

Situación Actual de los Repositorios Abiertos en México

Ana María González Mendoza
 Centro Universitario UAEM Texcoco
 Av. Jardín Zumpango s/n, Fraccionamiento El Tejocote,
 Texcoco, Estado de México, México
 merlinalove@hotmail.com

Irene Aguilar Juarez
 Centro Universitario UAEM Texcoco
 Av. Jardín Zumpango s/n, Fraccionamiento El Tejocote,
 Texcoco, Estado de México, México
 ireneico@gmail.com

Resumen Después de dos décadas de publicación masiva de contenidos digitales en la Web, ahora la comunidad científica se enfrenta al reto de aprovechar de forma fácil y eficiente toda la información valiosa disponible. Para ello entre otras estrategias se ha centralizado la producción académica en repositorios abiertos, los cuales representan la fuente de información de mejor calidad para el desarrollo de la educación y de la investigación científica pues en ellos se almacena todo tipo de documentos académicos sobre cualquier tema del conocimiento humano. En este trabajo se describe el software de licencia libre que permite la implementación de estos repositorios, haciendo un análisis comparativo de las herramientas de desarrollo disponibles, posteriormente se describe la situación actual de los repositorios abiertos en el mundo y finalmente se describe la situación de estos repositorios en México

Keywords— *EPrints, DSpace, Fedora, Greenstone, Omeka, OAI-PMH, Repositorios Mexicanos.*

I. INTRODUCCION

El movimiento de Acceso Abierto se basa en el convencimiento de que la libre circulación de las ideas no sólo es un derecho humano, sino que se trata de un instrumento de mejora de la propia ciencia. Surge de la necesidad imperiosa de hacer accesibles los conocimientos a todos los que puedan utilizarlos, aplicarlos o aumentarlos, su finalidad es hacer un mejor uso de los recursos, minimizar la duplicación de esfuerzos, encontrar fuentes que permitan fortalecer el discurso de la investigación. Se originó formalmente en el año 2001 y actualmente múltiples organizaciones reconocen y favorecen este movimiento, entre ellos el European Research Council y el European Research Advisory Board, los consejos de investigación científica del Reino Unido y el Irish Research Council.

Los repositorios de acceso abierto son portales que permiten el acceso a publicaciones científicas o técnicas, de forma gratuita, y de esta forma ser leídos, descargados, copiados, distribuidos, impresos, buscados o enlazados y utilizados con propósitos legítimos ligados a la investigación científica, a la educación o a la gestión de políticas públicas, sin otras barreras económicas, legales o técnicas que las que suponga Internet.

En los últimos 15 años han surgido una gran variedad de herramientas para la implementación de bibliotecas digitales y repositorios. Si bien esto trae consigo un amplio abanico de posibilidades para buscar la que mejor se adapte a nuestras necesidades; puede resultar una tarea ardua el seleccionar una de ellas.

En este trabajo se han analizado 5 de estas herramientas: Eprints, DSpace, Fedora, Greenstone y Omeka; todas ellas con dos características en común: Ser de software libre y compatible con el protocolo de interoperabilidad OAI (Open Access Initiative).

En la primera parte de este documento, hablamos de estas herramientas y sus características. Como resultado se pueden apreciar sus coincidencias y particularidades.

Posteriormente se muestra un análisis de repositorios desde una perspectiva global, conoceremos cual es la herramienta líder a nivel mundial y cuál es el país que cuenta con más repositorios; para terminar con una aproximación a Latinoamérica y finalmente la situación en la que se encuentra México. Cerramos este documento presentando las conclusiones.

II. HERRAMIENTAS DE SW LIBRE PARA LA CREACION DE REPOSITARIOS

Anteriormente otros autores como George Pirounakis,

A. EPrints

El software EPrints se convirtió en el primero y uno de los más utilizados softwares de acceso libre y gratuito para la implementación de repositorios institucionales y ha inspirado desde su desarrollo a otros software que cumplen propósitos similares.

EPrints es utilizado para la creación de repositorios y es compatible con el protocolo OAI (Open Archives Initiative) para la recolección de metadatos. Comparte muchas de las características comúnmente observadas en los sistemas de gestión de documentos, pero ha sido utilizado, principalmente, para la implementación de repositorios institucionales y revistas científicas. EPrints fue creado en la University of Southampton School of Electronics and Computer Science y liberado bajo una licencia GPL, General Public License por sus

siglas en inglés. Su primera versión es lanzada en el año 2000 como resultado de una reunión en el año 1999 en Santa Fe que puso en marcha lo que eventualmente se convirtió en el protocolo OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting).

EPrints es una aplicación web con arquitectura LAMP (donde la inicial P corresponde a Perl) y es multiplataforma (Linux, Solaris, Mac OS X y Windows).

La configuración de un repositorio EPrints implica la modificación de los archivos de configuración escritos en Perl o XML. La publicación del repositorio es en plantillas HTML, hojas de estilo CSS e imágenes en línea. Es una herramienta multilingüe; ha sido traducido a otros idiomas a través de archivos frase XML específicos del idioma, incluyen búlgaro, francés, alemán, húngaro, italiano, japonés, ruso, español y ucraniano.

Existen más de 300 repositorios de acceso abiertos registrados que han sido implementados con EPrints alrededor del mundo. [1]

B. DSpace

DSpace es un software de código abierto que provee herramientas para la administración de colecciones digitales, el uso más común es en la creación de repositorios institucionales.

La primera versión de DSpace la 1.0 fue liberada el cuatro de noviembre de 2002, como resultado de un esfuerzo conjunto por parte del Instituto Tecnológico de Massachusetts MIT y los laboratorios Hewlett-Packard, bajo una licencia GPL que permite a los usuarios personalizar o extender el software según se necesite. El uso de esta licencia no tiene ningún costo y en caso de hacer una mejora o modificación del código se sugiere compartirla a la comunidad DSpace como forma de retribución; los *committers* de esta comunidad son desarrolladores voluntarios seleccionados en un grupo específico y se encargaran de depurar y añadir estas mejoras.

DSpace está escrito en Java. Emplea Base de Datos Oracle o PostgreSQL. Es multiplataforma compatible con Linux, Mac OSX y Windows. Soporta los protocolos: OAI-PMH, OAI-ORE, SWORD, WebDAV, OpenSearch, OpenURL, RSS, ATOM. La interface es totalmente personalizable. Reconoce los tipos de archivo más habituales como son textos, imágenes, videos y audio; además de poder agregar nuevos formatos. Utiliza metadatos por default Dublin Core, pero es posible usar MARC Y MODS. Sus campos para realizar búsqueda son personalizables y se encuentra disponible en más de 20 idiomas.

DSpace enfatiza entre sus características, ser una herramienta para preservar a través del tiempo los contenidos, por lo que al utilizarla le concedemos la libertad de realizar copias de los ítems con fines de conservación y respaldo; por su parte DSpace siempre hará mención del autor y no hará modificaciones a los contenidos.

DSpace es el software libre para la implementación de repositorios más popular a nivel mundial; es utilizado por más de 1000 repositorios de acceso abierto en todo el mundo. [2]

C. Fedora

FEDORA (Flexible Extensible Digital Object Repository Architecture), también llamado Fedora Commons, es liderado por el Fedora Project Steering Group bajo la administración de DuraSpace. Fue originalmente desarrollado en 1997 por Sandy Payette and Carl Lagoze en la Universidad de Cornell y fue pensado para almacenar, gestionar y tener acceso a objetos digitales. En el 2000 Thornton Staples y Rosser Wayland, crearon el primer prototipo y para el 16 de mayo de 2003 es liberada la primera versión al público de Fedora; la 1.0.

Fedora define un conjunto de abstracciones para expresar a los objetos digitales; definiendo sus relaciones y sus comportamientos. Es de código abierto y ofrece un servicio central basado en APIs; además de poseer otras características como búsquedas, protocolo OAI-PMH, administración, etc. Una característica clave de Fedora es su flexibilidad para soportar todo tipo de contenido digital.

Ofrece soporte RDF y el software repositorio está integrado con la tecnología de almacenamiento de triple semántica, incluyendo la base de datos Mulgara RDF. Fedora ayuda a garantizar que el contenido digital es duradero, proporcionando características que apoyan la conservación digital.

Cuenta con una comunidad que ha puesto herramientas adicionales, aplicaciones y utilidades que aumentan las posibilidades del repositorio de Fedora incluyendo Hydra, Islandora y Avalon. Fedora es usado para la gestión de objetos digitales en colecciones digitales, bibliotecas digitales, repositorios institucionales, preservación de archivos, entre otros.

La arquitectura del servidor FEDORA se basa en 4 APIs principales: gestión, acceso, búsqueda y el Open Archival Initiative Service (para obtención de metadatos).

Existen más de 30 repositorios de acceso abiertos registrados alrededor del mundo que usan Fedora. [3]

D. Greenstone

Greenstone es un conjunto de programas para crear y distribuir colecciones de bibliotecas digitales en Internet o en CD-ROM. Es un software libre, multilingüe, publicado bajo los términos de la Licencia GPL. Greenstone lo produce el New Zealand Digital Library Project en la Universidad de Waikato, y ha sido desarrollado y distribuido en cooperación con la Unesco y la ONG Human Info de Bélgica.

Los desarrolladores de Greenstone recibieron el Premio Namur 2004 de la International Federation for Information Processing por contribuciones a la concienciación de las implicaciones sociales de las tecnologías de la información, y la necesidad de un enfoque holístico en el uso de tecnologías de la información que tenga en cuenta sus implicaciones sociales.

Greenstone se puede usar para crear grandes colecciones de documentos digitales en las que se pueden hacer búsquedas. Además de las herramientas en línea de órdenes para la construcción de bibliotecas digitales, Greenstone tiene la Interfaz para Bibliotecarios de Greenstone (GLI), gráfica, que se usa para construir colecciones y asignar metadatos.

Por medio de complementos seleccionados por el usuario, Greenstone puede importar documentos digitales en diversos formatos, incluyendo texto, HTML, jpg, tiff, MP3, PDF, video, y Microsoft Word, entre otros. El texto, PDF, HTML y documentos similares se convierten al Formato de Archivo de Greenstone (GAF) que es un formato XML equivalente. Greenstone cuenta con una interfaz en PHP llamada EmeraldView. [4]

Existen poco más de 50 repositorios abiertos registrados que han sido implementados con Greenstone.

E. Omeka

Omeka es un software libre, flexible y de código abierto pensado para la publicación en web de colecciones de bibliotecas digitales, archivos, museos o cualquier otra institución que desee difundir su patrimonio cultural. Omeka ha sido desarrollado por el Rosenzweig Center for History and New Media de la Universidad George Mason quien también es responsable del gestor bibliográfico Zotero. Se trata de una plataforma interoperable, que cumple con el estándar OAI-PMH.

La gran diferencia con otros programas pensados para construir repositorios digitales, es que Omeka no sólo es capaz de almacenar y mostrar colecciones digitales, sino que también permite crear exposiciones virtuales en torno a éstas. El sistema permite asociar una plantilla (aparición) diferente a cada exposición. El sitio Europeana www.europeana.eu utiliza Omeka para distribuir sus exposiciones virtuales.

Entre sus características soporta cualquier tipo de archivo, usa metadatos Dublin Core, creación de diferentes perfiles de usuarios.

Cuenta con una gran variedad de plug-ins para añadir funcionalidad. Permite personalizar las búsquedas. Cuenta con servicios 2.0 como Google Translate para que nuestros visitantes puedan traducir la interfaz de Omeka o Tweetster for Omeka que permite publicar tweets automáticamente cada vez que se publique un nuevo ítem en el repositorio.

Se pueden crear árboles jerárquicos de las exposiciones o personalizar el orden de aparición de los ítems.

Además se permite añadir geolocalización a nuestros ítems y situarlos en un mapa o crear líneas de tiempo. Entre otras muchas funcionalidades. [5]

III. COMPARATIVA

Una vez que hemos conocido algunos aspectos importantes de estas herramientas vale la pena realizar las siguientes observaciones.

Los primeros esfuerzos para crear repositorios abiertos no se dio en el continente americano y en contraste una de las herramientas que terminaría siendo la más utilizadas en el mundo sí; nos referimos a DSpace. Al tratarse de licencias GPL o GNU (General Public License) por sus siglas en inglés; todas estas herramientas cuentan con comunidades que dan soporte. La primera herramienta para desarrollar repositorios

de acceso abierto fue Eprints aunque en la actualidad no es la más popular y uno de las herramientas para la creación de repositorios de acceso abierto más actuales es Omeka.

Todas estas herramientas soportan los tipos de datos más comunes: .doc, .pdf, .jpg, .gif, .mp4, .mov, etc. y muchos de ellos pueden agregar nuevas extensiones de archivo para que nuevos formatos puedan ser almacenados. Todas ellas cuentan con soporte multilinguaje. Se pueden personalizar las búsquedas y permite registro y autenticación de usuarios. Además de contar con interfaces personalizables.

En la Tabla I podemos observar que existe diversidad en los lenguajes en los cuales han sido desarrolladas estas herramientas; así como las Bases de Datos que manejan. Por lo que debemos estar seguros que podemos encontrar una que se adapte a nuestros requerimientos específicos. Otra característica que vale la pena resaltar es que todas ellas son compatibles con el protocolo OAI-MPH y los metadatos Dublin Core. Una cosa a tener en cuenta al momento de pensar en la interoperabilidad.

TABLE I. COMPARATIVA

	EPrints	Green stone	DSpace	Fedora	Omeka
Lanzamiento	EPrints 1.0 Noviembre 2000	Green stone 2001	DSpace 1.0.1 4 de noviembre de 2002	Fedora 1.0 2003	Omeka 1.0 2 de Junio de 2009
Ultima version	EPrints 3.3.13	Green stone 2.86	DSpace 5.0	Fedora 4.1.0	Omeka 2.2.2
Almacena	Data Object	Docu ment	Item	Digital Object	Item
Langu age	Pearl	Java y C++	Java	Java	PHP
Base de Datos	MySQL , Oracle y Postgre SQL	Gestor propio	Oracle y Postgre SQL	Depen de del sw utilizad o en conjunt o	MySQL y PostgreS QL
Protoc olo OAI-PMH	Si	Si	Si	Si	Si
Metad atos Dublin Core	Si	Si	Si	Si	Si

IV. REPOSITARIOS A NIVEL MUNDIAL

Existen varios sitios web donde se pueden realizar consultas de los repositorios implementados con distintos softwares, sin embargo los datos presentados en este documento se obtuvieron de dos fuentes: Los sitios webs oficiales de los repositorios y datos referidos en la página OpenDOAR.

OpenDOAR es un prestigioso directorio de repositorios institucionales de acceso abierto; es un proyecto liderado por la Universidad de Nottingham, en Reino Unido. Cada repositorio que se agrega en el directorio es previamente analizado por un equipo de especialistas del proyecto, quienes evalúan rigurosamente y en forma manual la calidad de la información que contiene cada uno de ellos. Es importante hacer énfasis en esto, ya que muchos otros repositorios, si bien existen; no se encuentran incluidos este directorio, y por lo tanto queda fuera de este análisis.

OpenDOAR permite realizar búsquedas tomando en cuenta distintos criterios, como son: Materia, Tipo de contenido, Tipo de repositorio, País, Idioma y Software. Y nos permite ordenar los resultados por: Nombre, País, Número de ítems, OAI URL y Software. Cuenta con una sección para visualizar los datos en gráficas y proporciona información en profundidad sobre cada repositorio que puede ser utilizada para su análisis.

OpenDOAR fue identificado por KB Oliver y R. Swain en 2006 como el líder de directorios de repositorios en un estudio realizado por la Universidad de Johns Hopkins. Y fue uno de los servicios que contribuyeron para que SHERPA fuera galardonado con el Premio Europa SPARC 2007. [6]

Como podemos apreciar en la Figura 1, el software más utilizado a nivel mundial es DSpace dejando muy por detrás a Eprints ubicado en el segundo lugar. En tercer lugar se ubican repositorios que no especifican el software con el que han sido desarrollados; como por ejemplo el sitio de Redalyc en México el cual cuenta con más de 200,000 mil ítems.

En la cuarta posición está ubicado Digital Commons desarrollado por Bepress, su licencia es de tipo comercial y es líder en ofrecer repositorios con servicio de hosting. Es compatible con el protocolo OAI-PMH. Utiliza metadatos Dublin Core. Esta desarrollado en Pearl y utiliza una base de datos PostgreSQL. [7]

En el quinto lugar de esta grafica encontramos a OPUS; originalmente se desarrolló con el apoyo de la Deutsches Forschungsnetz en 1997 y 1998 en la Universidad de Stuttgart, desde 2010 su desarrollo quedo a cargo de Cooperative Library Network of Berlin-Brandenburg (KOBV), en Berlín Alemania. Está escrito en PHP, utiliza Base de Datos MySQL, es compatible con el protocolo OAI-PMH y utiliza metadatos XmetaDiss. [8]

En Alemania, el software OPUS es el más comúnmente utilizado para la implementación de repositorios de acceso abierto. Cabe señalar que de acuerdo a OpenDOAR en el continente Americano no hay repositorios registrados desarrollados con OPUS.

Otro dato interesante, que podemos observar en la Figura 2, es que de 2849 repositorios registrados en OpenDOAR el 45.5% de los repositorios a nivel mundial están en Europa.

En la Figura 3 se observar que el país que tiene más repositorios es Estados Unidos, y el país latinoamericano que sobresa es Brasil.

La mayoría de los repositorios a nivel mundial son de tipo institucional y el mayor tipo de documentos que almacenan son artículos científicos, seguidos por documentos de tesis. Mas de

la mitad de estos repositorios están en inglés, y el segundo idioma que predomina es el español. [6].

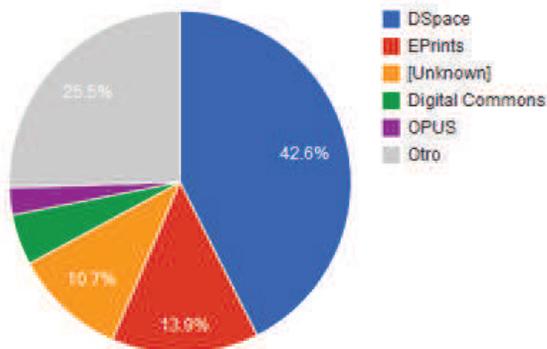


Fig. 1. Software utilizado a nivel mundial para la creación de repositorios de acceso abierto

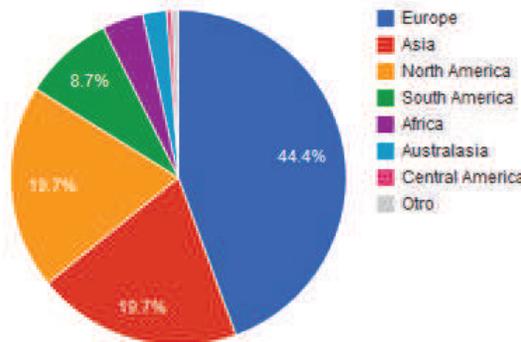


Fig. 2. Gráfica de los repositorios registrados de acuerdo al continente.

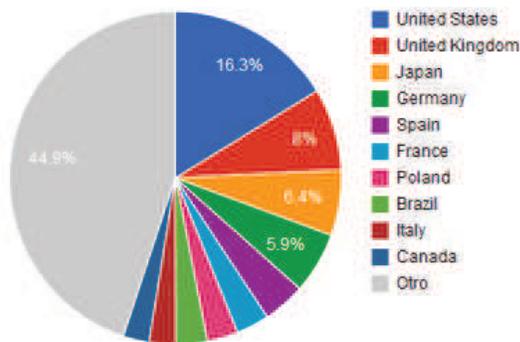


Fig. 3. Grafica de repositorios registrados por país.

V. REPOSITARIOS EN MEXICO

Siguiendo esta pauta quisimos ver la situación de los repositorios en Latinoamérica, podemos observar que Brasil es

el país latinoamericano con más repositorios. México se ubica en la quinta posición como podemos apreciar en la Figura 4.

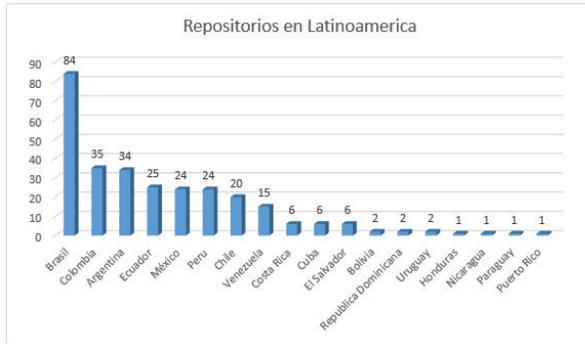


Fig. 4. Repositorios de acceso abierto en Latinoamérica.

Como se puede apreciar en la Figura 5 en México más del 50% de estos repositorios están desarrollados en DSpace repitiendo la tendencia que hemos visto reflejada a nivel mundial; 5 de ellos no proporcionan datos del software con el que se han desarrollado, 2 están implementados con EPrints, 1 con Catia, 1 con HTML, 1 con Open Repository, 1 más con SciELO y 1 desarrollo propio. La mayoría de estos repositorios manejan una base de datos PostgreSQL.

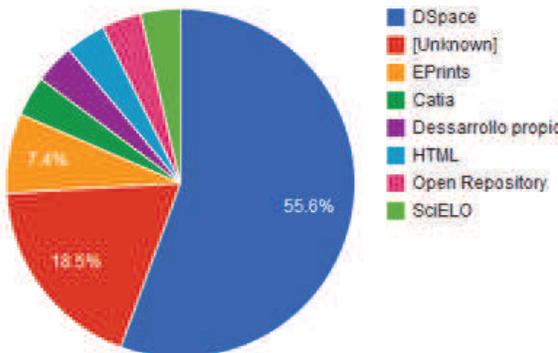


Fig. 5. Herramientas utilizadas en México para crear repositorios de acceso abierto.

En la Figura 6 podemos observar que la mayoría de los contenidos que albergan los repositorios en México consiste en artículos, seguido por los trabajos de tesis y disertaciones. Esto se debe a que los encargados de estos repositorios son las instituciones de educación superior. No es de extrañarse que estos contenidos sean los que prevalecen en los repositorios.

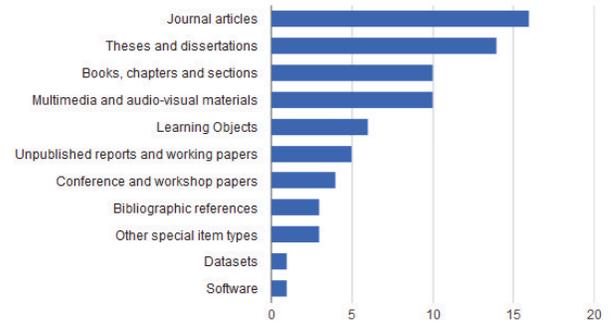


Fig. 6. Tipo de contenidos de los repositorios mexicanos.

En la Figura 7 observamos los temas que predominan en los repositorios mexicanos. En primer lugar encontramos los repositorios multidisciplinarios, seguidos por ciencias en general. Podemos intuir que conforme va madurando la idea de repositorios se van volviendo más especializados, no es una casualidad que los primeros repositorios al surgir en instituciones de educación superior fueran pensados para albergar toda la propuesta académica de los mismos.

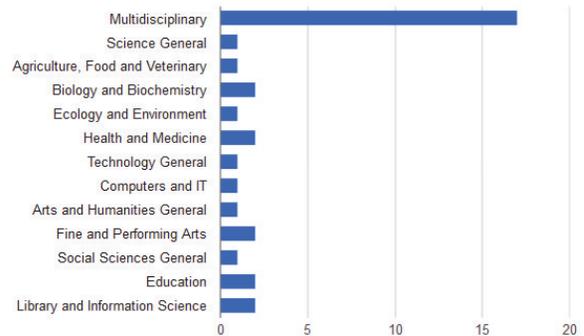


Fig. 7. Temas predominantes en los repositorios mexicanos.

Otro dato interesante y que vale la pena resaltar es que en 16 de estos repositorios, es posible recolectar datos mediante el protocolo OAI-PMH. Y que de 2010 a la fecha hubo un repunte en la implementación de repositorios pasando de 12 a 27 que son los que se encuentran registrados en OpenDOAR al día de hoy.

CONCLUSIONES

Como podemos observar es un tema muy amplio con muchas vertientes, existen más de 100 herramientas para la implementación de repositorios, todas ellas desarrolladas en distintos lenguajes, compatibles con distintos protocolos, tipos de metadatos y bases de datos.

Cada uno de ellos con distintos enfoques y filosofías. Es importante tener en cuenta que la tendencia apunta a crear repositorios de acceso abierto, sin restarle importancia a todas las implicaciones legales que pudiese traer consigo la

distribución de estos contenidos sin el pleno consentimiento de su creador. El reto ya no es únicamente implementar un repositorio, ya que con estas herramientas se facilita mucho la parte técnica; el verdadero reto está en hacerlo correctamente, empleando las técnicas adecuadas para el correcto meta etiquetado y más allá de eso que los contenidos estén correctamente almacenados, garantizar su preservación y su posterior recuperación, sin dejar de lado que estos contenidos deben ser de calidad. Solamente así es posible acercarnos al verdadero ideal de la interoperabilidad y alcanzar el verdadero objetivo de los repositorios de acceso abierto. Quizás es por ello que Alemania se ha preocupado tanto en que todos sus repositorios sean desarrollados con una misma herramienta. Ya no se trata de almacenar por almacenar. En el sitio Repositories Support Project [9], podemos encontrar algunas directrices para crear y gestionar repositorios exitosos.

La implementación de un repositorio que cumpla con los ideales para los que fue creado no es una tarea sencilla, por lo que se debe tener en claro los objetivos que se pretenden alcanzar y de este modo enfocarse y buscar las opciones que mejor cumplan nuestras expectativas; de lo contrario podría convertirse en un largo y complejo sendero.

Como trabajo futuro sería interesante aplicar métricas que nos permitan evaluar ciertas características de los repositorios implementados actualmente en México y de esta forma analizar qué tan cerca nos encontramos de alcanzar el ideal de interoperabilidad.

REFERENCIAS

- [1] EPrints, "EPrints," 2015. [Online]. Available: <http://www.eprints.org>. [Accessed 2015].
- [2] DSpace, "DSpace," 2015. [Online]. Available: <http://www.dspace.org/>. [Accessed 2015].
- [3] F. Commons, "Fedora Commons," 2015. [Online]. Available: <http://www.fedorarepository.org>. [Accessed 2015].
- [4] Greenstone, «Greenstone,» 2014. [En línea]. Available: www.greenstone.org. [Último acceso: 2014].
- [5] Omeka, "Omeka," 2015. [Online]. Available: www.omeka.org.
- [6] OpenDOAR, "OpenDOAR Directory of Open Access Repositories," 2015. [Online]. Available: www.opendoar.org. [Accessed 2015].
- [7] Bepress, "Digital Commons," 2015. [Online]. Available: <http://digitalcommons.bepress.com/>. [Accessed 2015].
- [8] K.-Z.-Z. f. I. B. ZIB, «OPUS 4,» 2015. [En línea]. Available: <http://www.kobv.de/opus4>. [Último acceso: 2015].
- [9] Jisc, "Repositories Support Project," 2015. [Online]. Available: <http://www.rsp.ac.uk/>. [Accessed 2015].
- [10] M. N. George Pirounakis, "Comparing Open Source Digital Library Software," Grecia, 2015.
- [11] I. P. D. S. D. P. A. R. G. Ileana Dayamina de la Cruz Santos, «Biblioteca virtual de la EcuRed: nueva fuente de información en la red nacional cubana,» 2011. [En línea]. Available: <http://www.redalyc.org/articulo.oa?id=181418901007>.
- [12] M. Á. G. M. Argueta, «LA INTEROPERABILIDAD Y EL INTERCAMBIO DE METADATOS EN LA RED,» 2012. [En línea]. Available: <http://www.revista.unam.mx/vol.13/num9/art95/index.html>.
- [13] R. S. B. M. Kathleen B. Oliver, "Directories of Institutional Repositories: Research Results & Recommendations," 2006.