de fuentes de datos que influirán en la validación de resultados y tiene incidencia sobre los recursos que se utilizan durante todo el ciclo.

El dominio de datos de estudio cubre el perfil de investigadores ecuatorianos, enfocándonos específicamente en los parámetros considerados para la categorización docente según lo determina la legislación del país. Se cuenta con datasets diversos a pesar de corresponder al mismo dominio universitario. A continuación se detallan las tareas llevadas a cabo para esta fase de análisis y descripción de las fuentes de datos dentro del proceso de publicación de datos enlazados.

2.2. Identificación de conceptos y proveedores de datos

Se ha considerado necesario partir del análisis la normativa institucional y estatal, que ha permitido extraer conceptos considerados como indicadores importantes de la actividad docente dentro del campo de la educación superior. La normativa que rige la educación superior se basa en algunos

documentos: Ley Orgánica de Educación Superior (LOES)⁸ del Ecuador, Reglamento para escalafón docente de la LOES, Reglamento de carrera docente de la Universidad Técnica Particular de Loja (UTPL)⁹, Normativa sobre formación docente UTPL y Normativa sobre producción intelectual UTPL. Para el análisis de los datos a publicar se revisaron los apartados de la LOES referentes a indicadores que se consideran para evaluación y categorización de Universidades, así como normativas referentes a evaluación docente en documentos de UTPL.

En la Fig. 1 se observa el proceso realizado para la descripción: Los documentos revisados, se resumen en los aspectos que se reconocen dentro del curriculum vitae que debe reunir un docente de una institución de educación superior, mismos que se resumen en: Grado académico, Experiencia académica, Publicaciones, Evaluación de desempeño, Capacitación y actualización profesional, Proyectos, Tesis, Suficiencia de lengua extranjera; y, Resultados de concursos de méritos y oposición.

Los conceptos extraídos de los documentos de la normativa universitaria, nos permiten dirigirnos hacia las dependencias que contienen las fuentes de datos relacionadas con los conceptos encontrados.

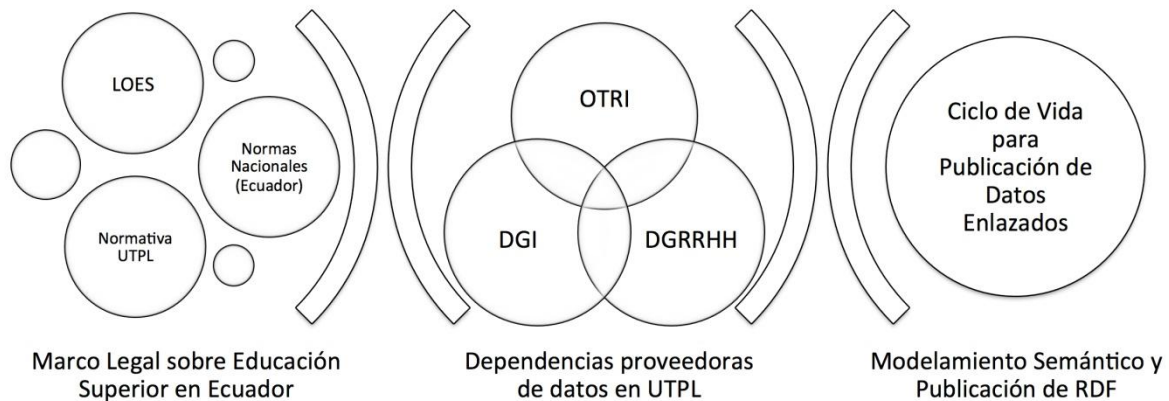


Figura 1. Recursos de información recogidos en UTPL para efectos de evaluación de un docente de educación superior según LOES.

Se identificaron tres dependencias institucionales tienen a su cargo información relacionada con los conceptos obtenidos inicialmente y éstas son: Oficina de Transferencia de Resultados de Investigación (OTRI), Dirección General de Investigación (DGI) y Dirección General de Recursos Humanos (DGRRHH). Cada una de las dependencias localizadas cuenta con un sistema de gestión de información y además con una persona responsable de la administración del mismo, a quien se ha considerado como el experto que nos ha brindado apoyo para entender la estructura de las fuentes de datos y su significado, además de proporcionarnos los datos necesarios y la respectiva documentación.

2.3. Análisis de las fuentes de datos

Las dependencias seleccionadas nos han proporcionado datasets y documentación, la siguiente tarea desarrollada para relacionar los conceptos seleccionados con los datos con los que contamos se realizó tomando en cuenta de cada base de datos: Tipo de recursos de información relacionados con los conceptos obtenidos en la tarea anterior, Entidad responsable, Nivel de granularidad de la información, Tipo de fuente de información, Frecuencia de actualización, Responsable de la información. En la Tabla 1 se puede observar el análisis para la base de datos correspondiente a Dirección General de Investigación, en la misma podemos identificar que se relaciona con los conceptos de *publicaciones científicas* y *formación docente*, tenemos información detallada por cada docente para los conceptos

⁸ Ley Orgánica de Educación Superior del Ecuador: <http://www.ces.gob.ec/descargas/ley-organica-de-educacion-superior>

⁹ Documentos institucionales de UTPL. Procuraduría universitaria: <http://procuraduria.utpl.edu.ec/institucionales>

identificados (producción en publicaciones y formación integral), la base de datos en MySql nos permitirá extraer los datos sin problema y los datos son actualizados constantemente por los docentes.

Tabla 1. Datos de identificación de la base de datos y relación con conceptos publicaciones y formación docente.

Tipo de recursos de información relacionada con los conceptos identificados:	Publicaciones científicas Formación Docente
Entidad responsable:	Dirección General de Investigación
Nivel de granularidad:	Información detallada por el docente – investigador
Responsable:	Ing. Diana Torres
Fuente de información:	Base de datos MySql alimentada por su sistema de información
Frecuencia de actualización:	La base de datos es actualizada según lo realicen los docentes, está habilitada durante todo el año
Número de tablas:	84
Número de filas:	66503

De igual manera se procedió con las otras dos dependencias seleccionadas (DGRRHH y OTRI). En la Fig. 2, en la cual se resume gráficamente el trabajo, la frecuencia de actualización es continua, se observan los conceptos relacionados con cada una de las bases de datos y también la cantidad de filas y tablas que comprenden la información con la que contamos.



Figura 2. Contexto de origen de fuentes de datos.

Como se observa en la Fig. 2, los conceptos obtenidos se han reducido a seis, esto se debe a que en la tarea de análisis los administradores de cada base de datos indicaron aspectos relacionados con las políticas de confidencialidad de datos, con lo cual se ha delimitado nuevamente el alcance de nuestro dominio, quedando finalmente los conceptos: a) datos personales del docente, b) producción intelectual, c) proyectos, d) actualización profesional, e) grado académico; y, f) experiencia laboral.

3. MODELAMIENTO SEMANTICO

Para este trabajo se han seguido las guías metodológicas dadas en Gómez-Pérez & Suárez-Figueroa (2009) y Suárez-Figueroa (2010) basado en escenarios para el modelamiento y desarrollo de ontologías. De los ocho escenarios propuestos, se ha trabajado con el escenario 1 para determinar la especificación de requisitos de la ontología y el escenario 3 para el trabajo con reutilización de recursos ontológicos.

Un correcto análisis y descripción de las fuentes de datos, facilitará el escenario de la construcción de un modelo ontológico que resulta en el vocabulario a utilizar para nuestros datos enlazados. Las guías metodológicas proporcionadas por diferentes autores se complementan con la definición de un modelo de datos que se refleja en una ontología o vocabulario que permite continuar con el proceso y facilita la conversión de datos en tripletas RDF.

3.1. Mapeo entre conceptos y entidades de almacenamiento físico

Las actividades anteriores han permitido identificar las tablas de las bases de datos analizadas que contienen información sobre los conceptos seleccionados, también se identifican los atributos que corresponden a cada concepto. De las tablas se extrajeron los atributos relacionados a los conceptos, y se realizó un mapeo entre atributos, conceptos y vocabularios ya existentes, con lo cual se obtuvo un documento de términos a utilizar que más adelante se utilizará al momento de construir la ontología en la fase de reutilización de recursos ontológicos. En la Tabla 2 se presentan la equivalencia de datos de docentes con propiedades semánticas basadas en el vocabulario FOAF; por ejemplo, el atributo “nombre de docente” puede ser representado mediante una propiedad de objeto “foaf:firstName”.

Tabla 2. Mapeo entre atributos de que describen a docentes en base de datos y propiedades en RDF.

Nombre tabla.campo en BD	Tipo	Ejemplo	Propiedad RDF	
docentes	Docente		foaf:Person	Class
docentes.primer_nombre	String	"first_name": "ANDRÉS",	foaf:firstName	Property
		"second_name":		
docentes.segundo_nombre	String	"FERNANDO",	foaf:giveName	Property
		"first_surname":		
docentes.primer_apellido	String	"SÁCHEZ",	foaf:lastName	Property
		"second_surname":		
docentes.segundo_apellido	String	"POMA",	foaf:giveName	Property
docente.fecha_de_nacimiento	Date	"birthday": "01/01/1975",	foaf:birthday	Property
docente.lugar_de_nacimiento	String	"birthplace": "Cuenca",	address:localityName	Property
docentes.ciudad	String	"city": "Loja",	address:localityName	Property
docentes.provincia	String	"province": "Loja",	address:región	Property

Se analizaron las tablas, revisando el número de filas, claves primarias, foráneas, clave única. Adicionalmente, se realizó un análisis sobre el nivel de privacidad los atributos de las tablas seleccionadas. Este proceso facilitó la selección de conceptos y atributos para el modelamiento del dominio que comprende el vocabulario. Se cuenta también con información de algunos atributos de las tablas que no pueden ser utilizados por considerarse de carácter privado, y que por tanto no podrán ser publicados.

3.2. Especificación de requisitos para desarrollo de ontología

El propósito de construcción de la Ontología de Referencia es proveer un modelo de conocimiento consensado para el dominio de actividad científica y académica universitaria, que podrá ser utilizada por docentes y autoridades y público en general que desee conocer este tipo de actividades. El proceso de desarrollo del modelo ontológico inicia con la especificación de requisitos, basado en un conjunto de preguntas y respuestas dentro del dominio que ya está determinado, con ello se ha logrado obtener una frecuencia de términos utilizados, contrastados con los resultados obtenidos en la actividad de

mapeo. La ontología tiene que enfocarse en el dominio de información científica y académica que realizan un docente universitario, los principales conceptos que la ontología debe representar son: docente, publicación, proyecto, actualización académica, universidad, organización, grado académico.

Como requerimientos no funcionales de la ontología resultante señalamos dos: a) debe ser modular, que permita la reutilización, agregación o edición de clases y propiedades, de ser necesario para otros trabajos o usos posteriores; b) el idioma en que debe desarrollarse es el inglés, que se puede considerar como un idioma genérico en el que están desarrolladas la mayoría de ontologías.

La ontología tiene como usuarios finales:

- Docentes universitarios.
- Líderes de grupo/ Jefes de departamento/ autoridades que desean conocer el desempeño de su grupo de trabajo y tomar decisiones de acuerdo a criterios institucionales.
- Pares académicos internos y externos.
- Estudiantes, que se insertan en el ámbito universitario y necesitan involucrarse en distintas áreas de investigación.
- Organizaciones gubernamentales, las cuales realizan un monitoreo y evaluación de instituciones educativas.

Entre los potenciales usos que la ontología permitirá dar respuesta a las preguntas de competencia realizadas inicialmente, por ejemplo:

- Poner a disposición del público en general la información de producción académica y científica universitaria a través de un portal.
- Búsqueda de proyectos de investigación que se encuentran desarrollando en la institución.
- Búsqueda de relación de docentes de UTPL con otros docentes de otras universidades al rededor del mundo.
- Búsqueda de áreas de investigación en las cuales se desempeñan los docentes de la institución.
- Búsqueda de información sobre el nivel de titulación de la planta docente de UTPL para fines de evaluación.

El lenguaje seleccionado para la implementación de la ontología es OWL-DL. El proceso de modelamiento de datos, está basado en el documento guía publicado por la W3C “Mejores Prácticas para seleccionar Vocabularios”¹⁰ el cual expone una lista de criterios a tomar en cuenta para seleccionar el vocabulario adecuado una vez se cuenta con el listado de términos y conceptos del análisis de dominio. Para este trabajo se ha trabajado con los criterios: multilingüismo, frecuencia de utilización, frecuencia de actualización, editor (Publisher), documentación disponible, análisis previo del vocabulario y a reutilizar identificando el dominio al cual está representando, función que desempeña una clase o relación dentro de la ontología diseñada; y, restricciones definidas para cada clase o relación. La principal herramienta utilizada para buscar los vocabularios a analizar “Linked Open Vocabularies” LOV¹¹, que ha permitido realizar el análisis de los parámetros seleccionados.

El primer enfoque es el modelo conceptual con las ontologías a utilizar para la red propuesta, como se indica en la Fig. 3.

El análisis y descripción de las fuentes de datos ha permitido que la tarea de reutilización de vocabularios sea más sencilla, los vocabularios dublin core¹², bibo¹³, vivo¹⁴, foaf¹⁵, contienen varias propiedades y conceptos de nuestro modelo. Entre los vocabularios relacionados al área de investigación y academia, contamos con:

- *Bibo*: Representa principales conceptos y propiedades para describir las citas y referencias bibliográficas, que es apta para usarse en el dominio de producción intelectual (publicaciones científicas) de los docentes universitarios. Utilizada para representar datos relacionados con “Publicaciones Científicas”.

¹⁰ http://www.w3.org/2011/gld/wiki/222_Best_Practices_for_Vocab_Selection

¹¹ LOV, Linked Open Vocabularies <http://lov.okfn.org/dataset/lov/>

¹² Vocabulario Dublin Core <http://dublincore.org/documents/2012/06/14/dcmi-terms/>

¹³ Bibliographic Ontology Specification <http://bibliontology.com>

¹⁴ VIVO-ISF Ontology version 1.6 <http://www.essepuntato.it/lode/owlapi/http://vivoweb.org/ontology/core#>

¹⁵ FOAF Vocabulary Specification 0.99 <http://xmlns.com/foaf/spec/>

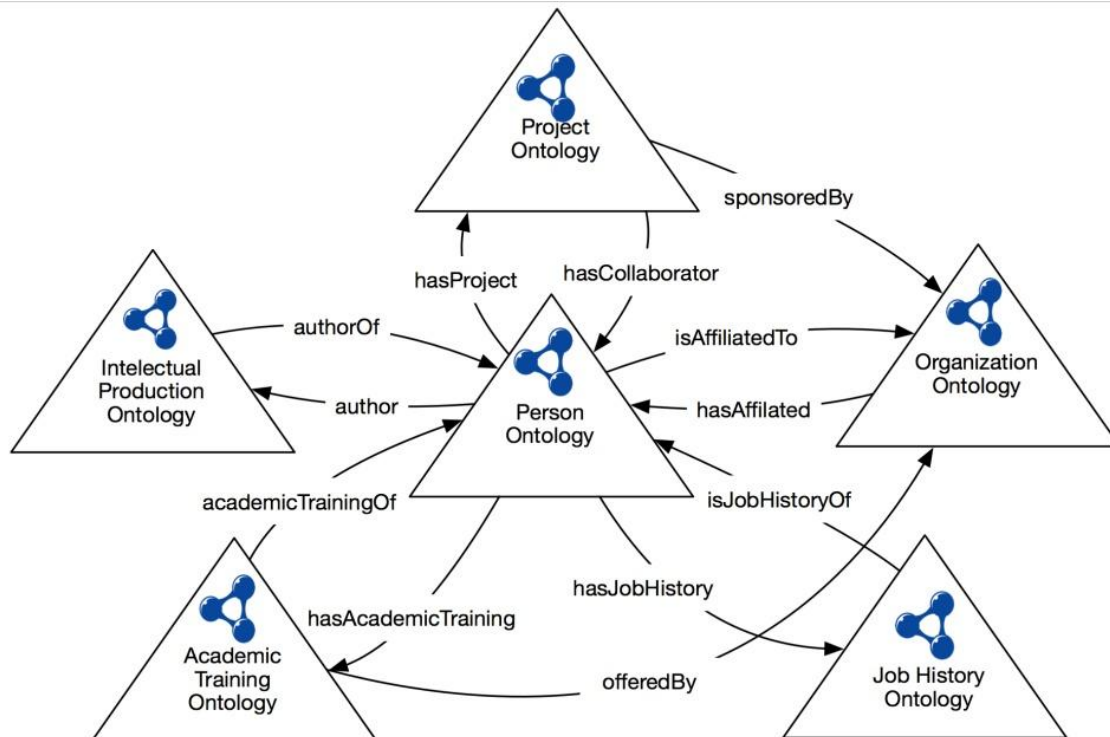


Figura 3. Modelo conceptual de la red de ontologías para categorización docente.

- *Vivo*: Para representar recursos de un entorno académico y de investigación universitario, algunos de ellos considerados dentro del dominio de este trabajo como “Grado Académico”.
- *DublinCore*: Dublin Core Metadata Initiative (DCMI) Metadata Terms desarrollado para representar recursos y atributos que se utiliza mucho en el ámbito bibliográfico será reutilizado para representar publicaciones, sus atributos como Documentos, título, creador, y propiedades con otras entidades, con las cuales es necesario relacionar dichos elementos.
- *Foaf - Vcard*: Vocabularios desarrollados para representar a Personas con sus atributos y relaciones hacia otras entidades, estos dos vocabularios nos permiten representar a los Docentes/Profesores y sus atributos tales como nombres, apellidos, dirección, teléfono, etc., que consisten en su descripción personal; así mismo permitirán establecer relaciones con otras clases a como puede ser la relación de afiliación.
- *Frapo*: Describe información administrativa de proyectos de investigación. Ayuda a representar datos relacionados con Proyectos realizados por los docentes.

A partir del modelo conceptual se diseñaron las ontologías planteadas para cada subdominio, mismas que se podrán trabajar por separado o integrarse para lograr los diversos fines de aplicación que se pueda seguir proporcionando.

La Fig. 4 representa el modelo correspondiente a actualización profesional. Se ha reutilizado mayormente los vocabularios dublin core, vivo y schema, además de que ha sido necesario modelar clases propias para este subdominio. Contiene las clases de actualización académica (:EducationalTraining), organización (org:Organization), docente (foaf:Person), grado académico (vivo:AcademicDegree), y las relaciones existentes reutilizando propiedades de datos (por ejemplo: vivo:relates) y de objeto (dcterms:date). En la Tabla 3 se describen de los términos correspondientes a las clases modeladas (obtencion de titulos universitaio y participacion del docente en eventos académicos), y su representación dentro del subdominio de actualización profesional.

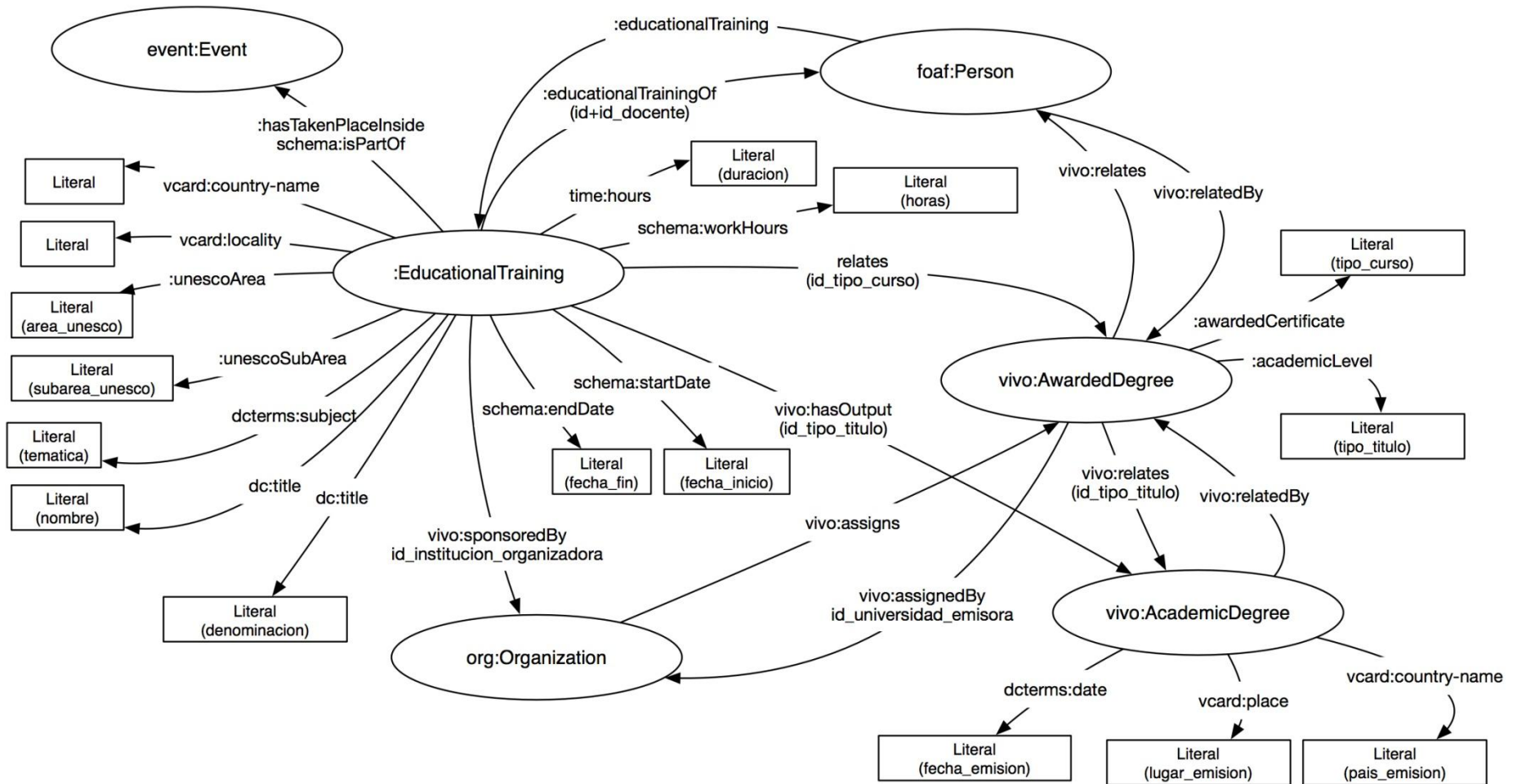


Figura 4. Modelo de datos para actualización profesional.

Tabla 3. Términos utilizados para modelar el subdominio de actualización profesional.

Clase	Descripción
Event	Representa el tipo de evento en el cual un docente ha participado
Educational training	Representa la preparación académica que constantemente realiza el docente en su área de especialización
Organization	Se ha utilizado para representar la organización de la cual el docente ha obtenido el aval de su actualización académica
Person	Representa al docente-investigador
Academic degree	Utilizada para modelar el grado académico que obtiene el docente

La Fig. 5 corresponde al modelo de datos para historial laboral de un docente. En este caso este subdominio se ha basado en tres conceptos: historia laboral (:WorkHistory), docente (foaf:Person) y organización (org:Organization) y las relaciones existentes como por ejemplo dcterms:description que representa la descripción de un trabajo realizado. También ha sido necesaria la creación de relaciones que permitan modelar la relación existente entre la historia laboral del docente y la organización a la que corresponde.

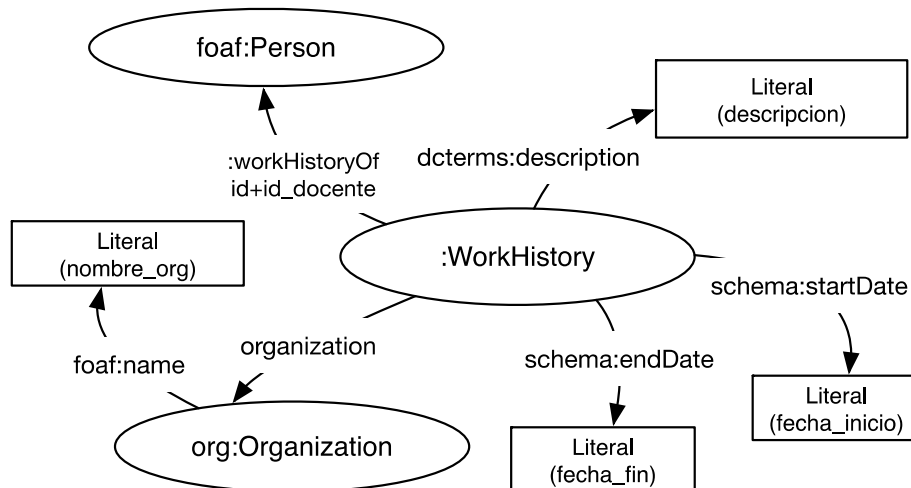


Figura 5. Modelo de datos para historia laboral.

La Tabla 4 describe las clases trabajadas para modelar el subdominio de historia laboral. Las propiedades se han modelado tomando en cuenta los atributos que corresponden a este dominio. Algunas de las propiedades corresponden a un vocabulario propio al no encontrarse modeladas en ninguno de los vocabularios consultados.

Tabla 4. Términos utilizados para modelar el subdominio de historia laboral.

Clase	Descripción
Work history	Utilizado para representar el historial de trabajo que tiene un docente
Organization	Se ha utilizado para representar la organización de la cual el docente ha obtenido el aval de su actualización académica
Person	Representa al docente-investigador

Para el subdominio proyectos, se refleja en el modelo que se muestra en la Fig. 6. Se han relacionado las clases que representan a proyecto (foaf:Project), docente (foaf:Person), rol (org:Role), principalmente. Y las propiedades que relacionan estas clases (org:member, org:role, etc.); así como

también relaciones a datos que describen estas clases (dcterms:title, frapo:startDate). En este caso se ha reutilizado vocabularios foaf, vivo, org, dcterms, principalmente.

Tabla 5. Términos utilizados para modelar el subdominio de proyectos.

Clase	Descripción
Project	Representa proyectos de investigación/extensión/desarrollo que ha realizado un docente investigador de una institución de educación superior
Person	Representa al docente-investigador
AreaUnesco	Modelada para representar el área de la Unesco dentro de la cual están ubicados los proyectos

La Tabla 5 una descripción de los términos que corresponden a las clases modeladas en el subdominio para proyectos, al igual que en los otros subdominios, clases y atributos se han modelado reutilizando los vocabularios que representan dominios similares.

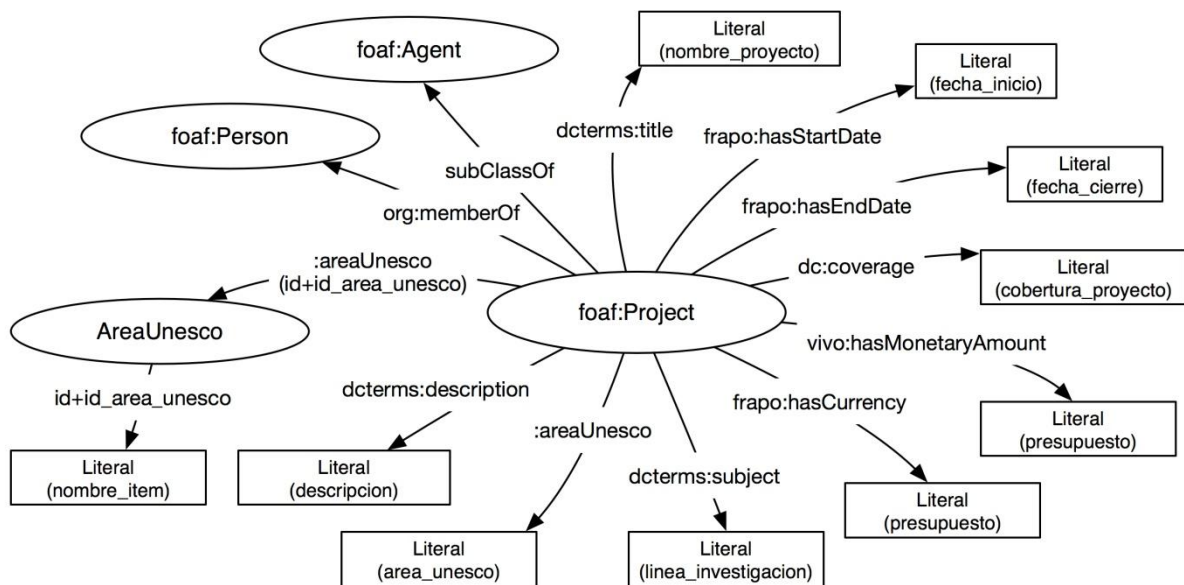


Figura 6. Modelo de datos para proyectos.

La Fig. 7, muestra el modelo que representa las publicaciones científicas, para modelar este subdominio se reutilizaron clases de vocabularios como bibo, vivo, dublin core, foaf, mismas que se han descrito en las diferente clases y relaciones del diagrama. Algo muy interesante en este modelado es poder encontrar la relación de 1 a n que representan los autores y sus publicaciones. En la Tabla 6 se puede observar las clases y la descripción de este subdominio.

El modelamiento de este subdominio de publicaciones, muestra la producción intelectual de los docentes y se o ha representado reutilizando las clases Artículo (bibo:Article), Co-autor (vivo:Authorship), Docente (foaf:Person), Organización (foaf:Organization), entre otras. Los vocabularios bibo, dublin core, foaf, vivo, han proporcionado las propiedades que relacionan clases (vivo:relates, dc:creator, etc.), relacionan a un objeto (bibo:abstract, dc:title, etc.), necesarias para modelar publicaciones científicas. Una vez establecido el modelo, se ha utilizado la herramienta Protégé para implementar la red de ontologías y puedan ser explotadas de diversas formas. El producto final es la red de ontologías que modela el dominio de información académica y científica de docentes investigadores.

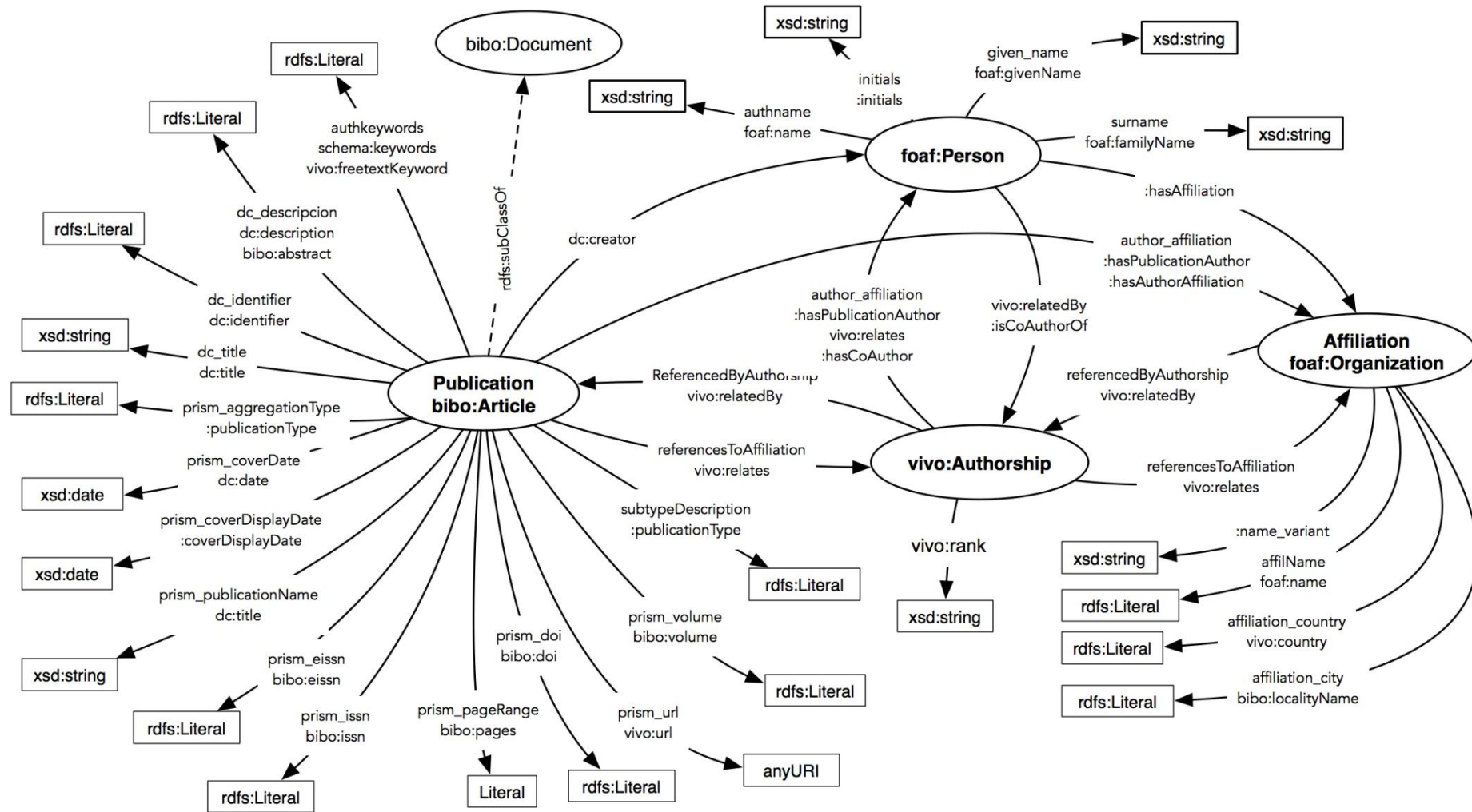


Figura 7. Modelo de datos para publicaciones científicas.

Tabla 6. Términos utilizados para modelar el subdominio de publicaciones.

Clase	Descripción
Publication	Representa los diferentes documentos que constituyen una publicación realizada en medios impresos o digitales, por un docente-investigador
Organization	Se ha utilizado para representar la organización al cual el docente está afiliado. (Afiliación)
Person	Representa al docente-investigador
Authorship	Utilizada para solucionar la cardinalidad entre docente y publicaciones. (Reificación)

4. CONCLUSIONES

Este modelo ontológico es fundamental para describir e integrar información de docentes universitarios y que es fundamental para crear aplicaciones, por ejemplo para detectar redes de colaboración, categorización docente, benchmarking universitario, sistemas de tomas de decisiones, entre otras aplicaciones. La red de ontologías desarrollada permite describir semánticamente conceptos almacenados en fuentes de datos heterogéneas, tanto estructuradas, como no estructuradas. Algunas de las lecciones aprendidas durante el desarrollo de este trabajo se centran en el análisis de fuentes de datos, muchas veces pasadas por alto. En este caso hemos logrado reducir nuestro dominio de datos al hacer un análisis más detallado, recurriendo no solamente a fuentes de datos, sino también a documentos relacionados y a personas clave dentro de la organización que son quienes tienen dominio de su área de trabajo.

AGRADECIMIENTOS

Este trabajo ha sido financiado por la Universidad Técnica Particular de Loja, a través de la convocatoria interna de proyectos de investigación. Agradecemos al apoyo del equipo de la Dirección de Investigación de UTPL, que nos ha proporcionado el apoyo requerido para llevar a cabo el proceso de análisis y descripción de datos.

REFERENCIAS

- Auer, S., 2011. LOD2. The Semantic Web. [Presentación de Slideshare] Disponible en <http://www.slideshare.net/lod2project/the-semantic-data-web-sren-auer-university-of-leipzig>.
- Berners-Lee, T., J. Hendler, O. Lassila, 2001. The semantic web. *Scientific American*, 284(5), 28-37.
- Gómez-Pérez, A., M.C. Suárez-Figueroa, 2009. *Scenarios for building ontology networks within the NeOn methodology*. In: Proceedings of the 5th International Conference on Knowledge Capture, ACM, pp. 183-184.
- Hausenblas, M., 2011. Linked data Life Cycles [Presentación de Slideshare]. Disponible en <http://www.slideshare.net/mediasemanticweb/linked-data-life-cycles>.
- LOES, 2010. *Ley Orgánica de Educación Superior*, Registro Oficial No. 298, Quito, Ecuador.
- Piedra, N., E.Tovar, R. Colomo-Palacios, J. Lopez-Vargas, J.A. Chicaiza, 2014. Consuming and producing linked open data: the case of Opencourseware Program. *Electronic Library and Information Systems*, 48(1), 16-40.

Suárez-Figueroa, M.C., 2010. *NeOn Methodology for building ontology networks: specification, scheduling and reuse*. Universidad Politécnica de Madrid, Disertación Doctorale.

Villazón-Terrazas, B., L.M. Vilches-Blázquez, O. Corcho, A. Gómez-Pérez, 2011. *Methodological guidelines for publishing government linked data. Linking Government Data*. Springer, New York, USA, pp. 27-49.