

El Open Source como facilitador del Open Access

Lic. Diego J. Spano

Miembro del Comité Coordinador del Grupo Latinoamericano de Greenstone

Prodigio Consultores

www.prodigioconsultores.com

Buenos Aires - Argentina

Resumen

Tanto Open Source como Open Access son conceptos que generalmente se ven juntos en la literatura referida al acceso a la información. Esta proximidad no es obra de la casualidad, sino que por el contrario, si bien ambos conceptos pueden aplicarse en forma separada, los mayores beneficios se logran cuando ambos conceptos se integran y se combinan: menores costos, mejor control de la aplicación, mayor visibilidad.

El objetivo de este documento es realzar el rol que tiene el profesional informático, quien junto a los autores y los bibliotecarios conforman el trípode sobre el cual se basa todo proyecto de Repositorio Institucional. El informático será el responsable de la plataforma tecnológica sobre la cual estará montado el repositorio institucional. Y el software Open Source le brindará innumerables opciones a la hora de evaluar las diferentes alternativas.

Introducción

La definición de Acceso Abierto queda perfectamente establecida en lo que se conoce como “la definición de las 3B”, haciendo referencia a las declaraciones de Budapest, Bethesda y Berlín. En la declaración de **Berlín** (2003) se menciona que “... *por acceso abierto a esta literatura (científica) queremos decir que está disponible gratis para el público en Internet, permitiendo a los usuarios su lectura, descarga, copia, distribución, impresión, búsqueda o enlazado a los contenidos completos de estos artículos, recolectarlos para su indexación, pasarlos como datos para software o utilizarlos para cualquier otro propósito legítimo, sin más barreras financieras, legales o técnicas que aquellas que supongan acceder a Internet. La única restricción a la reproducción y distribución, y el único rol para el copyright en este dominio, debería ser dar a los autores el control sobre la integridad de su trabajo y el derecho a ser debidamente reconocido y citado*”.

La declaración de **Budapest** (2002) expresa que “...Los académicos requieren herramientas y asistencia para depositar sus artículos referidos en archivos electrónicos abiertos, una prácticas comúnmente denominada "auto-archivo". Cuando estos archivos alcanzan los estándares creados por la Iniciativa de Acceso Abierto, los buscadores y otras herramientas pueden tratar los archivos separados como uno. Los usuarios no necesitan saber qué archivos existen o dónde se localizan para encontrarlos y usar su contenido.”

La declaración de **Bethesda** (2003) refiere que “... una versión completa de la obra y todos los materiales suplementarios, incluyendo una copia de los permisos citados anteriormente, en un formato electrónico estándar apropiado se depositará de forma inmediata a la publicación inicial en al menos un repositorio en línea apoyado por una institución académica, una sociedad de intelectuales, una agencia gubernamental, o cualquier otra organización debidamente establecida que persiga facilitar el acceso abierto, la distribución sin restricciones, la interoperabilidad y el archivado a largo plazo...”.

De estas tres definiciones puede determinarse que en todo proyecto de Repositorio Institucional o Biblioteca Digital hay una interrelación implícita, necesaria y exigible entre los autores, los bibliotecarios y los informáticos, interrelación que también está determinada por tres componentes básicos, como son la información, los procesos y la tecnología.

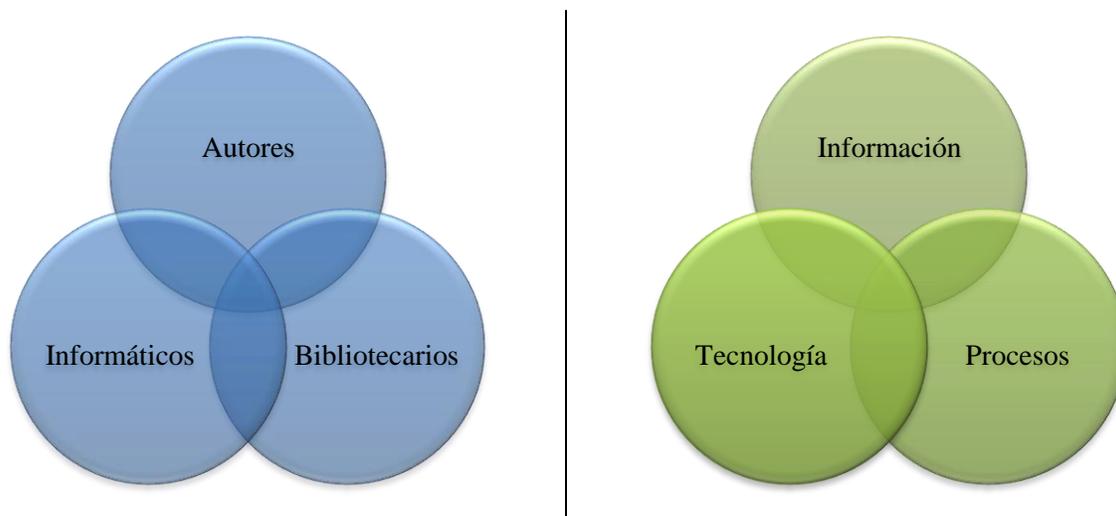


Figura 1 – Paralelismo entre roles y componentes de un repositorio digital

Los autores (investigadores, docentes, académicos) producen y requieren información, que los bibliotecarios a través de determinados procesos publican y comparten en el repositorio

utilizando tecnologías que los informáticos desarrollan y mantienen. Esta es la rueda sin fin que retroalimenta a los repositorios.

Los aspectos técnicos derivados de las declaraciones de Berlín, Budapest y Bethesda

Cómo mencioné anteriormente, el análisis de la “definición de las 3B” es mucho más amplio que solo limitarlo a entender qué es el Acceso Abierto. De dicha definición se desprenden una gran cantidad de cuestiones técnicas que no son responsabilidad del Autor pero que se deben tener en cuenta ya que su desconocimiento redundará en una subutilización del repositorio, o bien, su falta total de visibilidad. Es decir, el Open Access si bien es ya un logro por sí mismo, no estará completo hasta tanto no se lo acompañe con tecnologías abiertas y estándares abiertos.

Entre esas cuestiones técnicas están las siguientes:

Disponibilidad de Acceso

“...está disponible gratis para el público en Internet...”

Ya sea tanto para los autores como para los usuarios, es necesario que el acceso al repositorio sea posible desde cualquier ubicación, mediante la sola utilización de un navegador de Internet. Esto independiza el acceso del lugar físico donde reside el repositorio.

Herramientas para Autoarchivo

“... Los académicos requieren herramientas y asistencia para depositar sus artículos referidos en archivos electrónicos abiertos, una prácticas comúnmente denominada "autoarchivo"...”

Para el caso en que se opte por el autoarchivo o vía verde, se requiere el desarrollo de una interfaz sencilla de usar, con ayuda en línea y lo suficientemente intuitiva como para evitar problemas en la subida del documento.

Recuperación de la información

“...permitiendo a los usuarios su lectura, descarga, copia, distribución, impresión, búsqueda o enlazado a los contenidos completos de estos artículos, recolectarlos para su indexación...”

Este es quizás uno de los aspectos más importantes, ya que tiene un impacto altísimo en las estadísticas de uso del repositorio. La recuperación de la información debe ser posible tanto mediante búsquedas estructuradas como no estructuradas. En el primer caso, el objetivo se logra mediante una correcta definición de los metadatos asociados al documento. La utilización de estándares de metadatos como lo es por ejemplo el esquema Dublin Core,

facilita la recuperación de la información. El segundo caso, el de las búsquedas no estructuradas (también llamadas búsquedas por texto completo o full text), hace uso del contenido de los documentos, el cual es indexado y puesto a disposición para la realización de las búsquedas. Pensar en la implementación de un repositorio que no permita las búsquedas a texto completo, es acotar en un 90% el acceso a la información buscada.

A esto se le suma también las distintas variantes que existen respecto a las clases de búsquedas posibles sobre los metadatos o sobre el contenido. Las variantes van desde las simples búsquedas por palabras o frases, el uso de operadores booleanos – Y, O, NO -, búsquedas con operadores de proximidad y comodines, búsquedas semánticas, y por último, las búsquedas difusas, es decir, aquellas que apelan al grado de similitud en la escritura entre lo buscado y lo encontrado.

Formatos de archivo

“... una versión completa de la obra y todos los materiales suplementarios, incluyendo una copia de los permisos citados anteriormente, en un formato electrónico estándar apropiado...”

La gran variedad de formatos de archivo que actualmente está disponible es en muchos casos una complicación a la hora de diseñar un repositorio institucional. Por tal motivo, siempre que sea posible, se debe homogeneizar el formato de los archivos admitidos en el repositorio para que ello redunde en una mayor claridad en el acceso a la información contenida. Debe optarse por aquel formato que cumpla con la mayor cantidad de estándares y que no esté subordinado a una aplicación específica o propietaria.

Interoperabilidad y preservación

“... facilitar la distribución sin restricciones, la interoperabilidad y el archivado a largo plazo...”

Un repositorio institucional no debe ser tomado como un sistema estanco, aislado. Por el contrario, el mayor potencial se da en la posibilidad de interoperar con otros repositorios, en lo que se conoce como "harvest" o cosecha de metadatos. Junto a la iniciativa del Open Access está ligada la Iniciativa de Archivos Abiertos (OAI, por sus siglas en inglés) que contempla en su definición un protocolo de intercambio de información entre repositorios llamado OAI-PMH. La misión del OAI es desarrollar y promover estándares de interoperabilidad que faciliten la disseminación de contenidos.

La interoperabilidad tiene su punto de apoyo en la preservación a largo plazo. Esta preservación hace referencia al conjunto de políticas de copias de respaldo, el hardware y software utilizado, la inalterabilidad de la definición de las URIs y todo aquello que tenga como objetivo lograr que el documento sea accesible en el largo plazo. La preservación también estará dada por la multiplicidad de copias de un mismo documento que puede alcanzarse mediante el intercambio entre repositorios.

Interfaz con el usuario

“...Los usuarios no necesitan saber qué archivos existen o dónde se localizan para encontrarlos y usar su contenido...”

El usuario debe poder acceder al repositorio mediante una interfaz que contemple una amplia variedad de opciones de búsqueda. Además de las búsquedas, también es necesario que el usuario pueda recorrer visualmente el contenido del repositorio, siempre teniendo presente que el proceso de acceso al documento debe ser transparente, sin importar donde esté alojado.

Una solución al alcance de todos: software Open Source

El conjunto de requerimientos técnicos comentados anteriormente genera la necesidad de contar con aplicaciones de software que aporten soluciones a la hora de crear, implementar y compartir un repositorio institucional, ya sea que los documentos se ingresen mediante el autoarchivo o bien, que sea el resultado de un proceso de gestión de la edición científica. El camino más directo en busca de esas soluciones nos lleva al software Open Source.

El software Open Source (Código Abierto) es aquel que incluye el código fuente y está disponible, por lo general, sin cargo. Existen algunos requerimientos adicionales que un software debe cumplir para ser considerado open Source: el software debe ser libre de distribuir, se debe permitir modificaciones derivadas del mismo, la licencia no debe discriminar a ninguna persona y la licencia no debe limitar ningún campo de aplicación o emprendimiento.

El software que cuenta con una licencia Open Source permite que una comunidad de desarrolladores a lo largo de todo el mundo lo mejore, lo corrija, lo pruebe y lo mantenga.

Si bien el software de código abierto, al igual que cualquier otra aplicación comercial (o “cerrada”) requiere de recursos para poder instalarlo, configurarlo y mantenerlo, tiene como principales ventajas el hecho de que permite a quien lo use independizarse de quien lo desarrolló, usarlo sin necesidad del pago de licencias, probarlo en su totalidad ya que no hay funciones que estén bloqueadas, modificarlo para ajustarlo a requerimientos específicos, y por sobre todo, se tiene la posibilidad de comprobar cada una de sus funciones. El software cerrado, por el contrario, tiene la gran desventaja de ser una caja negra en donde se desconoce el uso interno que hace de la información.

Herramientas Open Source para edición científica

Antes de comenzar a mencionar las principales herramientas Open Source para la edición de documentos científicos, es necesario tener presente cuales son los criterios básicos con

los cuales se pueden comparar dichas aplicaciones. Toda comparación es válida siempre y cuando se haga bajo un conjunto de requerimientos específicos y mensurables de manera concreta. En los próximos párrafos se listarán una serie de aplicaciones, aunque no hace referencia al grado de cumplimiento de los criterios mencionados, ya que su análisis pormenorizado está fuera del alcance de este documento.

Estos criterios de evaluación pueden agruparse bajo las siguientes categorías:

Programa de edición

- Capacidad para soportar múltiples publicaciones
- Capacidad para manipular múltiples roles administrativos
- Capacidad para configurar flujos editoriales específicos para cada publicación en particular

Escalabilidad

- La escalabilidad de una aplicación se medirá en términos de su capacidad para soportar la demanda de consulta de múltiples publicaciones sin que ello perjudique la experiencia del usuario
- Compatibilidad con otras aplicaciones – es la capacidad que demuestra la aplicación para comunicarse con otras aplicaciones de publicación o con repositorios de documentos.

Subida de Contenidos

- Simplicidad de uso de la plataforma
- Capacidad de pegar o importar desde otro documento
- Captura automática/creación de metadatos (Autor, Tema, Palabras claves) para todos los contenidos.
- Capacidad para determinar metadatos opcionales y obligatorios

Administración online del proceso editorial

- Control de acceso basado en roles: autor, editor, revisor.
- Integración del workflow con sistemas de email o RSS para la recepción de notificaciones

Marcado de contenido y Upload

- Manejo de lotes de documentos
- Conversión automática a formato abierto (XML por ejemplo)
- Soporte de cualquier tipo de formato.

Publicación y presentación

- Disponibilidad de plantillas predefinidas
- Posibilidad de modificar la interfaz de presentación de los documentos.

Requerimientos adicionales

- Existencia de material de estudio
- Información de ayuda, Wiki, FAQ.
- Nivel de soporte del desarrollador o de la comunidad creada en torno al software
- Capacidad de extracción de estadísticas de uso

A continuación se mencionan los sistemas de edición científica más destacados:

DPubS

Sitio web: <http://dpubs.org/about.html>

DPubS (Digital Publishing System) es un software open Source diseñado para organizar, presentar y distribuir publicaciones científicas, monografías, presentaciones y cualquier otro tipo de publicaciones relacionadas. Es desarrollado por la Universidad Cornell .

Open Journal System

Sitio web: <http://pkp.sfu.ca/?q=ojs>

Open Journal Systems (OJS) es un sistema de gestión y publicación de revistas que ha sido desarrollado por el Public Knowledge Project de Canadá (<http://pkp.sfu.ca/>) para ampliar y mejorar el acceso a la investigación científica. Permite modelar el proceso completo de producción y publicación, desde la presentación inicial hasta su posterior archivo.

OJS está muy difundido en todo el mundo. En el siguiente gráfico puede observarse la distribución geográfica de las instalaciones de OJS por continente.

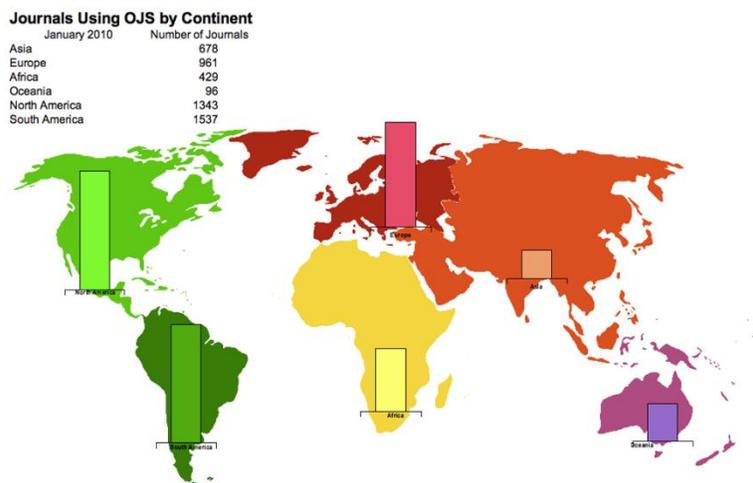


Figura 2 – Distribución geográfica de instalaciones de OJS. (Fuente: <http://pkp.sfu.ca/files/OJS-Journal-Map-Jan-10-lrg.png>)

E-Journal

Sitio web: <http://drupal.org/project/ejournal>

Esta aplicación es un módulo perteneciente a Drupal, un sistema de administración de contenidos. Este módulo es un potente sistema de edición que permite crear y controlar las publicaciones electrónicas dentro de Drupal, pudiendo definirse tantas publicaciones como se desee, agregar autores y editores, etc.

El módulo E-Journal está basado en el sistema Open Journal System.

Otras aplicaciones

También podemos mencionar otros sistemas como Topaz (<http://topazproject.org/>), DIVA Project (<http://www.diva-portal.org/about.xsql>) y Rhaptos (<http://rhaptos.org/>), pero el común denominador de estas aplicaciones es su falta de actualización con respecto a los otros sistemas mencionados.

Herramientas Open Source para repositorios digitales

Como mencioné anteriormente, la publicación en Acceso Abierto no siempre recae en un proceso editorial, sino que en muchos casos, los autores depositan sus trabajos en lo que se

llaman Repositorios Institucionales. Esta forma de publicación conocida como “autoarchivo” o “ruta verde” permite la rápida visibilidad de los trabajos, ya que una vez subido el documento, el mismo comenzará a ser cosechado y citado por otros repositorios, lográndose de esta manera una diseminación global.

A continuación se listan las principales aplicaciones para repositorios digitales.

Dspace

Sitio web: <http://www.dspace.org/>

Dspace es un sistema desarrollado por el MIT junto a Hewlett-Packard . El sistema fue expresamente creado como un repositorio digital para capturar la producción intelectual de una institución, integrando a las comunidades de usuarios junto a la estructura del sistema. Su diseño permite la participación de los autores, las instituciones, los departamentos y otras unidades típicas de una organización, todo ello guiado por un workflow que permite la aceptación, la revisión o el rechazo de un documento.

Dspace es una de las aplicaciones para repositorios más difundida a nivel mundial.

EPrints

Sitio web: <http://www.eprints.org/>

Eprints es desarrollado por el Departamento de Ciencias de la Computación y Electrónica de la Universidad de Southampton, en el Reino Unido. Es un software distribuido sobre la base de la Licencia Pública General GNU y tiene una amplia comunidad mundial.

EPrints es un sistema de repositorio que provee un acceso sencillo a trabajos ya publicados. Como tal, no modela el proceso completo de revisión por pares y la producción final.

Fedora

Sitio web: <http://www.fedora.info>

La Universidad de Virginia y la Universidad Cornell han desarrollado en conjunto esta aplicación que permite implementar un repositorio con seguimiento de políticas, versionado de documentos y gran escalabilidad.

Greenstone

Sitio web: <http://www.greenstone.org>

Greenstone es desarrollado en la Universidad de Waikato en Nueva Zelanda y cuenta con el apoyo de UNESCO y la ONG Human Info de Bélgica. Es un sistema de muy fácil instalación que permite crear y publicar bibliotecas digitales en donde pueden administrarse una amplia variedad de objetos. Si bien su principal campo de acción son las bibliotecas digitales, también permite la aplicación de un formulario web para el autoarchivo de documentos, emulando de esta manera un software para repositorios institucionales.

Conclusión

El Open Source y el Open Access son dos corrientes que van en la misma dirección. La sinergia entre ambas genera un sinnúmero de consecuencias positivas para los autores, las bibliotecas, las instituciones académicas y los usuarios finales, quienes en definitiva, logran acceder a un universo de información científica sin otra restricción más que una conexión a Internet. En definitiva, tanto el acceso, como el código como así también los estándares tienen su punto común en un adjetivo: “Abiertos”. Partiendo de esa base, cualquier persona podrá sentirse parte del universo digital.

Es importante resaltar que más allá de la metodología empleada para llegar a un artículo final, ya sea a través de alguna de las herramientas para edición indicadas anteriormente, o bien, por alguna otra metodología, siempre existirá la posibilidad de que ese artículo sea enviado a alguno de los portales multidisciplinarios de revistas, como lo son SCIELO, REDALYC, DOAJ, o portales temáticos como CLACSO.

Bibliografía

- Corrado, Edward M. 2005. The importance of open access, open Source, and open standards for libraries. Issues in Science and Technology Librarianship n. 42 (spring). [Online]. Disponible: <http://www.library.ucsb.edu/istl/05-spring/article2.html>
- Cyzyk Mark and Choudhury Sayeed. 2008. A Survey and Evaluation of Open-Source Electronic Publishing Systems (Consulta: julio 2010)
- Open Society Institute. 2004. A Guide to Institutional Repository Software. (Consulta: julio 2010)
- Declaración de Budapest. <http://www.soros.org/openaccess/esp/read.shtml> (Consulta: julio 2010)
- Declaración de Bethesda. http://ictlogy.net/articles/bethesda_es.html (Consulta: julio 2010)
- Declaración de Berlín. <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html> (Consulta: julio 2010)
- Babini, Dominique y Jorge Fraga (compiladores). 2006. Edición electrónica, bibliotecas virtuales y portales para las ciencias sociales en América Latina y El Caribe. Buenos Aires: Consejo Latinoamericano de Ciencias Sociales, 2006. 271 p. <<http://bibliotecavirtual.clacso.org.ar/ar/libros/secret/babini/babini.html>> (Consulta: julio 2010)
- Barton, Mary R. y Waters, Margaret M. 2004. Cómo crear un Repositorio Institucional : manual LEADIRS II. Cambridge: MIT Press. 169 p. <<http://www.recolecta.net/buscador/documentos/mit.pdf>> (Consulta: julio 2010)
- De Volder, Carolina. 2008. Los repositorios de acceso abierto en Argentina: situación actual. Información, cultura y sociedad. n.19. pp. 79-98. <http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-17402008000200005&lng=es&nrm=iso> (Consulta: julio 2010)
- Iniciativa de Archivos Abiertos – OAI < <http://www.openarchives.org/>>
- Witten, Ian y Diego Spano. 2010. Greenstone en América Latina: Una ola expansiva de más de 11.000km. En: Cyranek, Günther (editor). Greenstone: Un software libre de código abierto para la construcción de bibliotecas digitales. Experiencias en América Latina y el Caribe. Montevideo: UNESCO, 2010. pp 21-

32 <<http://www.unesco.org.uy/ci/fileadmin/comunicacion-informacion/Greenstone-paraweb.pdf>> [Consulta: 06 agosto 2010].