

Un generador de metadatos openaire conforme con el repositorio nacional de México

An openaire metadata generator according to the national repository of México

<https://doi.org/10.5281/zenodo.4451296>

AUTORES: Jared David Tadeo Guerrero Sosa¹

Daniel Sánchez Ferriz²

Víctor Hugo Menéndez Domínguez^{3*}

María Enriqueta Castellanos Bolaños⁴

Jorge Ricardo Gómez Montalvo⁵

DIRECCIÓN PARA CORRESPONDENCIA: mdoming@correo.uady.mx

Fecha de recepción: 15 / 09 / 2020

Fecha de aceptación: 28 / 12 / 2020

RESUMEN

La Ciencia Abierta tiene como objetivo incrementar el acceso a los trabajos científicos, tecnológicos, educativos y de innovación. Esta tarea ha sido posible gracias a los repositorios digitales y sus metadatos basados en estándares que facilitan la interoperabilidad. En México existe el repositorio nacional, compuesto por repositorios institucionales, mismo que ha establecido su propio esquema de metadatos basado en OpenAIRE. Sin embargo, se identificó que no existía una herramienta para la generación de metadatos compatible con dicho repositorio. Este trabajo presenta una herramienta

¹Maestro en Ciencias de la Computación, Universidad Autónoma de Yucatán, México, jared.guerrero@correo.uady.mx

²Maestro en Ciencias de la Computación, Universidad Autónoma de Yucatán, México, dsf8991@gmail.com

³Doctor en Tecnologías Informáticas Avanzadas, Universidad Autónoma de Yucatán, México, mdoming@correo.uady.mx

⁴Maestra en Gestión de Tecnología de Información, Universidad Autónoma de Yucatán, México, enriqueta.c@correo.uady.mx

⁵Doctor en Redes, Telecomunicaciones, Sistemas y Arquitecturas, Universidad Autónoma de Yucatán, México, jgomez@correo.uady.mx

tecnológica que soluciona la problemática, incorporando los catálogos de lenguajes controlados definidos por el repositorio nacional. La metodología consiste en el análisis de la documentación del repositorio nacional, la creación de las plantillas para metadatos y en la construcción de la interfaz del usuario encargada de generar los metadatos en formato XML, los cuales fueron validados exitosamente. Los beneficios aseguran una mayor consistencia con lo solicitado a los repositorios institucionales para la interoperabilidad de sus recursos.

Palabras clave: *Información/biblioteca, planificación, Informatización de archivos, Lenguaje de indexación.*

ABSTRACT

Open Science aims to increase access to scientific, technological, educational and innovation works. This task has been possible thanks to digital repositories and their standards-based metadata that facilitate interoperability. In Mexico there is a national repository, made up of institutional repositories, which has established its own metadata scheme based on OpenAIRE. However, it was identified that there was no tool for generating metadata compatible with said repository. This work presents a technological tool that solves the problem, incorporating the catalogs of controlled languages defined by the national repository. The methodology consists of the analysis of the national repository documentation, the creation of templates for metadata and the construction of the user interface in charge of generating the metadata in XML format, which were successfully validated. The benefits ensure greater consistency with what is requested from institutional repositories for the interoperability of their resources.

Keywords: *Information/library planning, Archive automation, Indexing languages.*

INTRODUCCIÓN

En los últimos años hemos sido testigos de la importancia de difundir el conocimiento generado en las instituciones, universidades y centros de investigación: muchas investigaciones se han desarrollado gracias al conocimiento generado y difundido de trabajos previos. La Ciencia Abierta (CONACYT, 2017b) promueve incrementar y facilitar el acceso a las investigaciones científicas, materiales e información resultado de estos

procesos, que hayan sido financiados con recursos públicos. En este sentido los repositorios digitales han tomado un papel fundamental en dicho proceso. Un repositorio es una plataforma digital centralizada que, siguiendo estándares internacionales, almacena, mantiene, preserva y disemina la información científica, tecnológica y de innovación, la cual se deriva de las investigaciones, productos educativos y académicos y desarrollos tecnológicos (CONACYT, 2014).

En México, el repositorio nacional (CONACYT, 2014) tiene como principal función el acopio, preservación, gestión y acceso electrónico de información y contenidos de calidad, incluyendo aquellos de interés social y cultural que se producen en México con recursos públicos. Mediante los estándares OAI-PMH y OpenAIRE recupera la información de los repositorios de las instituciones de los sectores social, privado y gubernamental que contienen información académica, científica, tecnológica y de innovación.

Sin embargo, generar las estructuras XML necesarias para la interacción con tales estándares no es una tarea sencilla. Además, se suelen cometer muchos errores para lograr la completa conformidad a la especificación dada por el CONACYT que está basada en OpenAIRE (INFOTEC, 2017).

En este trabajo se presenta una propuesta para la generación de metadatos OpenAIRE acordes a las especificaciones del repositorio nacional, empleando un formulario web que consume los catálogos REST definidos por CONACYT para emitir una estructura XML compatible con OpenAIRE y conforme con las especificaciones del repositorio nacional de México. Se pretende que la herramienta facilite la generación de metadatos de la información disponible en un repositorio para que sea compatible con el repositorio nacional.

MARCO CONCEPTUAL

La interoperabilidad en repositorios, los metadatos y los estándares Dublin Core y OpenAIRE son conceptos que están impulsando el crecimiento y el desarrollo de la Ciencia Abierta. Estos elementos están originando un cambio relevante en la forma de cómo se realiza el intercambio y la reutilización del conocimiento de una forma estandarizada y sostenida.

En esta sección se relatan, de forma abreviada, los conceptos junto con otras cuestiones actuales relativas a las áreas de conocimiento afines a la propuesta.

Interoperabilidad entre repositorios

Es la capacidad de intercambio de contenido entre dos o más sistemas. Se ha analizado este concepto desde cuatro aspectos fundamentales: el sintáctico, el semántico, el de infraestructura y el estructural (Gómez Dueñas, 2010), siendo el último el más importante en este trabajo, ya que se refiere a la comunicación entre repositorios heterogéneos por medio del uso de protocolos especializados, como OAI-PMH e IMS Digital Repositories Interoperability.

OAI-PMH es un mecanismo de baja barrera para la interoperabilidad del repositorio (Open Archives Initiative Organization, 2015). Los proveedores de datos son repositorios que exponen metadatos estructurados según un estándar (como la especificación Dublin-Core) a través de OAI-PMH. Los proveedores de servicios son los encargados de la realización de solicitudes de servicio OAI-PMH para recolectar metadatos.

Prácticamente, es un conjunto de servicios que son invocados dentro de HTTP, los cuales pueden hacer uso de argumentos requeridos, opcionales o exclusivos. Los verbos permiten identificar a un repositorio, obtener los identificadores de los recursos, presentar los metadatos de los recursos, obtener el registro de un recurso específico, identificar los esquemas de metadatos utilizados por el repositorio y recuperar los conjuntos utilizados por el repositorio para clasificar los registros. Además, existen herramientas que facilitan la recuperación de los metadatos de los recursos de los repositorios que utilizan OAI-PMH (Guerrero, Sánchez, Menéndez, Castellanos, & Gómez, 2019; Medina Moreno, Menéndez Domínguez, & Palacio Cinco, 2013).

Por su parte, IMS Digital Repositories Interoperability, es un estándar creado por el IMS Global Learning Consortium, cuya finalidad es facilitar el acceso a los contenidos en repositorios que cuenten con contexto educativo, pero igualmente con otro tipo de sistemas, como los sitios web de búsqueda (López Guzmán, 2005).

Metadatos

Al hablar de documentos en formato digital almacenados en un repositorio, son los metadatos los que cumplen las funciones de identificación y descripción, así como del registro de la información para el desarrollo adecuado de la preservación al aportar, entre

otros elementos, contenido sobre el tipo de formato y modificaciones realizadas sobre el mismo (Méndez Rodríguez, 2001).

El Consejo Internacional de Archivos (CIA) reconoce que los metadatos cuentan con la función principal de la preservación digital, es decir, conseguir que los documentos sean y permanezcan auténticos, completos, accesibles y potencialmente reusables. De igual manera, indica que la preservación correcta estará asegurada si los metadatos indican el nombre y la versión del formato de cada archivo y la identificación del software con el que ha sido creado y con el que se ha realizado la última modificación (Candás Romero, 2006).

Además, el CIA plantea tres tipos de metadatos, según sus funciones:

- *Metadatos descriptivos*: Son creados con los propios documentos o por la misma organización que los ha creado. Deben registrar al autor, la fecha de creación, el título y las palabras clave.
- *Metadatos de archivo*: Añadidos para facilitar la gestión de los documentos tras su creación, incluyen elementos como la última revisión, fecha de aprobación del archivo, condiciones de uso, entre otros.
- *Metadatos técnicos*: Necesarios para comprender y procesar los documentos, pueden ser, por ejemplo, formato de archivo y fecha de la última migración de formato.

A continuación, se presentan dos estándares de metadatos que toman relevancia para la interoperabilidad entre sistemas, especialmente el intercambio de datos entre repositorios de documentos digitales.

Dublin Core

Dublin Core nace como producto del trabajo cooperativo de ámbito internacional, promovido en su primera fase por la OCLC (Online Computer Library Center) y el NCSA (National Center for Supercomputing Applications), cuyo objetivo principal fue crear un conjunto de elementos que permitieran la descripción de recursos electrónicos con el fin de facilitar su búsqueda y recuperación (Julie, 2005).

Originalmente se diseñó como un conjunto de etiquetas que deberían ser generadas por el autor del documento (HTML) con el fin de facilitar su identificación y posterior recuperación en Internet. No obstante, este modelo ha llamado la atención de diversas

comunidades de profesionales interesados en la descripción de recursos en museos, bibliotecas y organizaciones comerciales.

En realidad, Dublin Core no es un sistema de conversión de metadatos, sino que es un sistema que define un conjunto básico de atributos que sirven para describir los recursos existentes de cualquier tipo en la red (DCMI, 2012).

El estándar Dublin Core define un conjunto base de 15 metadatos que constituyen el nivel simple. Estos metadatos describen el contenido del documento, la propiedad intelectual y la información para la creación de instancias. En la Tabla 1 se muestran los 15 metadatos del estándar Dublin Core (DCMI, 2012).

Tabla 1. Elementos del estándar Dublin Core.

Elemento	Etiqueta
Título	dc:title
Creador	dc:creator
Materia y palabras clave	dc:subject
Descripción	dc:description
Editor	dc:publisher
Colaborador	dc:contributor
Fecha de publicación	dc:date
Tipo de recurso	dc:type
Formato	dc:format
Identificador del recurso	dc:identifier
Fuente	dc:source
Idioma	dc:language
Relación	dc:relation
Cobertura	dc:coverage
Derechos de autor	dc:rights

Fuente: DCMI (2012), DCMI: Dublin Core Metadata Element Set, Version 1.1: Reference Description.

En Dublin Core, cada elemento es opcional y puede repetirse. No hay un orden específico de aparición. Su contenido puede regirse por esquemas de codificación, vocabularios controlados o estar abierto a las descripciones de texto en lenguaje natural. El estándar proporciona un conjunto de directrices para su implementación en los formatos XHTML, XML y RDF, lo que facilita el intercambio de los recursos, además de que sirve de base a otros protocolos de intercambio de datos como OAI-PMH.

OpenAIRE

En diciembre del 2010 surgió el estándar Open Access Infrastructure for Research in Europe (OpenAIRE). La vicepresidenta de la Comisión Europea, Neelie Kroes, mencionó que OpenAIRE, al permitir el acceso abierto, fomenta y permite una mayor participación de la sociedad en la ciencia (Rettberg & Schmidt, 2012).

OpenAIRE es una red de repositorios, cuyos principales objetivos se enfocan en el depósito en acceso abierto de toda la producción científica, principalmente la generada por proyectos de financiación de la Comisión Europea; construir las estructuras de apoyo a los investigadores para el depósito de publicaciones de investigación; y el establecimiento de una infraestructura digital para el acceso a la información científica (Rettberg & Schmidt, 2012).

Con el fin de lograr sus objetivos, OpenAIRE, establece lineamientos que permiten una generalización en la codificación de los metadatos. Los lineamientos de OpenAIRE ponen a disposición de los administradores de repositorios esquemas de codificación de metadatos a través del protocolo OAI-PMH para poder integrarse con la infraestructura de OpenAIRE.

OpenAIRE establece lineamientos para repositorios de documentos, repositorios de datos y sistemas CRIS. Para fines del presente trabajo, se exponen los asociados a los repositorios de documentos. Los campos de uso se indican de la siguiente manera (OpenAIRE, 2017):

- *M (Mandatory)*: El campo es obligatorio, es decir, requiere estar presente en todos los registros de metadatos.
- *MA (Mandatory when applicable)*: El campo es obligatorio cuando corresponda.
- *R (Recommended)*: El campo es recomendable.
- *(Optional)*: El campo es opcional, simplemente presenta información adicional.

La mayoría de los campos para la descripción de metadatos son utilizados por el conjunto base de Dublin Core.

OpenAIRE utiliza el prefijo de metadatos *oai_dc*, y es dependiente de una sintaxis en forma de URI y definida como el espacio de nombres *info:eu-repo* que en conjunto con otros elementos forman un vocabulario controlado para la identificación de proyectos, financiadores, publicaciones y conjuntos de datos (OpenAIRE, 2017). En la Tabla 2 se expone la nomenclatura de los elementos de OpenAIRE para repositorios de literatura en su versión 3. La versión 4, liberada a finales de 2018, extiende el listado de metadatos e incluye nuevos prefijos y espacios de nombres.

OpenAIRE para el repositorio nacional de México

En México existe el repositorio nacional, perteneciente al Consejo Nacional de Ciencia y Tecnología (CONACYT), el cual se define como una plataforma digital encargada de proporcionar acceso abierto en texto completo a una amplia variedad de recursos de información de tipo académica, científica y tecnológica (CONACYT, 2020).

Entre los materiales que se pueden consultar se encuentran artículos de revistas científicas, tesis elaboradas en instituciones de educación superior, protocolos de investigación, memorias de congresos y patentes, al igual que otros documentos producidos en la República Mexicana con el soporte de fondos públicos.

Tabla 2. Elementos del estándar OpenAIRE.

Elemento OpenAIRE	Uso	Etiqueta Dublin Core	Refinamiento por vocabulario
Título	M	dc:title	
Creador	M	dc:creator	
Identificador de proyecto	MA	dc:relation	info:eu-repo/grantAgreement/
Nivel de acceso	M	dc:rights	info:eu-repo/semantics/
Condición de licencia	R	dc:rights	
Fecha de fin de embargo	MA	dc:date	info:eu-repo/date/embargoEnd/
Identificador alternativo	R	dc:relation	info:eu-repo/semantics/altIdentifier/
Referencia de publicación	R	dc:relation	info:eu-repo/semantics/reference/

Referencia de conjunto de datos	R	dc:relation	info:eu-repo/semantics/dataset/
Materia	MA	dc:subject	
Descripción	MA	dc:description	
Editor	MA	dc:publisher	
Colaborador	R	dc:contributor	
Fecha de publicación	M	dc:date	
Tipo de publicación	M	dc:type	info:eu-repo/semantics/
Versión de la publicación	R	dc:type	info:eu-repo/semantics/
Formato	R	dc:format	
Identificador del recurso	M	dc:identifier	
Fuente	R	dc:source	
Idioma	R	dc:language	
Relación	O	dc:relation	
Cobertura	R	dc:coverage	
Audiencia	R	dc:audience	

Fuente: OpenAIRE (2017), OpenAIRE Guidelines Release 3.0.

El repositorio nacional está compuesto por más de 100 repositorios institucionales. Con la finalidad de consolidar el funcionamiento del repositorio nacional con los repositorios institucionales, la interoperabilidad se asegura en los niveles de metadatos, de contenido, de red, de estadísticas y datos de uso, de identificadores, de objeto y de semántica (CONACYT, 2014). A continuación, se describen los niveles más importantes para este trabajo.

- *A nivel de metadatos:* Se refiere a la integración de los metadatos bajo un esquema compatible con Dublin Core, y el protocolo para la interoperabilidad OAI-PMH, complementado con OpenAIRE.
- *A nivel de red:* Facilita la creación y el mantenimiento de una red de repositorios con la adopción de OAI-PMH y OpenAIRE.

Con el objetivo de establecer la consistencia en los datos y cumplir con la sintaxis de los metadatos, el CONACYT ha establecido una serie de catálogos, cuya responsabilidad de uso depende de los repositorios institucionales (CONACYT, 2017a; INFOTEC, 2018).

Tabla 3. Elementos exclusivos del repositorio nacional para la cosecha de metadatos.

Elemento repositorio nacional	Uso	Etiqueta Dublin Core
Identificador del autor	M	dc:creator
Identificador del colaborador	MA	dc:contributor
Citación	R	dc:relation

Fuente: INFOTEC (2017), Manual de mejores prácticas.

Las especificaciones del uso de OpenAIRE para el repositorio nacional abarca los campos presentados en la Tabla 2, además de elementos adicionales, exclusivos del CONACYT (INFOTEC, 2017), los cuales se muestran en la Tabla 3. Además se han establecido atributos adicionales para ciertos metadatos, como es el caso del atributo id para el metadato contributor. Es decir CONACYT ha establecido un perfil de aplicación de OpenAIRE tanto para los metadatos de literatura como para los metadatos para la publicación de datos.

Herramientas basadas en estándares de metadatos

Debido a la popularidad que goza Dublin Core, existen herramientas que facilitan la creación de metadatos basados en dicho estándar.

Por ejemplo, dublincoregenerator de Nick Steffel (Steffel, 2011) se basa en los 15 elementos básicos de Dublin Core, permitiendo la repetición de las etiquetas y emitiendo salidas en los formatos XML, HTML y XHTML.

Otro generador de metadatos Dublin Core es Dublin Core Metadata Gen, herramienta que prepara metadatos para su inserción en documentos HTML y XHTML (Bravo, 2005). Sus limitantes son la ausencia de los metadatos de cobertura, formato y origen, además de la falta de repetibilidad de los metadatos, en caso de ser necesario.

Por otra parte, en la plataforma GitHub (GitHub Inc., 2020) se encuentran más de 100 proyectos asociados con Dublin Core, siendo la mayoría desarrollados en Python y PHP.

También en GitHub se encuentran más de 50 proyectos relacionados con OpenAIRE, pero se encuentran en proceso o están orientados a realizar procedimientos específicos, por ejemplo, proceso de clasificación de textos científicos.

Sin embargo, no se han encontrado herramientas Dublin Core y OpenAIRE que se adapten a las necesidades para la cosecha de metadatos en el repositorio nacional del CONACYT por sus características particulares. Debido a esta necesidad, se ha construido una herramienta que cumple con este propósito, permitiendo llenar los metadatos dependientes de una sintaxis específica y del uso de catálogos de manera correcta.

METODOLOGÍA

A continuación, se describen los pasos para la elaboración del generador de metadatos, partiendo desde el análisis de la documentación del repositorio nacional de México hasta la implementación de la solución propuesta. En la Figura 1 se presenta la arquitectura basada en la metodología.

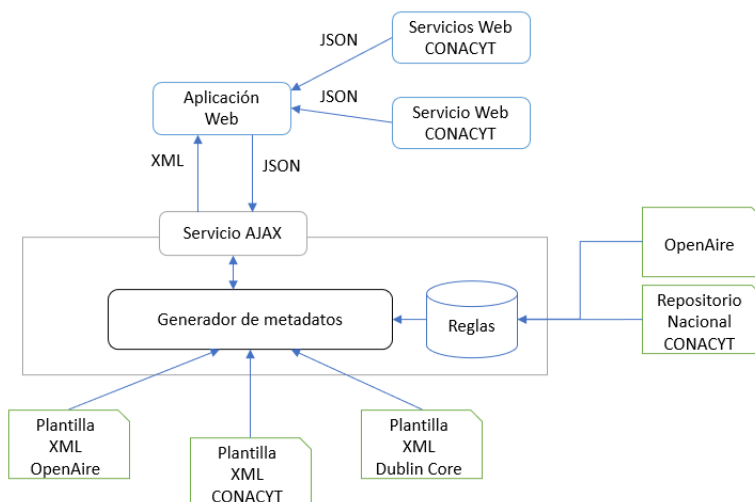


Figura 1. Arquitectura general propuesta.

Fuente: Autor

Análisis de la documentación del repositorio nacional

Un análisis empírico de las notas y observaciones expresadas en la documentación generada por CONACYT para la normatividad y los manuales para la interoperabilidad y desarrollo de los repositorios institucionales ha permitido generar un conjunto de reglas

relacionadas a la completitud y la corrección de los metadatos que son necesarios para el almacenamiento de un recurso en el repositorio nacional.

Las reglas permiten validar los valores y las relaciones existentes entre varios metadatos. En algunos casos, las reglas fueron extendidas para resaltar la obligatoriedad de ciertos metadatos. Por ejemplo, se debe indicar al menos un autor del recurso, lo que implica que también se debe proporcionar un identificador, que debe cumplir con un formato (su Clave Única de Registro de Población, Número de Identificación para Extranjeros o el ORCID).

Creación de plantillas para metadatos

También se ha elaborado una colección de plantillas asociadas a la documentación. Las plantillas, junto con las reglas, definen un perfil de aplicación que puede utilizarse para generar una estructura de metadatos conforme, en este caso, a la especificación del repositorio nacional de CONACYT. La herramienta es lo suficientemente flexible para adaptarla a nuevos contextos mediante la modificación de perfiles de aplicación, que está basado en XML. De esta forma se pueden generar estructuras de metadatos basadas en Dublin Core, OpenAIRE u otra más específica.

Construcción de la interfaz del usuario

A partir del análisis de la documentación y de las plantillas para metadatos, es necesario identificar los metadatos solicitados por el CONACYT y sus respectivos tipos de uso (obligatorio, obligatorio si es aplicable, recomendado u opcional), nivel de repetibilidad y si utiliza algún catálogo. Los catálogos que se utilizaron son:

- Niveles de acceso
- Referencias de publicación
- Referencias de conjuntos de datos
- Tipos de publicación
- Formatos de publicación
- Audiencias
- Áreas, campos, disciplinas y subdisciplinas de conocimiento

La interfaz web consiste en un formulario HTML que, en conjunto con funciones construidas en el lenguaje de programación JavaScript, es posible asignar valores a los metadatos y construir un archivo en formato XML siguiendo las reglas establecidas por el protocolo para la interoperabilidad OAI-PMH y OpenAIRE.

Los elementos del formulario se clasificaron en tres:

- **Metadato simple.** El valor ingresado en el campo por el usuario es insertado en el metadato correspondiente, sin añadir algún valor adicional. En esta clasificación, se encuentran los metadatos para el título, los autores, el identificador del proyecto, la condición de la licencia, las materias (palabras clave), la descripción o resumen, el editor, los colaboradores, la fecha de publicación, el identificador de publicación, el identificador del recurso, la fuente, el idioma, la relación y la cobertura. En la Figura 2 se presenta un ejemplo del llenado del metadato de título, y su salida.

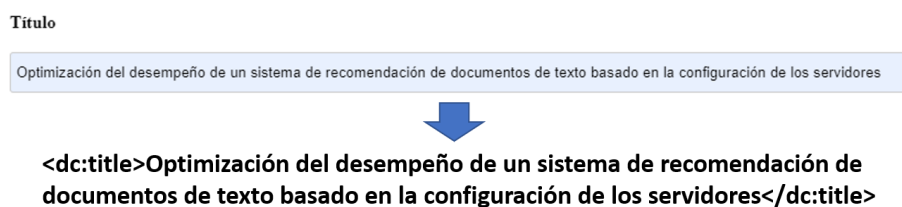


Figura 2. Ejemplo de metadato simple.

Fuente: Autor

- **Metadato con estructura específica simple.** El valor se compone de dos elementos: un vocabulario controlado y un término, obtenido del valor seleccionado por el usuario a partir de un menú de opciones. En esta clasificación se encuentran los metadatos para el nivel de acceso, el área de conocimiento, el campo de conocimiento, la disciplina, la subdisciplina, el tipo de resultado científico, la versión de la publicación, el formato y la audiencia. En la Figura 3 se muestra un ejemplo de llenado del metadato del nivel de acceso y su salida.



Figura 3. Ejemplo de metadato con estructura específica simple.

Fuente: Autor

- **Metadato con estructura específica compuesto.** Se utiliza para los identificadores alternativos, las referencias de publicación y las referencias de conjuntos de datos.

Se compone de un vocabulario controlado, un esquema el cual se obtiene del valor del elemento seleccionado por el usuario en un menú de opciones y el valor asignado por el usuario para el metadato. En la Figura 4 se presenta un ejemplo de llenado del metadato de la referencia de publicación y su salida.



Figura 4. Ejemplo de metadato con estructura específica compuesto.

Fuente: Autor

A continuación, se enlistan las tareas necesarias que se construyen en JavaScript:

- Consulta a los catálogos del repositorio nacional. Cada catálogo utilizado debe ser recuperado y almacenado en un menú de opciones. El valor de cada opción es el identificador de cada elemento recuperado por el catálogo. Además, los menús de opciones que utilizan valores de los catálogos deben llenarse al momento de la ejecución de la interfaz.
- Permitir ingresar más de un valor a los metadatos donde la repetibilidad es permitida.
- Validar si la publicación se encuentra en embargo. Si el usuario indicó que la publicación se encuentra en embargo, deberá ingresar la fecha de fin de embargo.
- Creación del archivo XML. Recuperar todos los valores asignados por el usuario en los metadatos y crear el archivo con las etiquetas y la estructura establecida por OAI-PMH y OpenAIRE.

RESULTADOS

Con base en la metodología presentada en la sección anterior, en la Figura 5 se presenta parte de la interfaz creada.

Con el objetivo de validar la funcionalidad de la herramienta propuesta, se replicó utilizando la interfaz diseñada el registro de los metadatos de un artículo del Repositorio Digital Institucional de la Universidad Autónoma de Yucatán (REDI-UADY). Este repositorio es uno de los más de 100 que componen al repositorio nacional. La salida generada se valida en tres niveles: sintaxis XML, interoperabilidad y sintaxis para el repositorio nacional.

El registro replicado se encuentra disponible en el repositorio nacional (CONACYT, 2018), y se presenta en la Figura 6, donde se indica, en el último renglón, que el recurso proviene de REDI-UADY.

The image shows a web form titled "Generador de metadatos OpenAIRE para el repositorio nacional". The form contains several input fields and a dropdown menu:

- Título:** A large empty text input field.
- Autor:** A large empty text input field.
- Añadir otro autor:** A small button.
- Identificador de proyecto:** A large empty text input field.
- Nivel de acceso:** A dropdown menu with "Acceso Abierto" selected.
- Fecha de fin de embargo:** A date input field containing "22/07/2018".
- Condiciones de licencia:** A large empty text input field.
- Referencia de identificador alternativo:** A large empty text input field.

Figura 5. Interfaz del usuario del generador de metadatos propuesto.

Fuente: Autor

REPOSITORIO NACIONAL		Documentos	Noticias	Colabora	Contacto	?	🌐	👤
Materias	INGENIERÍA Y TECNOLOGÍA - (CTI) CIENCIAS TECNOLÓGICAS - (CTI) Optimización - ([urn:issn:2007-5448]) Servidor web - ([urn:issn:2007-5448]) Apache - ([urn:issn:2007-5448]) MySql - ([urn:issn:2007-5448])							
Resumen o descripción	Cuando se habla de sistemas informáticos en la Web, una de las configuraciones más utilizadas por los desarrolladores es la combinación de Apache y MySql, independientemente del sistema operativo sobre el cuál se desarrolle. En este trabajo se muestra cómo optimizar los tiempos de procesamiento y despliegue de información en el navegador Web de un Sistema de Recomendación de trabajos de titulación, aplicando sencillas configuraciones en el servidor Apache y siguiendo algunas buenas prácticas de programación al momento de ejecutar llamadas a la base de datos. Las configuraciones realizadas son específicamente en cuanto a la compresión de datos y el manejo de caché. Este ahorro de tiempo repercute directamente en una mejor percepción por parte del usuario con respecto a la usabilidad del sistema.							
Fecha de publicación	31 de mayo de 2016							
Tipo de publicación	Artículo							
Recurso de información	http://redi.uady.mx:8080/handle/123456789/671							
Formato	application/pdf							
Fuente	urn:issn:2007-5448							
Idioma	Español							
Cobertura	Investigación aplicada							
Audiencia	Investigadores							
Repositorio Origen	Repositorio Digital Institucional de la Universidad Autónoma de Yucatán							

Figura 6. Ejemplo del registro de un recurso en el repositorio nacional.

Fuente: CONACYT (2018), Optimización del desempeño de un sistema de recomendación de documentos de texto basado en la configuración de los servidores.

En la Figura 7 se expone el archivo XML generado por la herramienta, el cual, cuenta con los metadatos asociados con el título, los autores, el nivel de acceso, la condición de la licencia, el área de conocimiento, el campo de conocimiento, las palabras clave, la descripción, la fecha de publicación, el tipo de resultado científico, el formato de la publicación, el identificador del recurso, la fuente, el lenguaje, la cobertura y la audiencia.


```

metadata ..
oai_dc:dc ..
  @xmlns:oai_dc: http://www.openarchives.org/OAI/2.0/oai_dc/
  @xmlns:dc: http://www.lyncode.com/soal
  @xmlns:xsi: http://www.w3.org/2001/XMLSchema-instance
  @xmlns:dc: http://purl.org/dc/elements/1.1/
  @xsi:schemaLocation: http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd
  dc:title Optimización del desempeño de un sistema de recomendación de documentos de texto basado en la configuración de los servidores
  dc:creator RODRIGO VENCES NAVA
  dc:creator VICTOR HUGO MENENDEZ DOMINGUEZ
  dc:creator ALFREDO ZAPATA GONZALEZ
  dc:rights info:eu-repo/semantics/openAccess
  dc:rights http://creativecommons.org/licenses/by-nc-nd/4.0
  dc:subject info:eu-repo/classification/cti/7
  dc:subject info:eu-repo/classification/cti/33
  dc:subject Optimización
  dc:subject Servidor web
  dc:subject Apache
  dc:subject MySql
  dc:description Cuando se habla de sistemas informáticos en la Web, una de las configuraciones más utilizadas por los desarrolladores es la combinación de Apache y MySql, independientemente del sistema operativo sobre el cual se desarrolle. En este trabajo se muestra cómo optimizar los tiempos de procesamiento y despliegue de información en el navegador Web de un Sistema de Recomendación de trabajos de titulación, aplicando sencillas configuraciones en el servidor Apache y siguiendo algunas buenas prácticas de programación al momento de ejecutar llamadas a la base de datos. Las configuraciones realizadas son específicamente en cuanto a la compresión de datos y el manejo de caché. Este ahorro de tiempo repercute directamente en una mejor percepción por parte del usuario con respecto a la usabilidad del sistema.
  dc:date 2016-05-31
  dc:type info:eu-repo/semantics/article
  dc:format application/pdf
  dc:identifier http://redi.uady.mx:8080/handle/123456789/671
  dc:source urn:issn:2007-5448
  dc:language spa
  dc:coverage Investigación aplicada
  dc:audience researchers
  
```

Figura 7. Metadatos generados a partir de la interfaz del usuario.

Fuente: Autor

A partir del archivo generado, se utilizó el validador de sintaxis XML de W3Schools (W3Schools, 2020). En la Figura 8 se presenta el resultado, donde una ventana emergente indica que no se identificaron errores.

Una vez que se ha verificado que el archivo generado cumple con la sintaxis XML, se procede a evaluar si REDI-UADY cumple con la capacidad de interoperabilidad a través de sus metadatos. En la Figura 9 se muestra el resultado de dicha validación a través de OAI-PMH Validator (Banos, 2020), donde, a partir de la URL de OAI-PMH de REDI-UADY, se identificó que los registros que utilizan los metadatos OAI-DC cumplen con el esquema XML para OAI-PMH. De esta forma, se garantiza que REDI-UADY permite intercambiar sus metadatos con otros repositorios, considerando, entre sus registros, el que se ha replicado por medio de la herramienta.

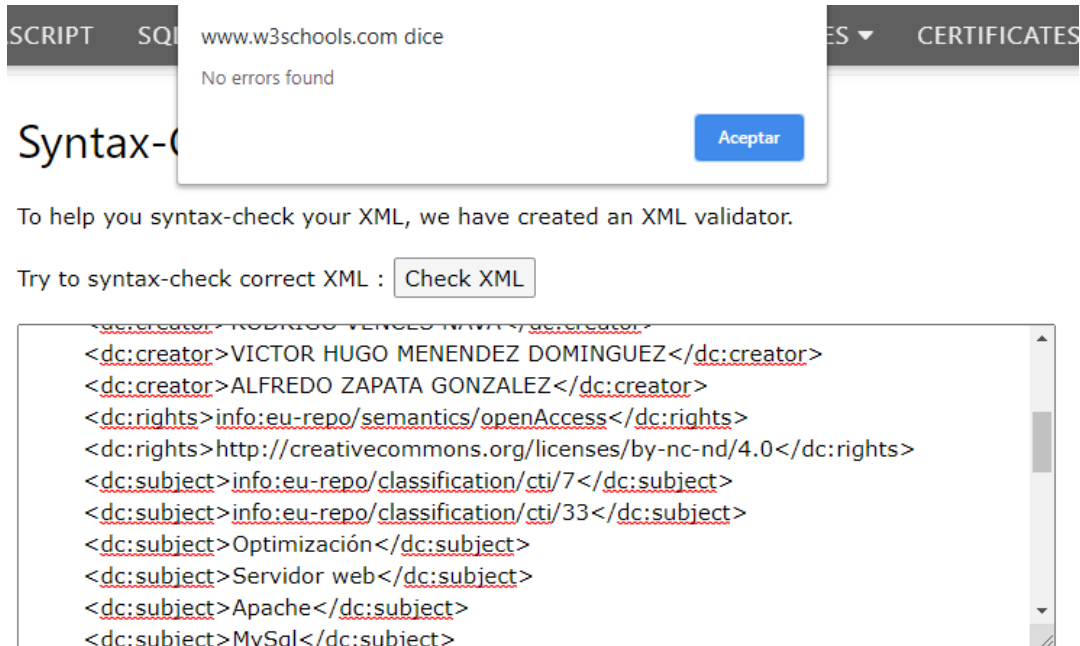


Figura 8. Validación de sintaxis de los metadatos generados.

Fuente: W3 Schools (2020), XML Validator.

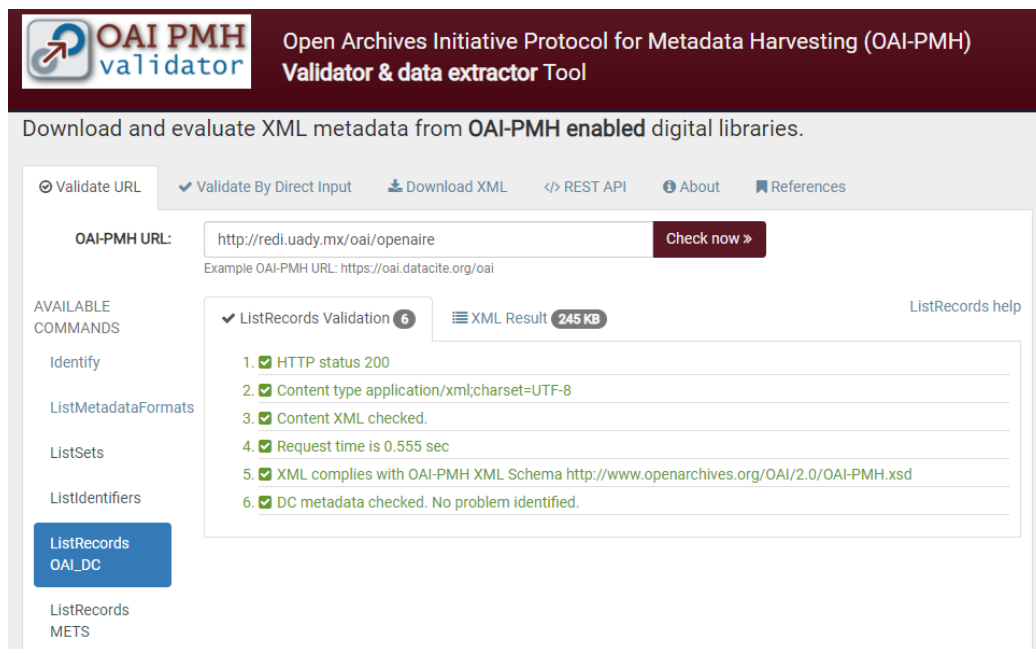


Figura 9. Resultados de la validación de OAI-PMH para REDI-UADY.

Fuente: Banos (2020), Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Validator & data extractor Tool.

Se ha comprobado que REDI-UADY cumple con los requerimientos de OAI-PMH para la interoperabilidad. Considerando que el repositorio nacional utiliza su propia versión de OpenAIRE para la descripción de sus recursos, se evalúa si REDI-UADY cumple con lo requerido a través del validador del repositorio nacional (CONACYT, s/f). En la Figura 10 se muestra el resultado de la validación de REDI-UADY a través del validador del repositorio nacional, donde se observa que REDI-UADY fue validado correctamente.

The screenshot shows the 'METADATA VALIDATOR' interface. At the top, there is a dark header with the title 'METADATA VALIDATOR' and a globe icon. Below the header, a text input field contains the URL 'http://redi.uady.mx/oai/openaire'. Underneath, an example URL is provided: 'Example: OAI-PMH URL: http://repository.parthenonfrieze.gr/frieza-oai/request'. A section titled 'Select an option' has two radio buttons: 'Literature (oai_dc)' (selected) and 'Dataset (oai_datacite)'. A 'Check' button is positioned to the right. A yellow warning box contains the text: 'Metadata validation only involves a validation test: you're not harvesting into Repositorio Nacional'. Below this, a 'Verb:' section shows 'ListRecords XOAI' and two buttons: 'Validation identifier' and 'XML'. The bottom part of the interface displays a list of three successful checks, each with a green checkmark: 1. HTTP Status 200, 2. Application content type xml/charset=UTF-8, and 3. Verify XML Content.

Figura 10. Validación del repositorio nacional para REDI-UADY.

Fuente: CONACYT (s.f), Metadata Validator.

DISCUSIÓN

La herramienta ha resultado una opción interesante en el proceso de construcción de la estructura de metadatos XML compatible con el openAIRE y que cumpla con el criterio de conformidad a las especificaciones establecidas por el Repositorio Nacional de México. Se han presentado los estudios de conformidad que avalan los resultados.

Un aspecto interesante en el desarrollo de la herramienta es su arquitectura. Al estar basado en componentes con funciones específicas posibilita su reutilización de forma sencilla para incorporarlo en otras soluciones de software.

Por otra parte, el uso de plantillas garantiza la vigencia del generador de metadatos. La estructura resultante del proceso puede adaptarse para ser compatible con nuevos perfiles de aplicación o nuevas versiones del estándar openAIRE. Más aún, permite generar una

estructura en otros formatos como CSV, HTML o cualquier otro requerida por un nuevo sistema de información.

CONCLUSIONES

Las herramientas que facilitan la creación de metadatos basados en Dublin Core son apropiadas para el uso general, debido a la naturaleza de la accesibilidad del uso, es decir, no son obligatorios y la repetibilidad es ilimitada.

Sin embargo, cuando se trata del uso de los elementos de dicho estándar para OpenAIRE, el cual tiene un nivel mayor de rigurosidad, si no se siguen los lineamientos establecidos, las herramientas existentes difícilmente ayudarán a permitir la promoción de las buenas prácticas para el registro y recuperación de la información de los repositorios digitales que interoperan a través de OAI-PMH y OpenAIRE.

Los beneficios de la herramienta propuesta para el caso del repositorio nacional del CONACYT, que también presenta sus lineamientos, se basan en la obligatoriedad del llenado de los campos que son elementales para el uso adecuado de OpenAIRE, además del uso de los catálogos validados por el CONACYT y que pueden ser consultados a través de servicios REST, asegurando una mayor consistencia con lo solicitado a los repositorios institucionales para la interoperabilidad de sus recursos.

La ventaja del consumo de los catálogos de CONACYT al momento del llenado de los metadatos es la facilidad de la integración de nuevos elementos, sin la necesidad de actualizar una base de datos que contenga los elementos actuales de cada catálogo.

A futuro, con el propósito de simplificar el desarrollo de sistemas gestores de información basada en Dublin Core u OpenAIRE, se pretende desarrollar una colección de componentes que faciliten la importación y exportación de información en estos estándares de intercambio de información de tal manera que puedan ser incluidas en procesos de generación automática de metadatos.

REFERENCIAS BIBLIOGRÁFICAS

- Banos, V. (2020). Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Validator & data extractor Tool. Recuperado el 1 de septiembre de 2020, de <http://validator.oaipmh.com/>

- Bravo, G. (2005). Dublin Core Metadata Gen: Generator of metadata using Dublin Core.
- Candás Romero, J. (2006). El papel de los metadatos en la preservación digital. *El Profesional de la Información*, 15(2), 126–136.
<https://doi.org/10.3145/epi.2006.mar.05>
- CONACYT. (s/f). Metadata Validator. Recuperado el 1 de septiembre de 2020, de <https://validator.repositorionacionalcti.mx/validator>
- CONACYT. (2014). *Lineamientos Técnicos para el Repositorio Nacional y los Repositorios Institucionales*. Recuperado de <http://www.siicyt.gob.mx/index.php/normatividad/2-conacyt/4-conacyt/1499-lineamientos-tecnicos-para-el-repositorio-nacional-y-los-repositorios-institucionales/file>
- CONACYT. (2017a). *Lineamientos específicos para repositorios*. Recuperado de <http://www.siicyt.gob.mx/index.php/normatividad/2-conacyt/1-programas-vigentes-normatividad/lineamientos/lineamientos-especificos-para-repositorios/4704-lineamientos-especificos-para-repositorios/file>
- CONACYT. (2017b). Lineamientos Jurídicos de Ciencia Abierta. Recuperado el 13 de septiembre de 2020, de <https://www.repositorionacionalcti.mx/documentos>
- CONACYT. (2018). Optimización del desempeño de un sistema de recomendación de documentos de texto basado en la configuración de los servidores. Recuperado el 1 de septiembre de 2020, de <https://www.repositorionacionalcti.mx/recurso/oai:redi.uady.mx:123456789/671>
- CONACYT. (2020). Repositorio nacional/Sección informativa. Recuperado el 30 de marzo de 2020, de <https://www.repositorionacionalcti.mx/>
- DCMI. (2012). DCMI: Dublin Core Metadata Element Set, Version 1.1: Reference Description. Recuperado el 6 de julio de 2020, de <http://dublincore.org/documents/dces/>
- GitHub Inc. (2020). GitHub. Recuperado el 3 de febrero de 2020, de <https://github.com/search?q=&type=>
- Gómez Dueñas, L. (2010). *Modelos de interoperabilidad en bibliotecas digitales y repositorios documentales: caso biblioteca digital colombiana*. Recuperado de http://eprints.rclis.org/14878/1/MODELOS_DE_interoperabilidad_BDCOL.pdf

- Guerrero, J., Sánchez, D., Menéndez, V., Castellanos, M. E., & Gómez, J. (2019). Tools for interoperability between repositories of digital resources. En L. Gómez Chova, A. López Martínez, & I. Candel Torres (Eds.), *Proceedings of INTED2019* (pp. 6292–6300). Valencia: IATED.
- INFOTEC. (2017). *Manual de mejores prácticas*. Recuperado de https://www.repositorionacionalcti.mx/docs/manualesInteroperabilidad/manual_para_mejores_practicas.pdf
- INFOTEC. (2018). *Descripción de servicios REST-Catálogos del Repositorio Nacional*. Recuperado de https://www.repositorionacionalcti.mx/docs/manualesInteroperabilidad/descripcion_de_Servicios_REST.pdf
- Julie, M. (2005). Electronic Records: A Workbook for Archivists. ICA Study 16. International Council on Archives. *Records Management Journal*, 15(3). <https://doi.org/10.1108/rmj.2005.28115cae.003>
- López Guzmán, C. (2005). Interoperabilidad de los Repositorios de Objetos de Aprendizaje a través de IMS. Recuperado el 20 de mayo de 2018, de <http://www.biblioweb.tic.unam.mx/libros/repositorios/interoperabilidad.htm>
- Medina Moreno, L. A., Menéndez Domínguez, V. H., & Palacio Cinco, R. R. (2013). Interoperabilidad entre Repositorios de Objetos de Aprendizaje mediante un Servicio Adaptable basado en Estándares de Acceso Abierto. *Academia Journals de Tuxtla Gutiérrez*. Tuxtla Gutiérrez.
- Méndez Rodríguez, E. M. (2001). *Metadatos y recuperación de información: Estándares, problemas y aplicabilidad en bibliotecas digitales*. Universidad Carlos III de Madrid.
- Open Archives Initiative Organization. (2015). The Open Archives Initiative Protocol for Metadata Harvesting. Recuperado el 30 de julio de 2020, de <https://www.openarchives.org/OAI/openarchivesprotocol.html>
- OpenAIRE. (2017). *OpenAIRE Guidelines Release 3.0*. Recuperado de https://guidelines.readthedocs.io/_/downloads/en/v3.0/pdf/
- Rettberg, N., & Schmidt, B. (2012). OpenAIRE — Building a collaborative open access infrastructure for european researchers. *LIBER Quarterly: The Journal of the Association of European research libraries*, 22(3), 160–175.

<https://doi.org/10.18352/lq.8110>

Steffel, N. (2011). Better, Faster, Stronger: Building a Better Dublin Core Generator. En T. Baker, D. I. Hillmann, & A. Isaac (Eds.), *International Conference on Dublin Core and Metadata Applications*. The Hague: Dublin Core Metadata Initiative. Recuperado de <https://dcevents.dublincore.org/index.php/IntConf/dc-2011/paper/view/5/23>

W3Schools. (2020). XML Validator. Recuperado el 1 de septiembre de 2020, de W3Schools.com website: https://www.w3schools.com/xml/xml_validator.asp