

Proceso de evaluación de Software para la creación de repositorios Open Source para alojar tesis y trabajos de investigación en el NEA. Propuesta de utilización de repositorios para publicar producciones TIC de nivel escolar primario y secundario.

Autores: Héctor Abel Bareiro; Oscar Alberto Estigarribia; Rubén Alberto Morenate

Área de Informática. Facultad de Humanidades y ciencias Sociales. Universidad Nacional de Misiones. Año 2015

lilibe@ymail.com, hectorbareir@gmail.com, pcskoki@yahoo.es, rubenmorenate@gmail.com

Área de Interés:

Educación en Tecnología: Articulación entre la Escuela Media y la Universidad en el área de TIC

Resumen

Este trabajo relata el recorrido efectuado para la selección del software, DSpace, utilizado para la creación del repositorio Institucional “Argos” de la Secretaría de Investigación y Posgrado de la Facultad de Humanidades y Ciencias Sociales de la Universidad Nacional de Misiones. Esta labor se desarrolló en el marco del proyecto de investigación “16H315- Repositorios de tesis de posgrados: Capacidad del sistema académico del NEA para la generación de depósitos de tesis de acceso libre”

Esta investigación se enfocó en la problemática del mundo académico en el que existe una gran cantidad de literatura científica y académica, producto de grandes esfuerzos, tanto del Estado como de otras instituciones y obviamente de los estudiosos e investigadores que generan, a diario, conocimiento especializado. La idea de dar visibilidad a dicha producción genera un sinnúmero de desafíos, tanto para las universidades como para los investigadores; problemas que se pueden resumir en la pérdida del anonimato de los autores; la visibilidad de proyectos

similares; la posibilidad de compartir recursos humanos y económicos, entre otros.

Como consecuencia de esta investigación se deduce que los repositorios tienen capacidad suficiente para alojar producciones TIC de los docentes de escuelas primarias y secundarias, con el fin de reunir estos materiales de manera ordenada ideas y experiencias para su mejor aprovechamiento, favoreciendo el crecimiento del conocimiento colectivo y la inclusión digital.

La capacidad que tiene los repositorios para organizar y facilitar el acceso a la producción académica y científica, ya sea por institución educativa, como por docentes, disciplinas, etc. Hacen que estas aplicaciones sean más que recomendables, puesto que inclusive permite incluir la revisión de los documentos por parte de otros miembros de la comunidad para asegurar la calidad de los documentos almacenados.

La factibilidad de todo lo expresado es alcanzable, puesto que las universidades nacionales han estado invitadas a desarrollar repositorios y, asumiendo el desafío, varias de ellas los vienen implementando desde hace algunos años, en consecuencia ya disponen del

conocimiento técnico y de los materiales necesarios para su creación, instalación, gestión y difusión mediante actividades de extensión o transferencia de conocimiento.

Durante el periodo de investigación, referente a la parte tecnológica del desarrollo de repositorios digitales, se analizaron los antecedentes teóricos y prácticos procedentes de otras investigaciones relativas a las distintas plataformas *open source* que pueden dar cabida al desarrollo de repositorios con sus distintas particularidades, hasta llegar a la que se consideró más adecuada.

El trabajo muestra el análisis de las distintas herramientas, la evaluación de estos, según sus características, grado de aceptación, difusión, cantidad de sitios web alojados, números de descargas, número de fallas, etc. reportados desde organizaciones como OpenDoar y Roar.

Si bien este artículo puede parecer muy técnico pretende poner a disposición del lector los pasos que permitirían repetir la experiencia de instalación de la plataforma seleccionada, DSpace.

Cómo referenciar éste artículo: HA Bareiro, OA Estigarribia, RA Morenate (2015). Proceso de evaluación de Software para la creación de repositorios Open Source para alojar tesis y trabajos de investigación en el NEA. Propuesta de utilización de repositorios para publicar producciones TIC de nivel escolar primario y secundario. 16H315. Congreso de tecnología en educación y educación en tecnología (TE&ET). (Corrientes 2015)

Palabras Claves: Tecnologías de la información y de la comunicación, Evaluación de software de código abierto; Administración de colecciones digitales; Repositorios Institucionales; E-tesis; Argentina NEA

Elección del software

Para la elección del software más adecuado para la creación del Repositorio Institucional de la Secretaría de Investigación y Posgrado de la Facultad de Humanidades y Ciencias Sociales de la Universidad Nacional de Misiones, se formó un equipo interdisciplinario integrado por bibliotecarios e informáticos, donde cada miembro aportó su experiencia sobre cómo debería funcionar el sistema y las condiciones necesarias, tanto las de servicio (metadatos, aportación de datos, tipos de contenido etc.) como las relacionadas con los servidores subyacentes (sistemas operativos, bases de datos, mecanismos de búsqueda, etc.). Se efectuó un minucioso análisis de los programas existentes, teniendo en cuenta los distintos aspectos a considerar: Tecnología básica de utilidad para un Repositorio digital Institucional, Modelos y características del producto, Otros aspectos técnicos sobre su funcionamiento, Pasos para la implementación, Costes, Características principales y proveedores del software.

Repositorios Analizados. Características

Entre los softwares libres disponibles en la web se optó por analizar los más utilizados, presentados a continuación, para comprobar sus ventajas.

Dspace: es un sistema de información con arquitectura de repositorio digital que captura, almacena, ordena, preserva y distribuye material de investigación digital. Está desarrollado en plataforma *opensource* y se puede personalizar según las necesidades. Es un proyecto desarrollado de manera conjunta por las bibliotecas del MIT (*Massachusetts Institute of Technology*) y

Hewlet-Packard Co. El software desarrollado por el MIT es de código fuente abierto y está disponible gratuitamente para cualquier institución que desee almacenar y administrar su producción digital. DSpace permite el almacenamiento estandarizado de documentos con un sistema de indexación y búsqueda sobre metadatos y de texto completo (Opcional) lo que convierte al repositorio de documentos digitales en una opción de preservación a largo plazo, ya que crea URLs permanentes para los materiales almacenados y permite la realización de copias de seguridad automáticas de los archivos de una institución a otra.

Eprints: es una herramienta basada en software libre, desarrollada por un equipo dirigido por Christopher Gutteridge, de la Universidad de Southampton, quien para su desarrollo tomó en cuenta que los repositorios de información se basan en una filosofía de colaboración, en la que los investigadores colocan a disposición de otros colegas los resultados de sus investigaciones o sus recursos. Es un software de almacenamiento genérico, que puede almacenar documentos en distintos formatos, imágenes, videos, audio, documentos en texto plano u otros. Para la incorporación de los documentos al sistema es necesario dividir las tareas entre los especialistas que trabajarán en él. Esta división de funciones es importante para sostener la calidad en el proceso de alojamiento de documentos y para la recuperación de la información. El software permite definir distintas funciones entre los participantes en un proyecto; cada uno de ellos posee ciertos permisos para realizar determinadas tareas en el sistema. De forma general, se asignan tres papeles fundamentales: usuario, editor y administrador.

Usuarios. Son las personas que emplean el repositorio para buscar información o proponer la incorporación de nuevos documentos. Cada material que se propone hospedar debe llevar una serie de metadatos para describirlo y representarlo.

Editores. Poseen las mismas facilidades de los usuarios pero, además, los que pertenezcan a este grupo deciden si aceptan o rechazan las propuestas de hospedaje; editan los metadatos; eliminan documentos con o sin notificación a quienes los propusieron, en determinadas circunstancias.

Administradores. Son quienes presentan las mayores responsabilidades. Además de poder realizar todas las acciones de los grupos anteriores, se ocupan de la administración y el buen funcionamiento del repositorio.

Cualquier otra persona que no esté en ninguno de estos grupos, puede consultar la documentación almacenada.

Zentity: Microsoft Zentity2.0 es un repositorio de información diseñado específicamente para instituciones académicas, gubernamentales y científicas que realizan y/o recopilan investigaciones, permitiendo guardar sus trabajos académicos digitales: *papers*, clases, presentaciones, videos. Está construido sobre la plataforma Microsoft.NET *Framework* 4.0, utiliza Microsoft *SQLServer* 2008. Sistema operativo *Windows Server*2008. Esta herramienta tiene la habilidad de encontrar y explorar interesantes relaciones de manera visual y mediante programación. Con estas características, los investigadores pueden acceder, analizar y descubrir estructuras y relaciones ocultas entre datos, así como extender los modelos de información existentes. Además permite descubrir las estructuras que antes estaban ocultas y las relaciones entre los elementos de datos, así como también, ampliar los modelos de datos existentes mediante la adición de las relaciones y propiedades adicionales a estas relaciones.

Incluye una plataforma Web, permitiendo a la interfaz de usuario consultar la base de datos, revisar y actualizar los registros, así como crear y modificar las relaciones entre los elementos que se almacenan.

Fedora: (*Flexible Extensible Architectur Digital Object Repository*) fue desarrollado por investigadores de la Universidad de Cornell, cuyo objetivo arquitectural es almacenar, administrar y acceder a los contenidos digitales en forma de objetos digitales. Es un robusto sistema de software de código abierto. Ofrece un servicio de repositorio central basado en WEB. Además de una amplia gama de servicios de apoyo y aplicaciones, incluyendo distintos tipos de búsquedas.

El Fedora *Commons* se refiere a la comunidad que rodea el proyecto de repositorio de Fedora. Es muy activa en la producción de herramientas adicionales, aplicaciones y servicios que aumentan el repositorio de Fedora. Muchas de estas creaciones están disponibles para toda la comunidad de código abierto.

Análisis Comparativo de Software para repositorio institucional

En la búsqueda de información acerca de pruebas o evaluaciones de repositorio, en una publicación acerca del trabajo de Siddharth Kumar Singh, presentada en OR2010 (5TA conferencia internacional de repositorios abiertos), utiliza los siguientes criterios para la evaluación de repositorios:

- 1) Adopción; 2) Lanzamiento de nuevas versiones; 3) Soporte; 4) Instalación; 5) Requerimientos del Sistema; 6) Globalización; 7) Escalabilidad; 8) Autenticación; 9) Control de Acceso; 10) Estándares de Metadatos; 11) Plugins y Scripts disponibles para extender las prestaciones del software; 12) Soporte de Base

- de Datos; 13) Sostenibilidad; 14) Interoperabilidad; 15) Ecosistema de Desarrolladores; 16) Optimización de Motores de Búsqueda para mejorar la visibilidad del repositorio en los buscadores web; 17) Actualización; 18) Búsqueda; 19) Almacenamiento; 20) Desempeño; 21) Migración

El análisis también se centra en los flujos de trabajo más comunes de los repositorios:

Consumo: Analizar la cantidad de navegación y búsqueda, facilidad en las búsquedas, estadísticas de uso, notificaciones de nuevos depósitos vía correo electrónico.

Depósito: Facilidad de registrarse en el sistema. Usabilidad de plantillas de depósito y disponibilidad de campos completados automáticamente, flexibilidad para agregar, eliminar o cambiar campos en los formularios. Facilidad de corregir errores durante o luego del depósito.

Aceptación: Distintos controles de calidad. Posibilidad de realizar cambios luego del depósito y quiénes pueden hacerlo. Si se rechaza un depósito, ver la posibilidad que brinda el software a quien deposita de hacer correcciones y enviar su trabajo sin tener que empezar nuevamente todo el proceso.

Importación por lotes: Facilidad dificultad de importar gran cantidad de ítems similares.

Una vez aplicado estos criterios de evaluación, los resultados obtenidos demostraron que el repositorio más usado es DSpace (Tabla1), esta información se obtuvo analizando información disponible en las organizaciones OPENDOAR¹ y ROAR². Una última revisión

¹ **Open DOAR** es un directorio de repositorios de documentos académicos en acceso abierto. Cada repositorio que aparece en la guía ha sido visitado por personal del proyecto para acreditar el tipo de documentos que se cuelgan, lo que garantiza un alto

indica que los parámetros observados en el 2012 se mantienen a abril del 2014 en ambas organizaciones.

Nivel de uso de los repositorios analizados

Repositorio	ROAR			Open DOAR		
	2012	2014	2015	2012	2014	2015
DSpace	640	1449	1535	589	1116	1227
EPrints	335	527	558	279	374	386
Fedora	19	55	57	15	32	37
Zenity	SD	1	1	sd	SD	sd

Tabla nº 1 - Fuentes: 22/04/2015 <http://www.opendoar.org/>; 22/04/2015 <http://roar.eprints.org/>

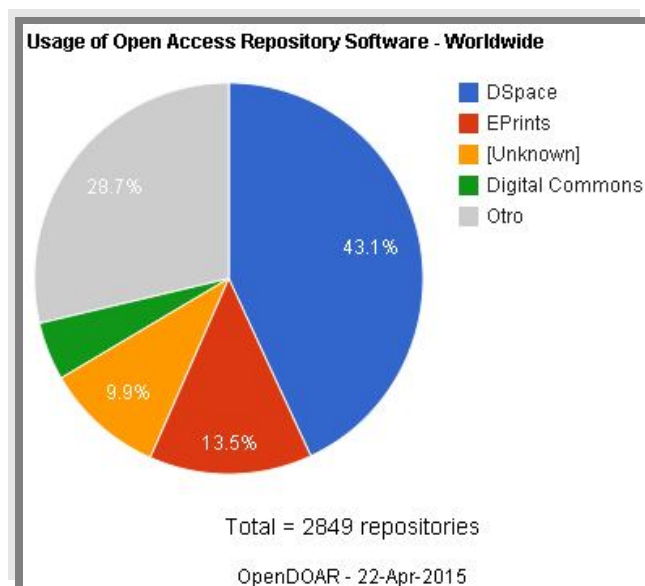


Gráfico 1: Fuentes:

nivel de calidad. Se Puede hacer búsquedas por país, por área temática, por contenido, por tipo de repositorio (institucional, gubernamental, académico).

² **ROAR:** El objetivo del informe anual es el de promover el desarrollo del acceso abierto al proporcionar información oportuna sobre el crecimiento y el estado de los depósitos en todo el mundo. El Gráfico 1 ilustra esta evolución según ROAR. La Tabla 1 presenta los datos provistos por ROAR y Open DOAR.

22/04/2015 <http://www.opendoar.org/>

Búsqueda de los componentes softwares

Se revisó la bibliografía existente sobre el tema con el fin de obtener una guía de instalación en que basarse para facilitar este proceso y se encontraron varias de ellas, entre los cuales se puede destacar “Creatingan Institutional Repository: LEADIRS Workbook” (<http://www.dspace.org/images/stories/leadirs.pdf>) y “Curso Dspace Para Bibliotecólogos (Gómez Dueñas, 2007). Se opto por la guía del Ing. Laureano Felipe Gómez Dueñas puesto que se adaptó muy bien a la metodología de trabajo del subgrupo de investigadores informáticos.

Se observó que los componentes para instalar Dspace son: la aplicación para el repositorio “DSpace”, el motor de base de datos “Postgres”, el compilador “Java JDK” y el servidor Web Java “Tomcat“. Las versiones disponibles para instalar deben estar de acuerdo a las características del servidor que se va a utilizar. Antes de comenzar a descargar los software, es necesario crear la siguiente estructura de carpetas como se indica en la imagen 1:



Imagen N° 1 “Esquema de estructura de carpetas para instalar Dspace”

Es recomendable seguir estrictamente esta estructura de carpetas puesto que los módulos del software de instaladores lo requieren de esta manera. Luego se procedió a descargar cada uno de los softwares dentro la carpeta “Instaladores”.

Instalación del software

Para completar la instalación realizamos la siguiente secuencia de trabajos:

Instalación del motor de base de datos

Postgres: algunos parámetros (tabla 2) que solicita la instalación son:

DATOS	DESCRIPCIÓN
Locale	Colocando nuestro Idioma y País
Encoding	Indicándole claramente que utilice codificación UNICODE UTF-8
Superusername	Dejamos el valor por defecto "postgres"
Password	Colocamos el valor "postgres" en ambas casillas de password.

Tabla n° 2: Parámetros instalación del motor identificados durante la investigación

También hay que tener especial cuidado durante la instalación cuando el sistema consulta si "Crea usuario administrador" se optó por "Sí", "La contraseña es muy corta" se optó por "No".

Instalación de JAVA JDK

Este kit es el más simple de instalar porque es auto asistido, solo se debe copiar el instalador en la carpeta correspondiente, y se instala sin asistencia.

Instalación de TOMCAT

Tener en cuenta que la primer parte se instala obligatoriamente en C:\Archivos de programa\Java\jre7, y luego la siguiente parte solicita un directorio y ahí se estableció en la carpeta de instalación C:\dspace\Tomcat.

Se inicia la instalación, cliqueando el botón "Siguiente" hasta la



Imagen 2: captura de pantalla durante la instalación de Tomcat

pantalla de administración donde se activan todos los módulos de instalación como figura en Imagen 2:

Al momento de solicitar usuario y contraseña se ingresaron para ambos datos: admin

Configuración de Variables de Entorno

Estos datos son del sistema operativo y se ingresan por el Panel de Control de Windows, "Opciones Avanzadas", "Variables de entorno" asignando los siguientes valores (tabla 3):

NOMBRE DE VARIABLE	VALOR
ANT_HOME	C:\dspace\ANT_HOME
CATALINA_HOME	C:\dspace\Tomcat
CLASSPATH	C:\dspace\jdk;C:\
JAVA_HOME	C:\dspace\jdk
Path	;C:\dspace\jdk\bin;C:\dspace\ANT_HOME\bin

Tabla 3: Variables de entorno de Windows identificadas durante la investigación

Se debe tener especial cuidado con la variable Path, normalmente existe en el sistema, en dicho caso accedemos a "Modificar". El mismo criterio se aplica si existen las otras variables.

Creación de la Base de Datos para DSpace

Esta etapa es todavía más técnica que la anterior, puesto que tiene una secuencia para completarla y es necesario algunas acciones con los software ya instalados. Se inició el pgAdmin III de Postgres, y en el icono rojo, con el botón derecho del mouse se eligió "Nuevo Rol de Login" aplicando nombre de rol y contraseña: dspace. A continuación se crea la "Nueva Base de Datos", y en "Codificado" hay que activar la opción UTF8. Quedando lista, de esta manera, la base de datos para DSpace.

Configuración de Parámetros

Se abrió el archivo de configuración dspace.cfg de la carpeta c:\dSPACE\dSPACE-1.4.2-source\config y se revisaron las líneas de parámetros para que estén correctamente indicadas las distintas opciones.

Compilación

Se inició copiando algunos archivos, estos se muestran en la siguiente tabla (3) que lo resume de manera muy práctica esta operación:

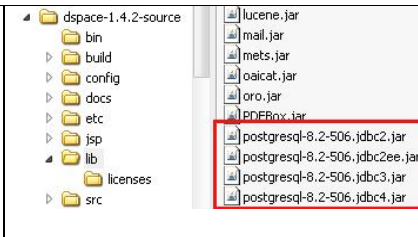
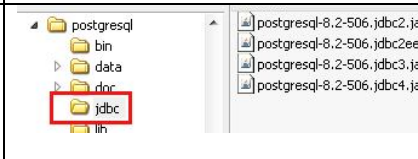
<p>Copia_mos los archivos:</p>	
<p>A:</p>	

Tabla 4: captura de pantalla durante la investigación

Luego de completar la copia de archivos se procedió a compilar DSpace. Este proceso se realizó en la carpeta que se creó con el archivo config.cfg que es: c:\dSPACE\dSPACE-1.4.2-source\config desde la ventana de comandos de “DOS³”, en ella se ejecutó el comando: antfresh_install. Si al finalizar la compilación aparece el mensaje “BUILD SUCCESSFUL” significa que se compiló correctamente.

Creación del Administrador de DSpace

Para esto es necesario trabajar dentro de la ventana de comandos de “DOS”, se ejecutó el comando: c:\dSPACE\bin\derun org.dspace.administer.CreateAdministrator

El comando anterior, antes de finalizar solicita confirme esta acción, por lo que hay que responder “Y”.

Para completar esta etapa es necesario hacer la siguiente copia de archivos (tabla 5) como se indica:


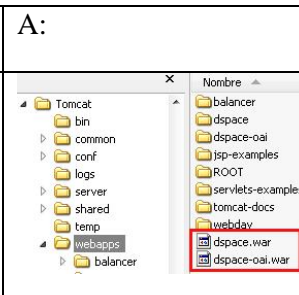
<p>Desde:</p> 	<p>A:</p> 
---	--

Tabla n° 4: Copia de Archivos. Obtenido de captura de pantalla durante la investigación.

Con el proceso copia realizado se tiene instalado DSpace y solo es necesario acceder a la plataforma para comenzar a trabajar en ella.

Resultados y Discusión

Inicio de DSpace por primera vez

Se inició el servidor Web “Apache Tomcat”. Se puede acceder a DSpace en forma local utilizando cualquier navegador Web ingresando la dirección: <http://127.0.0.1:8080/dspace/>

Si se quiere disponer de un nombre propio para la Internet, es necesario gestionar un nombre de dominio y realizar las configuraciones necesarias en la red local. Y ya se puede comenzar a operar el repositorio y hacerlo accesible al público

Comentarios sobre inconvenientes surgidos en la instalación

Inicialmente se cometió el error de no seguir paso a paso, lo escrito por el Ing. Gómez Dueña, con lo que se produjeron un montón de errores y no se pudo realizar la instalación, en un primer momento. Se reinició el proceso de instalación siguiendo estrictamente las instrucciones de este autor, respetando

versiones y organización de los directorios, con el objeto de obtener resultados positivos. Luego de superar estos primeros inconvenientes, se encontró que hay aplicaciones que son específicas para cada tipo de microprocesador, por lo que la organización de las carpetas con las aplicaciones instaladas para DSpace resultó ser un tema por demás importante.

La instalación de los software a utilizar, precisa que se configuren “variables del sistema”, que son necesarias para proveer parámetros a través del sistema operativo. La forma de realizar esto es diferente para cada versión de Windows.

Cada software a instalar requiere configuraciones especiales, como las que fueron detalladas más arriba. Al instalar el servidor Web “Tomcat”, este busca la “máquina virtual Java” que por defecto lo encuentra en “C:\Archivos de Programas” y debe respetarse esto.

Otro problema es que la versión 1.4.2 de Dspace no tenía incluido el archivo `compile.bat` que sirve para iniciar la compilación del sistema, paso fundamental para que luego funcione todo. Tampoco incluía el archivo `create_administrator.bat`. Estos dos casos se resolvieron con los comandos de “DOS” “`antfreshinstall`” dentro de la carpeta de Dspace. Y el comando “`c:\dSPACE\bin\dsrunorg.dSPACE.administer.CreateAdminister`”.

Resta realizar otros ensayos para descubrir dificultades de instalación y sus posibles soluciones, ampliándose así los antecedentes que puedan servir a otros para realizar este proceso bajo distintas condiciones. También resta probar, el enlace entre una computadora de uso común con Internet utilizando “Dyndns” o bien “Noip”.

Conclusiones

Esta investigación permitió optar por el software Dspace, para la creación del repositorio institucional “Argos”⁴, de la Secretaría de Investigación y Posgrado de la Facultad de Humanidades y Ciencias Sociales de la Universidad Nacional de Misiones, cuya misión es almacenar y mantener el acceso libre a la información científica y académica generada en sus distintos claustros.

El conocimiento adquirido en el tema y la experiencia práctica en la concreción y puesta en marcha de repositorios, permiten sugerir la extensión del uso de los repositorios digitales desarrollados por las universidades a las instituciones de otros niveles educativos mediante acuerdo o convenios que fueran necesarios.

Agradecimientos

A la Dra. Belarmina Benitez Vendrell directora del proyecto 16H315 y a todos los integrantes de este grupo de investigación que provienen de varias universidades del NEA.

Referencias

- Barton, Mary R.; Waters, Margaret M. *Cómo crear un Repositorio Institucional*. Manual LEADIRS II. 2004. The Cambridge-MIT Institute. Consultado 25 enero 2012. <http://www.recolecta.net/buscador/documentos/mit.pdf>
- Gómez Dueñas, Laureano Felipe. 2007. *Curso DSpace para bibliotecólogos: Manual de usuario*. Ecuador: Universidad de La Salle. Facultad Sistemas de Información y Documentación. URL http://eprints.rclis.org/11660/1/Instalacion_de_dspace_windows.pdf.

- Proyecto 16H315. Las tesis de posgrado: Capacidad del sistema académico NEA para la generación de depósitos de tesis de libre acceso / B. Benítez, dir. (2011-2013). Universidad Nacional de Misiones. Facultad de Humanidades y Ciencias Sociales. Secretaría de Investigación y Posgrado. Guía de Instalación de DSpace elaborada adhoc. Disponible en:
<https://docs.google.com/file/d/0ByKkDeuXaZYqdDBSYkNLcTJhWHM/edit?usp=sharing>