

LA REVOLUCIÓN DE LOS DATOS BIBLIOGRÁFICOS, CIENTÍFICOS Y CULTURALES

ARIEL ALEJANDRO RODRÍGUEZ GARCÍA

Coordinador



Z666.7

R47

La revolución de los datos bibliográficos, científicos y culturales / Coordinador Ariel Alejandro Rodríguez García.
- México : UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información, 2020.

xv, 346 p. – Colección: Metadatos

ISBN: 978-607-30-2996-4

1. Metadatos bibliográficos. 2. Datos vinculados. 3. Big data 4. Repositorios institucionales. 5. BIBFRAME (Modelo conceptual). I. Rodríguez García, Ariel Alejandro, coordinador. II. ser.

Diseño de cubierta: Oscar Daniel López Marín

Primera edición, 2020 D.R. © UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

Ciudad Universitaria, 04510, Ciudad de México

Impreso y hecho en México

ISBN: 978-607-30-2996-4

Publicación dictaminada

Contenido

INTRODUCCIÓN	ix
1. DATOS BIBLIOGRÁFICOS	1
I. Retro y prospectiva de la investigación bibliotecológica sobre metadatos	3
ARIEL ALEJANDRO RODRÍGUEZ GARCÍA	
II. <i>BIBFRAME</i> : Un modelo de metadatos para la web semántica	17
FILIBERTO FELIPE MARTÍNEZ ARELLANO	
III. Metadatos, datos enlazados e interoperabilidad: reflexiones en el dominio bibliográfico	33
FABIANO FERREIRA DE CASTRO	
IV. De los metadatos para la organización de la Información a la tecnología <i>middleware</i> para los servicios de las bibliotecas: la biblioteca inteligente	63
GERARDO BELMONT LUNA	
V. Enlazando datos: un modelo conceptual orientado a objetos para el diseño de bases bibliográficas	79
LIZBETH BERENICE HERRERA DELGADO	
VI. Informe desde las trincheras: Transformando unidades de catalogación para incluir servicios de repositorio	113
LISA FURUBOTTEN	

2. DATOS CIENTÍFICOS141

I. Factores determinantes para la implementación
del esquema de metadatos para repositorios de datos
de investigación de la Política de Ciencia Abierta en México143

MIGUEL ADOLFO GUAJARDO MENDOZA

II. Habilidades requeridas por el profesional de la Industria 4.0.
Aparición de un nuevo perfil para el área de producción,
en la cuarta Revolución Industrial.161

MARCO BRANDÃO

III. Implementación de un repositorio unificado para
conformar el Sistema de Inteligencia Institucional del INE183

ALEJANDRO VERGARA TORRES

IV. Calidad en los metadatos:
reto para asegurar la participación en repositorios
nacionales manteniendo los estándares institucionales205

ALMA BEATRIZ RIVERA AGUILERA, ELISA CRUZ ROJAS,

BENJAMÍN ALEJANDRO GUERRERO OLVERA Y ÓSCAR ALFONZO PEREYRA

V. Interoperabilidad en plataformas
de publicación y distribución de libros digitales.....223

JENNY TERESITA GUERRA GONZÁLEZ

3. DATOS CULTURALES	245
I. Metadatos para preservación digital	247
JUAN VOUTSSÁS MÁRQUEZ	
II. El Sistema de Información para el Registro Universitario de Espacios y Activos Culturales (SI-RUEyAC)	265
CATALINA NAUMIS PEÑA, NATALIA VELAZCO PLACENCIA Y ARIEL ALEJANDRO RODRÍGUEZ GARCÍA	
III. La ciudadanía digital y sus mediciones: el caso del estado de ánimo de los tuiteros en México	283
HÉCTOR ALEJANDRO RAMOS CHÁVEZ	
IV. El papel de los metadatos en la construcción de datos enlazados para bibliotecas	301
EDER ÁVILA BARRIENTOS	
V. La integración del usuario final en la creación de metadatos	325
PATRICIA HERNÁNDEZ SALAZAR	

Calidad en los metadatos: reto para asegurar la participación en repositorios nacionales manteniendo los estándares institucionales

ALMA BEATRIZ RIVERA-AGUILERA

ELISA CRUZ ROJAS

BENJAMÍN A. GUERRERO OLVERA

ÓSCAR ALFONZO PEREYRA

Universidad Iberoamericana Ciudad de México

INTRODUCCIÓN

La consolidación de un Repositorio Institucional (RI) involucra factores como gestión técnica y financiera, renovación de plataformas tecnológicas, nuevas estrategias de poblamiento; normalización y adecuación de metadatos a los lineamientos del Repositorio Nacional (RN) para asegurar la cosecha, consideración de diversos estándares, mantenimiento de la calidad de los registros del RI, respeto a los derechos de autor y mejora en estrategias de preservación. Compartir experiencias es de mucho valor cuando se abordan proyectos innovadores como suelen ser los relacionados con RI y, en especial, cuando los metadatos de los RI serán cosechados por un repositorio más amplio.

Llevar a cabo una reflexión sistemática de lo que ha sucedido en un proyecto suele quedarse en una buena intención, pues el proyecto mismo y otras actividades consumen

el tiempo para un escrito colectivo, como el que aquí se comparte. En este texto, revisaremos los antecedentes de la organización y publicación de colecciones digitales propias en nuestra biblioteca. Abordaremos el proyecto que posibilitó la experiencia descrita, el poblamiento del RI a ser consolidado, los lineamientos en relación a los metadatos en RI para ser cosechados por el RN, la normalización que algunas etiquetas han requerido y los resultados de las pruebas de cosecha.

ANTECEDENTES

En 1998 se iniciaron en la Biblioteca Francisco Xavier Clavigero los esfuerzos para la creación de una biblioteca digital, tarea que continuaba la tradición de uso de tecnología para ofrecer mejores servicios de información que arrancó en la biblioteca en los años setenta, tal como señala en detalle Rivera-Aguilera (2010).

En el año 2004, se inauguró la primera colección digital en texto completo, que de forma sistemática e integral (metadatos y texto completo) explotaba la información digital. Esta colección fue la de tesis digitales, la cual estuvo basada en estándares de metadatos descriptivos para tesis ETDMS (Electronic Thesis and Dissertation Metadata Standard) y DC (Dublin Core) y utilizó una combinación de software comercial (Aleph) y software libre (Harvest). La colección de tesis digitales de la Ibero se conformó originalmente en un sitio en HTML que ligaba en una estructura jerárquica organizada por departamentos académicos los archivos PDF de las tesis almacenadas en un servidor Linux. Los trabajos

de titulación podían recuperarse, también, a través del catálogo en línea en Aleph 500 vía la etiqueta 856 del formato MARC. Una vez puesta al servicio de los usuarios, la aplicación fue evaluada, desde la perspectiva de usabilidad, y encontrada satisfactoria por éstos. La explotación de los textos completos de los archivos PDF se hacía hasta el 2007 a través del buscador Harvest.

En 2008, se consideró ampliar —no sustituir— el concepto de biblioteca digital al de repositorio institucional. En el caso de la BFXC de la Ibero México, la concepción incluye desde entonces, además de las colecciones bibliográficas, la producción docente de investigación, documentos institucionales y de vinculación social en un solo interfaz de búsqueda. El administrador de colecciones digitales, seleccionado para implementar esta nueva etapa, fue Greenstone, el cual de forma integral almacena y explota los metadatos, y el texto completo de los objetos digitales textuales. Greenstone también administra con eficiencia las imágenes digitales y genera estructuras jerárquicas de acceso vía HTML dinámico. Este modelo implicaba, para las tesis digitales, una transferencia de metadatos desde MARC en Aleph a DC en Greenstone. Para las colecciones de comunicaciones oficiales, patrimonio natural y cultural (de uso local); producción científica (en pruebas) y material didáctico (en pruebas), una descripción basada en DC, Learning Object Metadata (LOM) y algunas etiquetas locales. Nuestra colección de tesis en Greenstone ha sido cosechada por REMERI. Sin embargo, dada la poca dinámica en nuevas versiones en Greenstone, la falta de integración de interfaces de autoguardado —adecuados para usuario final— y la escasez de usuarios de Greenstone en México, se consideró que el paso a DSpace fortalecía a nuestra biblioteca digital/RI.

EL PROYECTO

El Consejo Nacional de Ciencia y Tecnología (CONACYT) ha abierto, al menos en cuatro ocasiones (2014-2017), convocatorias para financiar proyectos de inicio o consolidación de RI. Obtener fondos en convocatorias evaluadas ayuda a las unidades de información a adquirir recursos que en la operación normal no estarían disponibles y a poner a la biblioteca, como un actor central en la innovación tecnológica institucional. Otro beneficio de participar es promover las políticas de acceso abierto en las universidades y posicionar a los bibliotecarios como pares académicos. Por ello, se decidió participar en la convocatoria 2016 y se obtuvo financiamiento para nuestro proyecto de consolidación en 2017.

La experiencia que se comparte en este capítulo se enmarca en el desarrollo del proyecto antes mencionado. Éste implicó una serie de tareas simultáneas en ciertas medida y necesariamente coordinadas:

1. Consolidar nuestro repositorio de tesis existente, diseñar y ejecutar procedimientos de acopio de publicaciones académicas de acceso abierto, generadas por comunidad universitaria desde diferentes plataformas abiertas y sus correspondientes metadatos.
2. Definir una estrategia de verificación de contenidos en acceso abierto y de obtención de permisos por parte de los autores.
3. Transferir archivos y metadatos disponibles, y adecuarlos a los lineamientos del repositorio nacional.

4. Llevar a cabo un ejercicio crítico y propositivo en relación con la normalización planteada por los lineamientos del repositorio nacional.
5. Obtener recursos de *hardware*, instalar adecuadamente nueva versión de *software* gestor de colecciones digitales, transferir metadatos y contenidos, e integrar una nueva forma de almacenamiento de parte de los recursos en formatos que aseguren la preservación de textos a largo plazo (XML).
6. Gestionar el equipo interno y la vinculación grupal con consultores o proveedores para lograr la concreción de un proyecto en un plazo de tiempo muy limitado.

En este texto, nos vamos a enfocar en el rol de los metadatos en un proyecto que pretende compartirlos desde un repositorio institucional a uno nacional y los enormes retos que implica adecuarse a unos lineamientos nacionales manteniendo los estándares calidad institucionales.

LINEAMIENTOS PARA COSECHAR METADATOS DE RI AL RN

El RN, administrado por Conacyt, brinda la normalización y los lineamientos para el ingreso de los metadatos que serán cosechados de los RI en su manual de Lineamientos Específicos para Repositorios (LER), se encuentran basados en estándares internacionales, tales como el protocolo OAI-PMH para la parte de interoperabilidad entre repositorios y los campos de OpenAIRE para la definición de las etiquetas y contenidos de los metadatos a ingresar en el RN.

Los lineamientos proponen veintisiete campos de metadatos DC: son nueve campos de carácter obligatorio, nueve campos obligatorios cuando aplique y otros nueve campos solo como recomendados.

Estos lineamientos brindados por CONACYT, en conjunto con la sintaxis necesaria para cada elemento, generan un “manual” sencillo y básico para la construcción del repositorio. El RN también ofrece catálogos de lenguajes o términos controlados a consultar para los campos que lo requieran para una mayor normalización de los repositorios.

Sin embargo, ya en la práctica del asentamiento de metadatos para la conformación del repositorio, nos encontramos con ciertas dificultades; por ejemplo, el interface rústico del catálogo de autores. Este catálogo, si bien es una gran herramienta para la normalización, no ofrece una búsqueda de autores de manera ágil ni fácil, ya que es necesario probar con variaciones la búsqueda de un nombre, pues al asentar un creador existen diferentes maneras en las que se separan los apellidos. En algunos casos, se utilizan guiones bajos para separar el primer del segundo apellido, lo que genera caracteres con los que se deben de realizar múltiples búsquedas para obtener el nombre que se esté buscando y, en algunos casos, los registros de los autores no cuentan con los indicadores que el mismo Conacyt pide. Bajo esta misma línea, no existe una normalización en el uso de mayúsculas y minúsculas, ya que existen casos de los tres tipos: únicamente con los usos de letras mayúsculas, con el uso de solamente letras minúsculas y el uso de ambas.

El otro problema más complejo es la obtención de los identificadores de los creadores que ya no tienen relación con la institución. Si bien integrar los identificadores parece una manera eficiente de ligar la información y ser

un principio de un control de autoridades de personas en el ámbito de ciencia y tecnología, todavía se muestra un esfuerzo muy precario y poco coordinado con esfuerzos internacionales como NACO o VIAF. En la práctica, obtener los identificadores Clave Única de Registro de Población en México (CURP), Número de Curriculum Vitae Único asignado por Conacyt (CVU), *Open Researcher and Contributor* ID (ORCID) y Documento Nacional de Identidad de cualquier país (DNI) de los autores de productos de investigación particulares, se vuelve complejo. Esto debido a que los materiales con los que se va a colaborar dentro del repositorio nacional son tesis, pues la mayoría pertenecen a autores que son estudiantes sin un registro para obtener un CVU y, si bien se puede asentar utilizando el CURP, se fuerza a las instituciones a revisar sus compromisos en materia de privacidad. Otro reto es la múltiple autoría y, más específico, la autoría en conjunto con investigadores de otras nacionalidades en las que en algunos casos no se cuenta con información en primera mano sobre algún número identificador para su registro.

Por cuestión de los tiempos en los que se deben entregar resultados del repositorio —tomando en cuenta la experiencia previa con la que ya se contaba en la Biblioteca Francisco Xavier Clavigero (BFXC) en torno a la administración y asignación de metadatos, posterior a un mapeo—, se tomó la decisión de reducir los campos Dublin Core (DC) a doce elementos, los cuales fueron elegidos por su importancia dentro de la descripción y por ser los campos con los que se facilitaba la cosecha de los datos, y considerando la experiencia y las recomendaciones de los consultores externos que brindaron su apoyo.

POBLAMIENTO DEL REPOSITORIO CONSOLIDADO

Tesis: la transferencia de metadatos desde Aleph.

La transferencia de registros bibliográficos de tesis se presentó en dos fases: de nuestro catálogo bibliográfico hacia el RI y del RI al RN. Se preparó el mapeo de los campos MARC21 a DC para el paso al RI y se adecuó paso de los metadatos DC a LER en los registros de transferencia basados en el protocolo OAI para ser cosechados en el RN a fin de lograr la interoperabilidad entre diversos repositorios (Roy y Chandra 2017,74). De esta manera, los contenidos de las tesis tienen mayor visibilidad a nivel nacional e internacional al participar en el repositorio nacional CONACYT. La preparación de los campos MARC21 utilizados en la descripción de las tesis digitales de posgrado se llevó a cabo de manera detallada por medio de un mapeo con una correlación exacta de los contenidos requeridos en los Lineamientos Específicos para Repositorios (LER), para lo cual se realizaron varias versiones y se fueron afinando los detalles. De manera simultánea, se llevó a cabo un barrido de control de calidad de los registros de tesis en ALEPH con la intención de reducir el margen de error para lograr una transferencia lo más limpia posible.

Cabe mencionar la importancia de los metadatos integrados y fusionados de diversos esquemas adaptados a las necesidades para obtener un perfil de aplicación específico en este proyecto. Este perfil se ve reflejado en la transferencia de las tesis a la versión DSpace 6.0 y, así también, en la cosecha hacia el repositorio CONACYT. Recordemos que los metadatos apoyan la construcción de servicios coherentes para los usuarios a partir de componentes que son técnicamente distintas y administrados por diferentes organizaciones (Potvin 2016, 100).

El proyecto permitió examinar las divergencias que existen entre lo dicho por los lineamientos y la realidad. Nuestro ejercicio demostró la importancia de desarrollar conjuntos de metadatos eficientes en la conformación de repositorios colectivos con la finalidad de mantener la interoperabilidad, autenticidad y confiabilidad. Tal como lo señala la literatura especializada, los conjuntos de esquemas o perfiles de aplicación deben de ser capaces de aplicarse a diversas instituciones con la finalidad de obtener recolecciones y cosechas a través de protocolos de acceso

abierto. Los metadatos enfatizan la base para la recuperación eficiente entre recursos, así como, dando la debida importancia a descriptivos, técnicos, administrativos y de preservación a través del tiempo y las tecnologías (Potvin 2016, 101).

Para lograr el objetivo de la transferencia en dos niveles, es decir, a DSpace RI y Repositorio CONACYT, se desarrolló un archivo de Excel con los campos requeridos para ambos casos. Aproximadamente en novecientos registros de tesis se enriquecieron los campos y se agregaron los requeridos por LER. Algunos de estos campos fueron los autores y directores con los identificadores requeridos por Conacyt y el área de conocimiento con su identificador, tipo de documento, es decir, “dc.contributor.author”, “dc.creator”, “dc.contributor.director”, “dc.type”, entre otros.

Artículos, ponencias, capítulos de libros
y libros: identificación de metadatos en Internet.

La recuperación de documentos requirió una búsqueda en las páginas especializadas en investigación científica. La plataforma que nos dio la mayor cantidad de documentos subidos por los propios investigadores fue Researchgate.net,

en la que encontramos veintisiete perfiles de investigadores de la universidad con bastante producción; las cuentas pertenecían tanto a profesores miembros del Sistema Nacional de Investigadores, como a profesores de tiempo. También buscamos documentos en Academia.edu, en los que dimos con cuarenta y siete perfiles de investigadores, pero muchos de ellos sin producción reportada en esa plataforma. Es importante recalcar que al momento de descargar los archivos, nos encontramos con varios errores y no había una forma sencilla de recuperar los metadatos. Más tarde, observamos que este error provenía del tamaño de los archivos subidos a Researchgate.

El siguiente reto fue observar los derechos de autor de muchos de los documentos descargados. Diversas casas editoriales tienen reglas para la distribución de los artículos publicados por ellos. Para entender las reglas de embargo y los tiempos en los cuales es permitido distribuir un artículo, recurrimos a SHERPA/RoMEO, página en la que se pueden consultar las reglas de *copyright* según la editorial. Esto nos llevó a tener que eliminar hasta un 15 por ciento de los documentos encontrados. Para poder compensar esta pérdida, se espera enviar emails a los investigadores invitándolos a enviar una versión *preprint* de la cual tiene permiso de distribución.

Durante la recuperación de archivos en Google Académico, también tuvimos algunos problemas en la descarga. En ocasiones, un mismo artículo se encontraba en diversas páginas y el nombre podía llegar a ser ligeramente diferente o el autor tenía un nombre homónimo que la del investigador de la universidad. Gracias a la página de Google, nos dimos cuenta de que había más investigadores en la plataforma de Researchgate; el problema al momento de buscarlos fue que sus nombres o apellidos no estaban escritos por completo en el perfil del investigador.

Una vez que localizamos más investigadores en las diversas plataformas con un total de mil recursos de ciencia y tecnología que era la meta, nos dimos a la tarea de almacenar los datos en una tabla de Excel que sirviera como base para identificar la cantidad de documentos encontrados, en qué plataforma y qué tipo de material se trataba. A este archivo de control, se le fueron agregando datos que nos permitieron llevar un seguimiento del avance del proyecto. El archivo también sirve de referencia para saber qué profesores cuentan con material de investigación, la cantidad y la fuente de estos documentos.

Otro aspecto importante a reportar fue la eliminación de documentos que no contaban con la suficiente calidad para ser subidos a los repositorios. Esto debido a que los documentos estaban mal escaneados, era solamente la portada del libro en donde presentaron un artículo o simplemente no contaban con los suficientes metadatos. En varias ocasiones, se requirió una nueva búsqueda del mismo documento a fin de encontrar metadatos faltantes o rectificar si el documento era del autor. Cada área disciplinar supuso un reto debido a que en las ciencias naturales es común participar con una gran cantidad de autores, mientras que, por otro lado, en ciencias sociales se puede contener tanto una como varias autorías.

Una vez recuperados los documentos, el siguiente objetivo fue identificar la mejor forma de abordar la captura de metadatos. En ánimos de acelerar el proceso, probamos extraer los datos de forma automática mediante el uso del software Zotero. Gracias a la aplicación, logramos extraer los datos de varios documentos descargados desde las páginas académicas que nos permitieron la libre obtención de los datos. El problema que nos causó este método fue que nos arrojaba demasiados datos que no eran necesariamente solicitados

por los lineamientos del CONACYT, además de generar una búsqueda extra de cada archivo a fin de conseguir los metadatos de la página. Lo anterior nos llevó a descartar este método y, en su lugar, optamos por capturar los datos de forma manual.

Mediante el listado de control en Excel, logramos ubicar a los investigadores con la mayor cantidad de documentos recopilados, por lo que se empezó la captura de los metadatos de acuerdo con este orden. Tuvimos la oportunidad de revisar qué profesores aún se encontraban laborando en la universidad y quiénes se acaban de integrar al Sistema Nacional de Investigadores (SNI). También integramos los correos electrónicos de éstos y otros datos que nos ayudaron a generar los archivos con los metadatos.

Durante el trabajo de la captura de metadatos, generamos dos archivos de Excel. En uno de estos archivos, se introdujeron todos los datos de importancia sobre el investigador y el archivo. Al mismo tiempo de generar metadatos para este primer archivo, revisamos por segunda vez los documentos para verificar si contaba con alguna restricción y si el archivo presentaba la calidad suficiente para estar en el repositorio. También, se realizó una modificación a los nombres de archivos recuperados. Se utilizaron las iniciales del investigador, más un identificador que permitiera saber si el archivo era un artículo, libro, conferencia o datos de investigación. Para hacer la modificación de archivos en grupo, utilizamos el *software* de Advanced Renamer, gracias al cual pudimos acelerar los cambios de nombre.

Gracias a todos estos procesos generados durante el proyecto, automatizamos la futura inserción de nuevos archivos en el repositorio que encontramos interesantes sobre el estado de investigación de la universidad, las áreas que producen más y cuáles son los investigadores más prolíficos.

ADECUACIÓN A LOS LINEAMIENTOS LER PARA EL CASO DE LOS AUTORES DE TESIS

Como ya se comentó en el apartado de lineamientos para cosechar metadatos, el principal reto fue la normalización de los nombres de autores, directores de tesis y su localización dentro del catálogo CONACYT. El proceso de búsqueda en dicho catálogo tomó cierto tiempo en la curva de aprendizaje. Nuestra base de tesis, por ejemplo, era de novecientos registros, entonces eran aproximadamente mil ochocientos o más nombres que normalizar y agregar identificadores en un lapso de tiempo de dos meses.

Así pues, para adecuar los nombres e identificadores en el archivo Excel conformado a partir de los registros bibliográficos de tesis en Aleph, se llevaron varias acciones. En una primera etapa se solicitó el apoyo de ocho catalogadores para la investigación de autores e identificadores y el enriquecimiento del archivo de Excel. Esto fue posible revisando y transcribiendo información de manera simultánea de los registros bibliográficos en ALEPH y catálogo CONACYT hacia el archivo Excel. Para esto, se elaboraron procedimientos, flujos de trabajo, controles de calidad; entre otros, para optimizar el tiempo y las actividades de los catalogadores.

En una segunda etapa del proceso, se identificaron los autores no localizados en el catálogo de autores CONACYT y se solicitó que se dieran de alta. Siguiendo los lineamientos establecidos, se llenaron otros archivos de Excel y se subieron a través de la interfaz del portal de Conacyt de alta de autores. Posteriormente, se fue completando el archivo de Excel con los campos LER requeridos.

Hay que destacar la importancia de implementar un estándar común nacional, en este caso, aplicado a las tesis

electrónicas, lo cual ha favorecido la visibilidad y cooperación entre diversas instituciones hacia la conformación de un repositorio nacional. Por otro lado, ha sido una oportunidad de reevaluar los desacuerdos entre lo ideal y la realidad en la aplicación de los esquemas de metadatos en la descripción, preservación, herramientas y transferencias de las tesis electrónicas; asimismo, su impacto en los sistemas, implementación de mejores servicios y difusión de los contenidos en acceso abierto.

RESULTADOS PRUEBAS DE COSECHA

Se llevaron a cabo siete pruebas de cosecha de metadatos, en su mayoría tesis y artículos, retroalimentadas por el área informática y bibliotecológica del RN. Los resultados de las pruebas iniciales reportaron dos tipos de problemas:

1. Autores
 - a. Falta de un identificador válido para los autores en al menos uno de ellos. Este tema se ha desarrollado ampliamente en los apartados referidos a los LER. Si bien conocíamos de este requerimiento, al momento de las pruebas de cosecha nos dimos cuenta de que para el caso de las tesis no sólo se refería a que al menos un autor tuviese CVU, CURP, etcétera, sino que todos los autores o colaboradores de un recurso (por ejemplo, para tesis dc.contributor con datos de director o sinodales) que no tengan un identificador único, deberían tener un identificador de nombre llamado RN que se considera un Registro de Autor CONACYT, para lo cual hay que dar de alta a todos los *contri-*

butor. Nótese que en el caso de un registro que solamente tenga RN, este será rechazado si no cuenta con algún dc.creator o dc.contributor que posea identificador del tipo CVU, CURP, ORCID o DNI. Con respecto a este punto, cabe mencionar que la política de CONACYT sobre identificadores únicos tuvo algunas modificaciones que facilitaron la cosecha durante el transcurso del proyecto; sin embargo, hay una oportunidad de mejora para los facilitadores del RN en la redacción de las políticas y ejemplificación de la captura de metadatos en relación a autores, y en general en relación a todos los metadatos LER.

- b. Inconsistencias en el catálogo de autores CONACYT. Por ejemplo, dos registros de un mismo autor, asentados de forma distinta y que a pesar de tener el mismo CVU eran rechazados.

2. El tipo de material para los estudios de caso debió ser ajustado, pues el catálogo de captura indicaba que se podía usar **dc.type** con el valor **masterDegreeWork**, pero nos rechazaba esos trabajos de titulación. El reporte de la cosecha nos indicó que necesitamos una repetición de este metadato previa con el valor **other**. Este mismo metadato seguía dando problema hasta que se detectó que las listas de valores válidos se albergaban en repositorios semánticos diferentes.

`<dc:type>info:eu-repo/semantics/other</dc:type>`

`<dc:type>info:mx-repo/semantics/masterDegreeWork</dc:type>`

En la séptima prueba, finalmente, se indicó que se podía solicitar la cosecha oficial.

CONCLUSIONES Y PROSPECTIVA

La experiencia del proyecto aquí reportado nos ha dejado muchas enseñanzas y retos a futuro, pues los nueve meses de duraición nos dejan la satisfacción de ver el repositorio consolidado con una nueva plataforma e interoperando con otros repositorios como el RN y el repositorio AUSJAL (<http://repositorio.ausjal.org/>). Sin embargo, quedan todavía muchas tareas sobre todo las relacionadas con afinar las políticas de uso del Dublin Core y perfiles de OAI, enriquecer los metadatos y asegurar el uso correcto de las etiquetas para mantenernos en los estándares y al mismo tiempo responder a los requerimientos de algunos cosechadores como el RN.

BIBLIOGRAFÍA

- Potvin, Sarah y Santi Thompson. "An Analysis of Evolving Metadata Influences, Standards, and Practices in Electronic Theses and Dissertations", *Library Resources & Technical Services* vol. 60, núm. 2: 99–114. Disponible el 29 de julio de 2018 en <https://journals.ala.org/index.php/lrts/article/view/5963/7585>.
- Rivera-Aguilera, Alma Beatriz. 2010. "Perfiles de aplicación en Greenstone: integración de Diversos Esquemas de Metadatos Para La Catalogación de Las Colecciones Digitales de La Biblioteca Francisco Xavier Clavigero." En *IV Encuentro de Catalogación y Metadatos. Memoria 4 Al 6 de noviembre de 2009*, 215–32. México: UNAM/Centro de Investigaciones Bibliotecológicas. Disponible en <http://132.248.242.6/~publica/conmutarl.php?arch=3&idx=247>.
- Rivera Aguilera *et al.* 2011. "Propuesta de Un Perfil de Aplicación Para Material Didáctico Digital Basado En El Marco de Singapur." En *VI Encuentro de Catalogación y Metadatos*, 100–113. México: UNAM/Centro de Investigaciones Bibliotecológicas.
- Rivera Aguilera, Alma Beatriz y José Juan Téllez Bertadillo. 2011a. "Elementos Contextuales, Técnicos y de Diseño Que Posibilitan El Acceso Abierto a Colecciones Digitales Universitarias." En *Coloquio Acceso Abierto a La Información en las Bibliotecas Académicas de América Latina y El Caribe*, Mesa Redonda Acceso En Las Instituciones Latinoamericanas, 309–334. México: UNAM/Centro de Investigaciones Bibliotecológicas.

- Rivera-Aguilera, Alma Beatriz *et al.* 2011b. “Conjunto Descriptivo de Un Perfil de Aplicación En Lo Referente a Creadores de Material Didáctico Digital En Ambientes de Biblioteca 2.0.” En *5o Encuentro de Catalogación y Metadatos*, 321–47. México: UNAM/Centro de Investigaciones Bibliotecológicas. <http://132.248.242.6/~publica/conmutarl.php?arch=1&idx=258>
- Roy, Bijan Kumar; Subal Chandra Biswas y Parthasarathi Mukhopadhyay. 2017. “Designing Metadata Harvesting Framework for OAI-Based LIS Repositories: A Prototype”, *International Journal of Information Science and Management* vol. 15, núm. 1: 73–88. <https://ijism.ricest.ac.ir/index.php/ijism/article/view/938/294>

La Revolución de los Datos Bibliográficos, Científicos y Culturales. La edición consta de 100 ejemplares. Coordinación editorial, Israel Chávez Reséndiz; revisión especializada, Valeria Guzmán González, revisión de pruebas, Carlos Ceballos Sosa, formación editorial, Oscar Daniel López Marín. Instituto de Investigaciones Bibliotecológicas y de la Información /UNAM. Fue impreso en papel cultural de 90 gr. en los talleres de Grupo Fogra. Año de Juárez 223. Col. Granjas San Antonio. Alcaldía Iztapalapa. Ciudad de México. Se terminó de imprimir en julio de 2020.