

# Marco de Trabajo para la Integración de Recursos Digitales Basado en un Enfoque de Web Semántica

Nelson Piedra <sup>1</sup>, Janneth Chicaiza <sup>1</sup>, Pricila Quichimbo <sup>1</sup>, Víctor Saquicela <sup>2</sup>, Elizabeth Cadme <sup>1</sup>, Jorge López <sup>1</sup>, Mauricio Espinoza <sup>2</sup> y Edmundo Tovar <sup>3</sup>

nopiedra@utpl.edu.ec, jachicaiza@utpl.edu.ec, pvquichimbo@utpl.edu.ec, victor.saquicela@ucuenca.edu.ec, iecadme@utpl.edu.ec, jalopez2@utpl.edu.ec, mauricio.espinoza@ucuenca.edu.ec, edmundo.tovar@upm.es

<sup>1</sup> Universidad Técnica Particular de Loja, San Cayetano Alto S/N, 1101608, Loja, Ecuador.

<sup>2</sup> Universidad de Cuenca, Av. 12 de Abril, 010150, Cuenca, Ecuador.

<sup>3</sup> Universidad Politécnica de Madrid, Avda. Montepríncipe S/N, 28660, Madrid, España.

DOI: 10.17013/risti.e3.55-70

**Resumen:** En un entorno abierto como la Web, no es posible estandarizar los procesos de descripción y publicación de metadatos, cada institución puede manejar diferentes formatos o modelos de datos. Para mejorar la interoperabilidad semántica entre repositorios heterogéneos, se están adoptando enfoques basados en tecnologías de la Web Semántica; de esta manera, cada Librería Digital puede conservar sus cualidades locales específicas y no requerirá resignarlas para poder normalizar el intercambio, reuso o la cosecha de recursos digitales. En este trabajo, se presenta un ciclo que cubre los procesos de: extracción de metadatos desde repositorios OAI-PMH, y la generación y publicación de datos enlazados, con el propósito de mejorar la integración e interoperabilidad de recursos almacenados en Librerías Digitales. La propuesta descrita facilita la existencia de diversidad de métodos y estándares en los procesos de cada proveedor de recursos digitales.

**Palabras-clave:** Repositorios Digitales, Datos Enlazados, Web Semántica, RDF, OAI-MPH, Ontología.

## *Framework for the integration of digital resources based-on a Semantic Web approach*

**Abstract:** From the point of view of access to metadata from distributed repositories, the Open Archives Initiative (OAI) proposed a protocol for the interchange and harvesting of metadata called OAI-PMH. This protocol provide a low degree of interoperability, however, with an approach based on semantic technologies, the interoperability of data can reach a higher level; thus, each digital library can retain their specific local qualities and will not require reassign them in order to standardize the exchange, re-use or harvesting of digital resources. In this paper, the process for the extraction of metadata, generation and publishing linked data in order to improve integration and interoperability between resources stored

on Digital Libraries is presented. The proposal facilitates the existence of diversity of methods and standards in the processes of each supplier of digital resources.

**Keywords:** Digital Library, Linked Data, Semantic Web, RDF, OAI-PMH, Ontology.

## 1. Introducción

Internet genera un escenario global en el que las condiciones de interoperabilidad (IEEE, 1990) son las que garantizan el descubrimiento, la distribución y re-uso de recursos digitales, más allá de condiciones de gestión local, de modelos tecnológicos y de herramientas de cualquier tipo. Para lograr la interoperabilidad entre repositorios de librerías digitales la Iniciativa de Archivos Abiertos (OAI) propone el protocolo para la recolección de metadatos denominado OAI-PMH (OAI, 2015). OAI-PMH se basa en estándares abiertos, por tanto, garantiza la interoperabilidad automática entre emisores y receptores de recursos digitales, con independencia del software utilizado e incentivando la neutralidad tecnológica y la innovación. Aunque OAI-PMH facilita el intercambio de metadatos en la Web, aún pueden persistir problemas para integrar los datos extraídos desde diversos repositorios. En un entorno abierto como la Web, no es posible estandarizar los procesos de descripción y publicación de metadatos, por tanto, cada institución puede manejar diferentes formatos de datos o esquemas de metadatos o vocabularios.

Con el objetivo de reducir las barreras para integrar la información de las librerías digitales, en un entorno heterogéneo, se está apostando por enfoques basados en tecnologías de la Web Semántica, específicamente Datos Enlazados (*Linked Data*). Con el concepto de la Web como repositorio global de datos enlazados se han conseguido significativos avances al momento de extraer y recuperar información útil para los usuarios, procesar el significado o semántica de la información, recuperar el conocimiento que forma parte de las páginas web, o incluso cuando se trata de interpretar sentencias de búsqueda en el contexto de la necesidad del usuario.

En este trabajo, se presenta un marco de trabajo para la publicación de los datos bibliográficos extraídos de repositorios digitales que usan OAI-PMH, siguiendo los principios de Linked Data y el ciclo de vida para la publicación de datos enlazados. En el siguiente apartado, se describen las tecnologías y las propuestas actuales para mejorar la interoperabilidad de metadatos. En la Sección 3, se describe el marco de trabajo propuesto para la cosecha de metadatos y posterior publicación de datos enlazados; un piloto aplicado a repositorios de universidades ecuatorianas es detallado en la sección 4. A continuación, en la Sección 5, se describen ciertos escenarios de uso de los datos generados. Finalmente, en la sección 6 se presentan las respectivas conclusiones y líneas de trabajo futuro.

## 2. Antecedentes y Trabajos Relacionados

En esta sección se presenta una breve introducción al protocolo OAI-PMH, Dublin Core y Datos Enlazados. Además se describen los enfoques existentes relacionados con este trabajo.

## 2.1. Protocolo OAI-PMH.

Desde el punto de vista tecnológico, las librerías digitales son repositorios que almacenan objetos digitales y utilizan OAI-PMH para exponer sus metadatos. OAI-PMH es un protocolo propuesto por la *Open Archives Initiative* que facilita la extracción de metadatos (descritos según un formato y esquema de metadatos) desde repositorios digitales. Para obtener los metadatos se utilizan los servidores de datos, que realizan solicitudes conocidas como verbos. OAI-PMH soporta 6 verbos que permiten obtener información relacionada a: el repositorio digital, el formato de metadatos, las colecciones de recursos y la descripción detallada de cada recurso.

OAI-PMH es compatible con muchas herramientas que permiten la creación de repositorios institucionales como Eprints<sup>1</sup>, Dspace<sup>2</sup>, Fedora<sup>3</sup>, entre otros. OAI-PMH divide este fenómeno en proveedores de datos y proveedores de servicios; los primeros son repositorios que exponen sus metadatos a través de OAI-PMH; los segundos, también llamados “harvesters o recolectores”, desarrollan servicios de valor agregado basados en los metadatos obtenidos de los proveedores. En OAI-MPH cada repositorio almacena sus objetos digitales de manera independiente.

## 2.2. Esquema de Metadatos Dublin Core (DC).

*Dublin Core* o la Iniciativa de Metadatos Dublin Core (DCMI<sup>4</sup>) es el esquema de meta-información más utilizado a nivel mundial<sup>5</sup>, para describir los metadatos de los recursos digitales. Para maximizar las posibilidades de interoperar con otras colecciones de datos, se ha utilizado a DC como base este esquema de metadatos. En los últimos años, el conjunto de elementos DC se ha convertido en una infraestructura operacional del desarrollo de la Web Semántica. Entre los metadatos DC para describir contenido web, están: *Title, Subject, Description, Source, Language, Creator, Publisher y Rights*. Además de los elementos básicos (ninguno obligatorio y todos repetibles) existen otros mecanismos que sirven para adaptar DC a las necesidades concretas de información y que hacen que este modelo de metadatos sea aplicable a cualquier proyecto de sistema o servicio de información digital.

## 2.3. Hacia la integración de repositorios a través de Datos Enlazados.

La Web Semántica visionada por Berners-Lee (2001) añade a la Web de Documentos del significado que le hace falta para disponer de un entorno en el que sea posible acceder a los datos contenidos en sitios Web y procesar automáticamente la información de un modo más exacto y completo. En este trabajo, los autores se centran en el concepto de Web Semántica, desde una perspectiva del procesamiento de grandes volúmenes de datos enlazados (*Linked Data*). Esta visión implica que los datos están almacenados en una base de datos global distribuida (Heath & Bizer, 2011).

<sup>1</sup> <http://www.eprints.org/>

<sup>2</sup> <http://www.dspace.org/>

<sup>3</sup> <http://fedorarepository.org/>

<sup>4</sup> <http://www.dublincore.org>

<sup>5</sup> <http://www.sedc.es/autoformacion/metadatos/tema7.htm>

La publicación de Datos Enlazados se fundamenta en cuatro principios básicos de diseño propuestos por Tim Berners-Lee (2006):

1. Usar URIs para identificar los recursos de la Web.
2. Usar URIs-HTTP para que los usuarios puedan localizar y consultar estos recursos.
3. Proporcionar información útil acerca del recurso cuando la URI haya sido consultada, utilizando RDF<sup>6</sup> para describir recursos y SPARQL<sup>7</sup> para consultarlos.
4. Incluir enlaces a otras URIs relacionadas con los datos contenidos en el recurso, de forma que se potencie el descubrimiento de información en la Web.

El enfoque de Linked Data ofrece ventajas significativas sobre las prácticas actuales de publicación de datos, pues mediante el uso de identificadores únicos (URIs) las bibliotecas permitirán que los recursos sean más fácilmente accesibles. Este enfoque tiene ventajas sobre un contexto basado solamente en OAI-PMH, en el que la Web es la infraestructura de transporte de datos y metadatos, mientras que desde Linked Data, los datos y su semántica son parte de la misma Web, están identificados a través de URIs y descritos en un lenguaje que permite la lectura y procesamiento automático por parte de agentes máquina (Ver Figura 1).

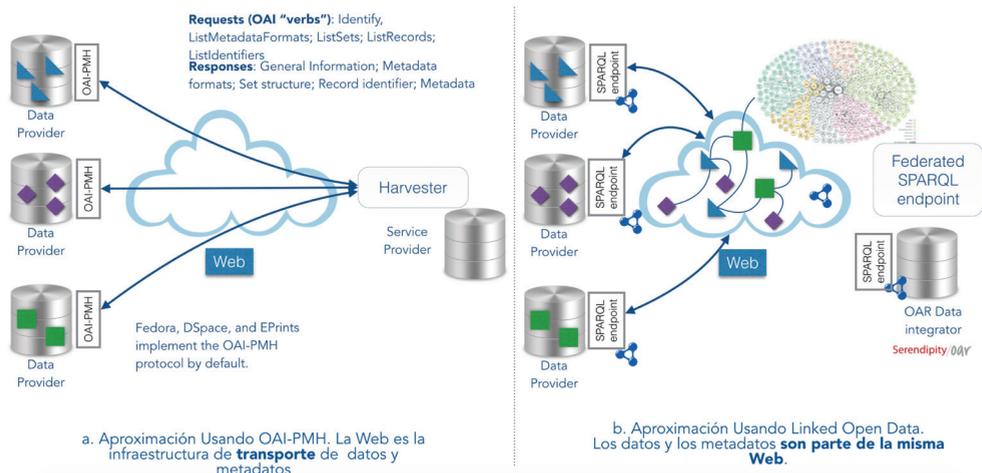


Figura 1 – Integración e interoperabilidad de metadatos bibliográficos de recursos digitales a través de Linked Data. Este enfoque tiene ventajas sobre un contexto basado solamente en OAI-PMH, en el que la Web es la infraestructura de transporte de datos y metadatos, mientras que desde Linked Data, los datos y su semántica son parte de la misma Web. Los recursos digitales pueden reusarse y adaptarse a nuevas necesidades educativas.

<sup>6</sup> Resource Framework Description, es un lenguaje para representar y publicar datos estructurados en la Web.

<sup>7</sup> SPARQL es el lenguaje para consultar grafos modelados con RDF.

Diferentes comunidades pueden beneficiarse de la adopción de un enfoque de integración e interoperabilidad semántica, así: (a) el movimiento de recursos educativos abiertos (Tovar, E. & Piedra, N. 2014); (b) usuarios en general que dispondrán de acceso a un conjunto más rico de información proveniente desde diferentes repositorios (búsquedas federadas); (c) los bibliotecarios y archivistas que pueden tener acceso a datos compartidos para describir sus recursos y reducir la redundancia; (e) Desarrolladores Web que enfrentan menores problemas de heterogeneidad en formatos y semántica de datos (Baker y otros, 2011); (f) investigadores que realizan extracción del conocimiento desde repositorios abiertos, por ejemplo para la búsqueda de expertos en un dominio determinado, detección de nuevos temas de investigación, análisis de redes científicas, (Rowlands & Bawden, 1999); entre otros.

#### **2.4. Trabajos Relacionados.**

Un escenario de integración hace posible la combinación de recursos de información existentes en diversas fuentes, esto proporciona al usuario una vista unificada de dichos recursos y también puede actuar como una fuente de datos para diversas aplicaciones (Lenzerini, 2002). Para conseguir este propósito, primero se deben resolver problemas de heterogeneidad a nivel de: (a) repositorios, (b) formatos de datos, (c) esquemas de metadatos, y (d) vocabularios o diccionarios de datos. Para abordar cada una de estas dimensiones, existen diferentes propuestas. El grado más bajo de interoperabilidad (a nivel de repositorios) se consigue con protocolos como OAI-PMH, el cual facilita la extracción e intercambio de metadatos, aunque su uso no asegura el reuso e integración de la información. Para asegurar la interoperabilidad semántica entre diferentes colecciones o esquemas (Gendt et al. 2006 y Francesconi et al. 2008) proponen crear links entre objetos equivalentes. En este sentido, el uso de vocabularios abiertos y la generación de datos abiertos en formatos legibles para máquinas, pueden ser claves para integrar elementos de diferentes repositorios, más allá de estándares y acuerdos previos entre los proveedores. En (Zengenene et al. 2014) se propone un marco teórico enfocado a proyectos de librerías digitales que intentan publicar sus metadatos mediante un enfoque de datos enlazados. Casos recientes de generación de datos enlazados son presentados en trabajos como (Malakhov et al., 2014, Anibaldi et al. 2015 y Szytler et al., 2014). En esta misma línea, mediante el presente trabajo, los autores intentan contribuir a mejorar el intercambio, el reuso, la compartición, el enriquecimiento de datos y la colaboración institucional y académica a nivel de librerías digitales.

### **3. Marco de Trabajo para la Publicación de Datos Enlazados.**

El ciclo de publicación de datos enlazados comprende una serie de componentes y actividades interrelacionadas. En la Figura 2, se presenta el marco de trabajo propuesto para de extracción de metadatos usando OAI-PMH y su publicación como datos enlazados (Piedra y otros, 2014).

- 1. Selección de fuentes de datos**, se refiere a las librerías digitales que son de interés para un determinado proyecto. La Federación de Librerías Digitales define a las bibliotecas como organizaciones que proporcionan los recursos y el personal especializado para seleccionar, estructurar, ofrecer acceso intelectual, interpretar, distribuir, preservar la integridad y asegurar la persistencia en el

tiempo de las colecciones de obras digitales, de tal manera que sean de fácil acceso y económicamente disponibles para su uso por una comunidad o por un conjunto de comunidades (Chinwe & Majesty, 2011).

**2. Cosecha de metadatos desde repositorios:**

2.1 Uso de librería Harvester 2.0 para extraer los metadatos de los repositorios, a través del protocolo OAI-PMH.

2.2 Almacenamiento de datos cosechados en un repositorio relacional y en formato de tripletas OAI.

**3. Modelamiento del vocabulario u ontología.** En esta fase se establecen relaciones de mapeado con otros vocabularios. El re-uso de recursos ontológicos y no-ontológicos es clave para incrementar el grado de interoperabilidad en el escenario de Linked Open Data.

**4. Conversión de datos a formato RDF:**

4.1 Conversión de datos extraídos a formatos estándar, abiertos e interoperables, de manera que se contribuya a facilitar su acceso y reutilización, y se resuelva el problema de disponer de recursos digitales aislados en silos.

4.2 Limpieza de datos generados: actividades que buscan reducir la ambigüedad y purgar la información extraída y generada durante el proceso de conversión.

**5. Enlazado de datos a través de sus relaciones semánticas** (idiomas, organismos, conceptos) con fuentes existentes, de este modo se podrá establecer vínculos entre conjuntos de datos abiertos, y contribuir a una **integración a escala global y generar el efecto de red.**

**6. Publicación y Explotación de datos.** La publicación de datos en la Web mediante tecnologías estándar propuestas por la W3C mejora la accesibilidad, disponibilidad e integración de estos recursos a otros.

## **4. Publicación de Datos Enlazados de repositorios institucionales**

En base al marco de trabajo definido en el apartado anterior, en este punto, se describe el proceso de publicación de datos enlazados de un conjunto de repositorios seleccionados.

### **4.1. Selección de fuentes de datos**

La colección de librerías digitales corresponde a las universidades miembro de la Red de Repositorios del Consorcio Ecuatoriano para el Desarrollo de Internet Avanzado (CEDIA<sup>8</sup>). CEDIA está integrado por más de 30 instituciones de educación superior del Ecuador y tiene múltiples alianzas internacionales con otras redes. El material incluido en los repositorios seleccionados comprende: atlas, CDs, DVDs, ebooks, enciclopedias, folletos, juegos, libros, memorias, revistas, tesis; esto representa una oportunidad inmejorable para analizar los datos contenidos.

<sup>8</sup> <https://www.cedia.org.ec/>

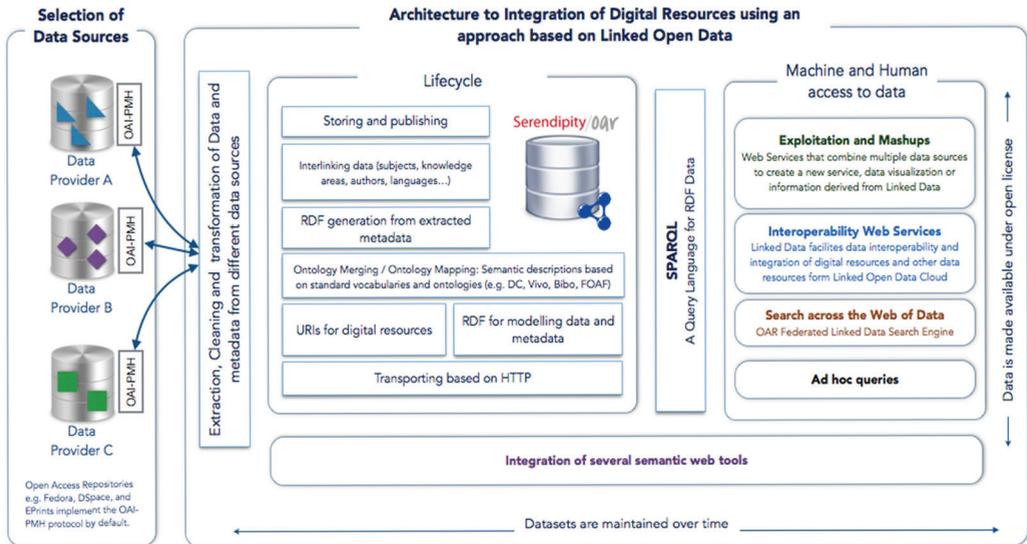


Figura 2 – Visión General del Marco de Trabajo

#### 4.2. Cosecha de metadatos desde repositorios OAI

El protocolo OAI-PMH es la base del modelo de cosecha de metadatos de recursos académicos, definidos según el esquema Dublin Core. La aplicación *Harvester2* fue utilizada para cosechar metadatos a través de OAI-PMH a través del verbo “listRecords”. Los metadatos extraídos se almacenan en una base de datos relacional en forma de tripletas.

#### 4.3. Limpieza de datos

La limpieza de los datos extraídos se realiza con el objetivo de detectar y corregir datos corruptos o erróneos. El proceso consiste en analizar patrones inconsistentes en los datos y ejecutar un esquema de limpieza. Entre los casos detectados está la variación en el formato de ciertos metadatos. Un caso concreto se detectó en el lenguaje de los recursos; se encontraron términos como: “en”, “en\_US”, “eng” e “English” para referirse al lenguaje inglés. También se detectaron variaciones en la descripción de autores. Se consideró como patrón válido a los nombres con la secuencia: <Apellidos, Nombres>. En la Figura 3, se describen ejemplos de patrones inconsistentes y la respectiva corrección.

o	o_syn
Velasco Vaca.Grijalva Rodas, Angelica Marisol.Ruth Cecilia	Velasco Vaca, Angelica Marisol
Velasco Vaca.Grijalva Rodas, Angelica Marisol.Ruth Cecilia	Grijalva Rodas, Ruth Cecilia
Aguilar Salazar, Byron Gloria Ximena Bustamante	Aguilar Salazar, Gloria Ximena
Aguilar Salazar, Byron Gloria Ximena Bustamante	Bustamante, Byron
Cueva Rivera Ximena Edditta	Cueva Rivera, Ximena Edditta
Alvarez V. Achig A., Jacqueline A. Vickys.S	Alvarez V., Jacqueline A.
Alvarez V. Achig A., Jacqueline A. Vickys.S	Achig A., Vickys.S

Figura 3 – Limpieza de Autores

Finalmente, se eliminaron problemas de ambigüedad en los tópicos (dc:subject) añadidos a cada recurso digital. Un esquema de corrección semi-automática pudo ser implementado en base a limpieza de patrones detectados.

#### 4.4. Modelamiento del Vocabulario

Las ontologías y los vocabularios abiertos constituyen el esquema base a partir del cual se describen los recursos y entidades de la Web. Por tanto, se han examinado vocabularios para representar recursos digitales, clasificación de tópicos, descripción de organizaciones, autores, catálogos de datos y repositorios. Las actividades que comprende el modelamiento de una ontología son: (a) mapeo de conceptos, (b) búsqueda de recursos ontológicos y no ontológicos a reusar y (c) Diseño de URIs persistentes.

##### 4.4.1. Mapeo de conceptos.

Esta actividad trata de mapear los principales términos relacionados al dominio de los repositorios digitales institucionales. Entre los conceptos identificados están: (a) Recurso bibliográfico: entidad bibliográfica; (b) Persona: creador o contribuyente de la entidad bibliográfica; (c) Corporación: organización que realizó la publicación de la entidad bibliográfica; (d) Tópicos o temas: tópicos relacionados a la entidad bibliográfica. En el contexto de este trabajo, los metadatos de una obra digital se clasificaron en tres grupos. Cada grupo indica la clase o el ámbito de la información que se guarda en ellos: elementos relacionados al contenido del recurso, elementos relacionados al recurso cuando es visto como propiedad intelectual y elementos relacionados con la instanciación del recurso (ver Tabla 1).

Tabla 1 – Mapeo de metadatos Dublin Core

<b>Contenido</b>		
<i>Título</i>	dc:title	Título del recurso
<i>Tópico</i>	dc:subject	Frases que describen el título o contenido del recurso
<i>Descripción</i>	dc:description	Resumen o descripción del contenido
<i>Fuente</i>	dc:source	Secuencia de caracteres usados para identificar unívocamente un trabajo a partir del cual proviene el recurso actual
<i>Tipo</i>	dc:type	Categoría del recurso
<i>Relación</i>	dc:relation	Es un identificador de un segundo recurso y su relación con el recurso actual
<i>Cobertura</i>	dc:coverage	Característica de cobertura espacial o temporal del contenido intelectual del recurso
<b>Propiedad intelectual</b>		
<i>Autor o creador</i>	dc:creator	Persona responsable de la creación del contenido intelectual del recurso
<i>Editor</i>	dc:publisher	Entidad responsable de hacer que el recurso se encuentre disponible en la red en su formato actual
<i>Otros colaboradores</i>	dc:contributor	Persona u organización que haya tenido una contribución intelectual significativa, pero que esta sea secundaria
<i>Derechos</i>	dc:rights	Nota sobre los derechos de autor

Instanciación		
<i>Fecha</i>	dc:date	Fecha en la cual el recurso se puso a disposición del usuario
<i>Formato</i>	dc:format	Es el formato de datos de un recurso
<i>Identificador</i>	dc:identifier	Secuencia de caracteres para identificar unívocamente al recurso, por ejemplo: URN, URL, ISBN, etc.
<i>Lenguaje</i>	dc:language	Lenguaje del contenido intelectual del recurso

En cuanto a la clasificación temática de los materiales bibliográficos, se mapearon estos conceptos mediante el vocabulario recomendado por la W3C para organizar conocimiento, SKOS. SKOS modela conceptos y esquemas de conceptos, etiquetas léxicas, relaciones semánticas, documentación, colecciones de conceptos, propiedades de mapeado y anotaciones.

#### 4.4.2. *Búsqueda de recursos ontológicos y no ontológicos a reusar*

El reuso de recursos ontológicos y no-ontológicos reduce el tiempo de desarrollo y los costes asociados en esta fase, también contribuye a la calidad de la ontología (Villazón-Terrazas, 2011). Debido a que existen varios vocabularios para modelar recursos bibliográficos, se han seleccionado a aquellos que favorecen la integración e interoperabilidad de datos en la Web:

- RDF Schema<sup>9</sup>, y OWL<sup>10</sup> para describir conceptos del vocabulario.
- Simple Knowledge Organization System (SKOS) para establecer un modelo de organización del conocimiento, taxonomías, y otras jerarquías temáticas. El vocabulario SKOS permite representar conceptos o temas tratados por el recurso.
- DCAT<sup>11</sup> diseñado para facilitar la interoperabilidad entre catálogos de datos publicados en la Web.
- FOAF<sup>12</sup>, desarrollado para representar a personas y organizaciones con sus atributos y relaciones hacia otros conceptos.
- Dublin Core (DC) proporciona un vocabulario de características “base”, capaces de proporcionar información descriptiva básica sobre cualquier recurso.
- Dublin Core Metadata Initiative (DCMI) Metadata Terms<sup>13</sup> usado para representar documentos, así como sus atributos como título, creador, y relaciones con otras entidades.
- Bibliographic Ontology Specification<sup>14</sup> (BIBO) provee los principales conceptos y propiedades para describir citas y referencias bibliográficas.
- VIVO<sup>15</sup> que permite describir diferentes conceptos en el dominio académico y científico.

<sup>9</sup> W3C Recommendation: <http://www.w3.org/TR/rdf-schema/>

<sup>10</sup> W3C Recommendation: <http://www.w3.org/TR/owl2-overview/>

<sup>11</sup> <http://www.w3.org/TR/vocab-dcat>

<sup>12</sup> <http://xmlns.com/foaf/spec/>

<sup>13</sup> <http://dublincore.org/documents/2012/06/14/dcmi-terms/>

<sup>14</sup> <http://purl.org/ontology/bibo/>

<sup>15</sup> <http://vivoweb.org/ontology/core>



- Categories: <http://data.utpl.edu.ec/serendipity/oar/category/>
- Graph: <http://data.utpl.edu.ec/serendipity/oar>
- SPARQL endpoint: <http://data.utpl.edu.ec/serendipity/oar/sparql>

#### 4.5. Generación y Publicación de Datos Enlazados

Para la generación en RDF de los datos cosechados, se desarrolló un generador propio basado en Jena. Un paso importante en el proceso de generación de datos RDF es asignar URIs a los textos extraídos; en este sentido, los metadatos de los recursos bibliográficos se mapearon con los URIs de los términos más apropiados, esto permite que los recursos puedan interoperarse e integrarse con otros conjuntos de datos. En la Tabla 2, se muestra un ejemplo del mapeo de texto correspondiente a tipo de material bibliográfico con su correspondiente URI.

Tabla 2 –Mapping entre términos del dominio y conceptos relacionados

Tipo de recurso	URI del concepto equivalente
Article	<a href="http://purl.org/ontology/bibo/AcademicArticle">http://purl.org/ontology/bibo/AcademicArticle</a>
Article	<a href="http://schema.org/Article">http://schema.org/Article</a>
Book	<a href="http://purl.org/ontology/bibo/Book">http://purl.org/ontology/bibo/Book</a>
Book	<a href="http://schema.org/Book">http://schema.org/Book</a>
Other	<a href="http://purl.org/ontology/bibo/Document">http://purl.org/ontology/bibo/Document</a>
Presentation	<a href="http://purl.org/ontology/bibo/Slideshow">http://purl.org/ontology/bibo/Slideshow</a>
Presentation	<a href="http://vivoweb.org/ontology/core#Presentation">http://vivoweb.org/ontology/core#Presentation</a>
Technical Report	<a href="http://purl.org/net/nknouf/ns/bibtex#Techreport">http://purl.org/net/nknouf/ns/bibtex#Techreport</a>
Technical Report	<a href="http://purl.org/ontology/bibo/Report">http://purl.org/ontology/bibo/Report</a>
Thesis	<a href="http://purl.org/ontology/bibo/Thesis">http://purl.org/ontology/bibo/Thesis</a>
Thesis	<a href="http://purl.org/ontology/bibo/ThesisDegree">http://purl.org/ontology/bibo/ThesisDegree</a>

Una vez que se han generado y almacenado los datos RDF, se pueden elegir algunas interfaces o aplicaciones web para mostrar estos datos en formato legible para las personas. Existen algunas herramientas que se conectan al repositorio RDF y presentan los datos recuperados como páginas Web o como esquemas gráficos. Una de las herramientas más populares para mostrar datos RDF en formato tabular es Pubby<sup>18</sup>, aplicación java que es utilizada por reconocidos repositorios de tripletas. Como parte de la validación del proceso realizado, se implementaron algunas consultas SPARQL.

*Consulta 01: Material bibliográfico relacionado a temáticas de software*

Para obtener los recursos académicos relacionados a temáticas de software, se utilizan

<sup>18</sup>Pubby está disponible en: <http://wif05-03.informatik.uni-mannheim.de/pubby/>

diferentes filtros sobre el título, la descripción y los tópicos de cada obra.

```

SELECT DISTINCT ?link ?title ?subject
WHERE {
    ?bibresource a <http://purl.org/dc/terms/BibliographicResource> .
    ?bibresource <http://purl.org/dc/terms/title> ?title .
    ?bibresource <http://purl.org/dc/terms/subject> ?subjectresource .
    ?bibresource <http://purl.org/dc/terms/description> ?desc .
    OPTIONAL {
    ?bibresource <http://purl.org/dc/terms/identifier> ?link .
        ?subjectresource rdfs:label ?subject.
    }
    FILTER ((REGEX(?title, „software“, „i“)
        || REGEX(?subject, „software“, „i“)) &&
        (REGEX(?desc, „engenharia“, „i“) ||
        REGEX(?desc, „engineering“, „i“) ||
        REGEX(?desc, „Ingeniería“, „i“)))
} GROUP BY ?bibresource ORDER BY ?title

```

### *Consulta 02: Top 10 de los temas más referidos en el campo de Software*

Cada recurso digital es asociado a uno o varios tópicos, la relación es establecida a través del metadato DC: <http://purl.org/dc/terms/subject>.

```

SELECT DISTINCT ?subject
WHERE {
    ?bibresource a <http://purl.org/dc/terms/BibliographicResource>.
    ?bibresource <http://purl.org/dc/terms/subject> ?subjectresource .
    OPTIONAL{
        ?subjectresource rdfs:label ?subject .
    }
    FILTER REGEX(?subject, „software“, „i“)
} ORDER BY DESC(count(?subject))

LIMIT 10

```

**Resultados:** Software-Desarrollo, Software libre, Software-diseño, Software educativo, Software WASP 7, Modelo matemático, Software MVC, Software Gravmag V.1.7-Utilización, Software para bibliotecas, Software web.

#### 4.6. Enriquecimiento y reconciliación de datos

Para mejorar el descubrimiento de los recursos es necesario crear enlaces RDF hacia recursos externos publicados en la nube de Linked Open Data y con datos de repositorios externos. Se han enlazado los tópicos y palabras claves de cada recurso digital con el contenido de vocabularios controlados y esquemas de clasificación. Se usó SILK y un componente de enlazado desarrollado a medida. De forma concreta, se han creado enlaces externos con la Nomenclatura de la UNESCO, y con DBpedia en inglés, DBpedia en español y DBpedia Latinoamérica. En la Tabla 3, se indica un subconjunto de los datos mapeados.

Tabla 3 – Ejemplo de enlazado de Universidades con fuentes externas

Texto a enlazar	URI del recurso externo relacionado
Universidad Técnica Particular de Loja	<a href="http://dbpedia.org/resource/Universidad_T%C3%A9cnica_Particular_de_Loja">http://dbpedia.org/resource/Universidad_T%C3%A9cnica_Particular_de_Loja</a> <a href="http://es-la.dbpedia.org/resource/Universidad_T%C3%A9cnica_Particular_de_Loja">http://es-la.dbpedia.org/resource/Universidad_T%C3%A9cnica_Particular_de_Loja</a>
Universidade Federal do Rio Grande do Sul	<a href="http://dbpedia.org/resource/Universidade_Federal_do_Rio_Grande_do_Sul">http://dbpedia.org/resource/Universidade_Federal_do_Rio_Grande_do_Sul</a>
Universidade de São Paulo	<a href="http://es-la.dbpedia.org/resource/Universidad_de_S%C3%A3o_Paulo">http://es-la.dbpedia.org/resource/Universidad_de_S%C3%A3o_Paulo</a>

### 5. Explotación de Datos Enlazados

La gestión de grandes colecciones de recursos digitales debe asegurar la localización exitosa de los contenidos más adecuados para un usuario. En este punto, se presentan dos servicios que aprovechan el potencial de los datos enlazados y de las tecnologías de la Web Semántica con el objetivo de mejorar la recuperación de recursos Web.

#### 5.1. Sugerencia de tópicos

Mediante un algoritmo iterativo, basado en consultas SPARQL, es posible recorrer la red de conceptos relacionados a un determinado término (tópico o área de conocimiento).

En la Tabla 4 se muestra un conjunto de términos cercanos al concepto “Software Engineering.” En este caso, se ha aprovechado el contenido dinámico del repositorio DBPedia y se han utilizado las relaciones jerárquicas establecidas entre dos conceptos SKOS. DBPedia tiene datos estructurados sobre tipo de cosas: lenguajes, localizaciones, personas, conceptos, organizaciones, etc. Se alimenta del contenido creado socialmente en la enciclopedia más grande del mundo, la Wikipedia.



## 6. Conclusiones y Trabajos Futuros

En un contexto de acceso abierto, los proveedores de recursos digitales abiertos deben adoptar un modelo que mejore la integración de repositorios, de manera que puedan soportar estándares de metadatos formales y abiertos para la descripción de recursos, y cuyo nivel de especificidad, granularidad y complejidad sea realizable. El modelo a adoptar debe respetar y garantizar condiciones de autonomía local en un marco de interoperabilidad tecnológica global. El enfoque de Linked Data potencia la interoperabilidad e integración en un contexto como la Web: alta heterogeneidad en un contexto distribuido.

En este trabajo se presentó un marco tecnológico y un método para la publicación y enlazado de datos bibliográficos digitales, siguiendo las buenas prácticas de publicación de datos enlazados en un espacio global. El proceso definido fue aplicado a un conjunto de repositorios digitales de instituciones ecuatorianas; las actividades desarrolladas aseguran la reproducibilidad del ciclo de publicación de datos enlazados sobre cualquier otro repositorio OAI.

Es de particular interés de los autores, contribuir a mejorar la visibilidad y acceso a los materiales digitales que se usan o generan en el ámbito académico y científico, por esta razón, se continúan diseñando servicios que explotan el potencial de los datos enlazados. De forma específica, se están explorando otras redes de repositorios federados de Latinoamérica, con el objetivo de conformar un *hub* central que proporcione información integrada de las obras y líneas de trabajo de cada institución de la región; además está en construcción un buscador semántico basado en facetas, que permite la exploración de contenido a través de categorías que han sido asociadas a los recursos. Este tipo de navegación permite especificar filtros sobre distintos atributos de los documentos y así cualquier usuario podrá encontrar más fácilmente recursos relevantes.

## Agradecimientos

Este trabajo ha sido desarrollado con el apoyo de la Comisión Europea a través del proyecto ESVI-AL - “Educación Superior Virtual Inclusiva – América Latina” del programa ALFA III, y el Consorcio Ecuatoriano para Desarrollo de Internet Avanzado (CEDIA) a través de la convocatoria CEPRA y el Grupo de Trabajo de Repositorios Abiertos. Los autores agradecen el soporte recibido por el equipo de Tecnologías Avanzadas de la Web y Sistemas Basados en Conocimiento de la UTPL.

## Referencias Bibliográficas

- Anibaldi, S., Jaques, Y., Celli, F., Stellato, A., Keizer, J. (2015) Migrating bibliographic datasets to the Semantic Web: The AGRIS case. *Semantic Web*, 6(2), 113-120.
- Baker, T., Bermés, E., Coyle, k., Dunsire, G., Isaac, A., Murray, P., y otros (2011). Library Linked Data Incubator Group Final Report. Obtenido de [http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/#Benefits\\_of\\_the\\_Linked\\_Data\\_Approach](http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/#Benefits_of_the_Linked_Data_Approach).
- Berners-Lee, Tim; Hendler, J.; Lassila, Ora (2001). *The Semantic Web*, *Scientific American*, 284(5), 34-43.

- Berners-Lee, Tim (2006). Linked Data-Design Issues. Obtenido de: <http://www.w3.org/DesignIssues/LinkedData.html>
- Chinwe, V. A., & Majesty, I. E. (2011). Digital library deployment in a university. *Library Hi Tech*, 29(2), 373-386. doi:<http://dx.doi.org/10.1108/0737883111138233>
- Gendt, M.; Isaac, A.; Meij, L. & Schlobach, S. (2006). Semantic Web Techniques for Multiple Views on Heterogeneous Collections: A Case Study. *Research and Advanced Technology for Digital Libraries*. Springer Berlin Heidelberg, vol. 4172.
- Francesconi, E.; Faro, S. ; Marinai, E. & Perugi, G. (1008). A Methodological Framework for Thesaurus Semantic Interoperability. *Proceeding of the Fifth European Semantic Web Conference*, 76-87.
- Heath, Tom & Bizer, Christian (2011). Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1), 1-136. Morgan & Claypool.
- Ian Rowlands and David Bawden (1999), „Digital Libraries: a conceptual framework,“ *Libri Journal*, vol. 49, pp. 192-202.
- IEEE (1990). IEEE Standard Glossary of Software Engineering Terminology. IEEE Standards Board, New York.
- Lenzerini, M. (2002). Data Integration: A Theoretical Perspective. In *Proceedings of the Twenty-first ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, June 3-5, Madison, Wisconsin, USA. doi: 10.1145/543613.543644
- Malakhov, D., Serebryakov, V., Teymurazov, K., Shorin, O. (2014) Semantic integration of bibliographic records. *CEUR Workshop Proceedings*, 35-41.
- Piedra, N. Tovar, E. López, J. Chicaiza, J. (2014). Consuming and producing linked Open Data: The case of Opencourseware. *Emerald EarlyCite*. DOI: 10.1108/PROG-07-2012-0045
- OAI (2015). The Open Arhives Initiative Protocolo for Metadata Harvesting, Implementation GuideLines. Obtenido de: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.
- Sztyler, T., Huber, J., Noessner, J., Murdock, J., Allen, C., Niepert, M.(2014). LODÉ: Linking digital humanities content to the web of data. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 423-424.
- Tovar, E. & Piedra, N. (2014). Open Educational Resources in Engineering Education: a perspective to improve Reusability of Resources and Data. *IEEE Transactions on Education*, 57(4). DOI: <http://10.1109/TE.2014.2359257>
- Villazón-Terrazas, B. (2011). A Method for Reusing and Re-engineering Non-ontological Resources for Building Ontologies.
- Zengenene, D., Casarosa, V., Meghini, C. (2014). Towards a Methodology for Publishing Library Linked Data. *Communications in Computer and Information Science*, 385 CCIS, 81-92.