

**COMISIÓN NACIONAL DE ENERGÍA ATÓMICA
CENTRO ATÓMICO BARILOCHE**

**EVALUACIÓN DE SOFTWARE PARA ORGANIZACIÓN DE
RECURSOS DIGITALES**

DSPACE 1.5.2

AUTOR: BELLADONNA, MARIANO ANDRÉS

**Número de versión 01
2 de Octubre de 2009**

GLOSARIO

- *Abstract*: síntesis, sumario o resumen que informa específicamente de los contenidos y alcances de una propuesta de investigación sea esta parte de un informe final, de una tesis u otro tipo de documento afín.
- *Bit*: unidad mínima de información empleada en informática, en cualquier dispositivo digital, o en la teoría de la información.
- *Carga*: conjunto de acciones a realizar para ingresar un ítem en el repositorio.
- *Checksum*: (suma de verificación) medida simple para proteger la integridad de datos, verificado que no hayan sido corrompidos.
- *Framework*: es una estructura de soporte mediante la cual otro proyecto puede ser organizado y desarrollado. Modela las relaciones generales de las entidades del dominio. Provee una estructura y una metodología de trabajo la cual extiende o utiliza las aplicaciones del dominio.
- *Google Analytics*: servicio gratuito de estadísticas de sitios web.
- *Handles*: Sistema de identificadores persistentes que identifica direcciones de recursos asignadas por un servicio común de registración.
- *Harvester*: Recolector automático de metadatos.
- *Implementador(es)*: Persona(s) responsable(s) de la puesta en marcha, personalización y/o el mantenimiento del Repositorio institucional.

- *Indexado*: En informática, tiene como propósito la elaboración de un índice que contenga de forma ordenada la información, esto con la finalidad de obtener resultados de forma sustancialmente más rápida y relevante al momento de realizar una búsqueda.
- *Ítem*: metadatos + conjuntos de archivos en varios formatos.
- *Java*: lenguaje de programación orientado a objetos desarrollado por Sun Microsystems a principios de los años 90. El lenguaje en sí mismo toma mucha de su sintaxis de C y C++, pero tiene un modelo de objetos más simple y elimina herramientas de bajo nivel.
- *JavaServer Pages*: tecnología que permite generar contenido dinámico para web, en forma de documentos HTML, XML o de otro tipo.
- *Java Servlet*: programa que se ejecuta en un servidor dentro del contexto de un contenedor. Su uso más común es generar páginas web de forma dinámica a partir de los parámetros de la petición que envíe el navegador web.
- *Plugin*: aplicación que se relaciona con otra para aportarle una función nueva y generalmente muy específica.
- *Script*: conjunto de instrucciones que permite la automatización de tareas creando pequeñas utilidades.
- *Spam*: mensajes no solicitados, habitualmente de tipo publicitario, enviados en grandes cantidades (incluso masivas) que perjudican de alguna o varias maneras al receptor.

- *Wiki*: sitio web cuyas páginas pueden ser editadas por múltiples voluntarios a través del navegador web. La aplicación de mayor peso y a la que le debe su mayor fama hasta el momento ha sido la creación de enciclopedias colaborativas. Existen muchas otras aplicaciones más cercanas a la coordinación de informaciones y acciones, o la puesta en común de conocimientos o textos dentro de grupos.

ÍNDICE DE ABREVIATURAS Y/O SIGLAS

- *API*: (Application Programming Interface) interfaz de comunicación entre componentes de software.
- *BSD*: (Berkeley Software Distribution) Distribución de Software Berkeley.
- *CAB*: Centro Atómico Bariloche.
- *CSS*: Cascading Style Sheets.
- *CNRI*: Corporation for National Research Initiative.
- *HTML*: HyperText Markup Language.
- *IB*: Instituto Balseiro.
- *MD5*: Message-Digest Algorithm 5. Algoritmo de reducción criptográfico.
- *MIT*: (Massachusetts Institute of Technology) Instituto Tecnológico de Massachusetts.
- *OAI-PMH*: (Open Archives Initiative Protocol for Metadata Harvesting) protocolo desarrollado por Open Archive Initiative. Se usa para recolectar la descripción de metadatos de los registros en un archivo de forma tal que puedan construirse servicios usando metadatos de diversos archivos. El protocolo es también conocido como protocolo OAI.
- *RSS*: (*Really Simple Syndicate*), se utiliza para suministrar a suscriptores de información actualizada frecuentemente. El formato permite distribuir contenido sin necesidad de un navegador, utilizando un software diseñado para leer estos contenidos

ÍNDICE DE CONTENIDOS

Introducción	1
Organización jerárquica de contenidos	1
Objetivos del RI del <i>CAB-IB</i>	5
Funcionalidad del producto	7
Usuarios	8
Anónimo	8
Normal	8
Administrador	8
Grupos	8
Autorización	9
Supervisión y colaboración	10
Proceso de autoarchivo y publicación	11
Proceso de carga	13
Formatos de archivos soportados	14
Establecer fecha de embargo	16
Autocompletado	16
Otras facilidades de carga	16
Navegación del repositorio y búsqueda de objetos	18
Suscripción	19
Plataforma de interoperatividad	21
Administración de metadatos	23
Soporte para preservación	25
Identificadores persistentes	25
Soporte OAI-PMH	27
Reusabilidad de registros actuales	28
Obtención de estadísticas	30
Google Analytics	30
Otras características del producto	32

Longevidad	33
Soporte técnico	33
Comunidad de usuarios	33
Soporte pago	34
Entrenamiento	34
Documentación	35
Instalación	35
Backups	36
Interfaces gráficas	36
Conclusión	38
Bibliografía	40
Otras fuentes de información consultadas	41

INTRODUCCIÓN

El presente documento describe el proceso de evaluación de DSpace (Versión 1.5.2) como posible software para la implementación de un Repositorio Institucional de la producción intelectual local.

DSpace es un software desarrollado en conjunto por MIT Libraries y Hewlett-Packard para la construcción de repositorios digitales abiertos. Provee soporte para captura y almacenamiento de datos en formatos digital, indexado para su posterior recuperación, distribución de contenido a través de la web y preservación a largo plazo. El sistema se encuentra disponible bajo licencia BSD, lo que significa que puede ser usado, modificado e incluso integrado a una aplicación comercial sin cargos en concepto de licencia.

DSpace ofrece un sitio de demostración en línea para probar el producto <<http://demo.dspace.org/>> [18 de Septiembre 2009] y una versión en CD obtenible desde <<http://cadair.aber.ac.uk/dspace/handle/2160/565>> [18 de Septiembre 2009].

Debido a limitaciones de tiempo y disponibilidad de recursos la presente evaluación se hará sobre instancias ya instaladas (demostración en línea y live CD). Esta decisión implica sacrificar la posibilidad de evaluar determinadas características en pos de la practicidad y la duración de la evaluación.

Por esta razón, se centrará la evaluación en las características y facilidades que el software ofrece a los usuarios, considerándose solo brevemente las posibilidades de administración, instalación y/o implementación, sobre todo en la medida que estas tengan impacto en los requerimientos de los usuarios.

Organización jerárquica de contenidos

DSpace impone una estructura de organización de la información que pretende ser un reflejo de la estructura de la institución que implementa el sistema. La organización jerárquica de contenidos es el punto central del modelo de repositorio propuesto por esta herramienta y condiciona directamente a muchas de las funcionalidades a ser analizadas más adelante en este informe.

Cada instancia de DSpace esta compuesta por comunidades, las cuales representan el nivel más alto de la jerarquía de contenido. Una comunidad es una agrupación lógica de una o más colecciones y puede a su vez ser dividida en sub-comunidades. Las comunidades están compuestas por:

- Un conjunto de colecciones o sub-comunidades.
- Una página de inicio personalizable.
- Canal *RSS*, conjunto *OAI-PMH*, búsqueda y navegación dedicada.
- Un grupo de usuarios con roles de manejo y administrativos.

Las comunidades contienen colecciones, que son agrupaciones de contenido relacionado lógicamente. Una colección puede aparecer en más de una comunidad. Están compuestas de:

- Un conjunto de ítems pertenecientes a esta u otra colección.
- Un grupo de usuarios con permisos para depositar en la colección.
- Un grupo de usuarios con acceso a texto completo de los ítems.
- Un proceso de revisión / aprobación para los depósitos.
- Posible delegación de autorización para compartir, modificar y depositar.
- Página de inicio personalizada.
- Canal *RSS*, conjunto *OAI-PMH*, búsqueda y navegación dedicada.

Cada colección está compuesta por ítems, los cuales representan el

elemento básico de almacenamiento del repositorio. Cada ítem pertenece a una única colección, aunque puede aparecer en colecciones adicionales.

Los ítems están subdivididos en paquetes de archivos que reúnen archivos lógicamente relacionados (por ejemplo archivos HTML e imágenes que componen un documento HTML se podrían organizar en un mismo paquete). En la práctica, la mayoría de los ítems se divide en los siguientes paquetes:

- Original: paquete con los archivos tal cual como fueron depositados
- Miniaturas: miniaturas de cualquier archivo de imagen depositado.
- Texto: texto completo extraído de los archivos en el paquete original para *indexado*.
- Licencia: contiene la licencia de depósito que el depositante le otorga a la institución.
- CC_Licencia: contiene la licencia de distribución asociada con el ítem, si la hubiese.

Organización del informe

Se comienza estableciendo los objetivos que fueron relevados junto con el personal de la biblioteca, y, a partir de allí, se divide el informe en dos secciones principales:

- La primera, titulada "*Funcionalidad del producto*" reúne toda la información resultante de la evaluación de aquellas características funcionales de DSpace que se corresponden directamente con los objetivos antes establecidos. Así, se presenta la revisión de diferentes funcionalidades categorizadas según el objetivo que abarcan. A modo de ejemplo véase como la categoría "*Usuarios*" de dicha sección se corresponde con el objetivo "*Brindar un sistema de registro seguro de*

usuarios para proteger la identidad de los autores.”

- La segunda, titulada “*Otras características del producto*” se encarga de analizar los resultados de todas aquellas funcionalidades probadas que no fueran directamente asociables con alguno de los objetivos, pero que valga la pena considerar al momento de decidir si DSpace resulta adecuado e incluso al momento de compararlo con otras soluciones de software.

Finalmente, se presenta la conclusión de la evaluación.

OBJETIVOS DEL REPOSITORIO INSTITUCIONAL DEL CAB-IB

- Establecer una infraestructura flexible, robusta, escalable y sostenible que sirva de archivo permanente para la producción intelectual de la institución (considerando la posibilidad de que el proyecto aquí iniciado se replique en los otros centros).
- Brindar un sistema de registro seguro de usuarios para proteger la identidad de los autores.
- Que permita la autocarga/autoarchivo por parte de autores, para esto los metadatos no deberán ser muchos ni complejos, la interface lo más simple de usar posible, de fácil navegación.
- Deberá permitir la publicación de distintos objetos (variedad de contenido) por ejemplo: libros, conferencias, conjuntos de datos, objetos de aprendizaje, multimedia, patentes pre-prints, posprints, publicaciones, referencias, software, tesis, documentos no publicados, etc.
- Que permita la publicación de Tesis por parte de la biblioteca, estos requerirán más metadatos normalizados.
- Deberá permitir la obtención de un elemento cargado mediante un motor de búsqueda en el cual se pueda buscar por la mayor cantidad de criterios posibles.
- Deberá contemplar la preservación a largo plazo, esto implica, entre otras cosas, que los metadatos puedan ser exportados a distintos formatos existentes para poder realizar migraciones de versión del mismo producto y migraciones a otros productos.

- Deberá adherir al OAI-PMH, entorno Web, utilizar software Open Source gratuito, dentro de lo posible que tenga una administración del contenido sencilla, etc.
- Que permita reusar, en la mayor medida posible, los registros de la producción intelectual que la institución posee actualmente. Son características deseables a tal fin la posibilidad de importar registros en bloque y la exigencia del menor esfuerzo posible en la adaptación de los registros actuales previo a la importación (tener en cuenta que a través de los años la institución ha ido almacenando su producción intelectual en diferentes formatos).
- Que permita la obtención de estadísticas en relación a la actividad del repositorio y su contenido. Es deseable, además de la provisión de soporte nativo para recuperación de estadísticas, la posibilidad de personalizar y/o extender tanto las estadísticas disponibles como su forma de presentación.

DSPACE 1.5.2

FUNCIONALIDAD DEL PRODUCTO

USUARIOS

DSpace distingue entre tres tipos de usuarios, por orden de privilegios: anónimo, normal y administrador. Por defecto el sistema asigna un conjunto de privilegios a cada uno, los cuales pueden ser administrados directamente desde la interfase web.

Anónimo

Usuario no autenticado en el sistema.

Normal

Una vez identificado en el sistema, los usuarios normales pueden recibir privilegios según lo disponga el administrador.

Las cuentas de usuarios pueden ser creadas por el administrador o solicitadas directamente por el usuario a través de la interfase web. En el último caso y como medida de prevención contra *spam* el sistema no creará la cuenta hasta tanto no se confirme la registración mediante respuesta a un mail de confirmación enviado automáticamente a la casilla de mail provista por el usuario.

Una cuenta de usuario recién creada tendrá los mismos permisos que un usuario anónimo más la posibilidad de suscribirse a colecciones para recibir alertas ante la aparición nuevos ítems.

Administrador

Acceso irrestricto a todas las funciones en todos los ítems de DSpace.

Grupos

Son otro tipo de entidad a la que puede asignársele permisos en el sistema de autorización. Un grupo está compuesto usualmente por un conjunto de usuarios. Así, cualquier usuario identificado como miembro de un grupo obtiene los privilegios asignados al grupo. Un mismo usuario puede pertenecer a más de un grupo, obteniendo los privilegios de ambos e incluso un grupo puede ser miembro de otro grupo.

La creación y administración de tanto grupos como usuarios individuales puede realizarse directamente desde la interfase web de administración.

Autorización

El sistema de autorización en DSpace está basado en asociaciones entre objetos, acciones y listas de usuarios que pueden llevar a cabo dichas acciones sobre tales objetos. Los permisos deben ser explícitos, la falta de un permiso explícito resulta en la aplicación de la política por defecto: “denegar”.

Las políticas de autorización se pueden definir, desde más específico a más general, a nivel de:

- Ítem
 - agregar / eliminar / leer / escribir: solo asignable a grupos, no a usuarios individuales.
- Colección
 - agregar / eliminar / leer / escribir ítems de la colección.
 - leer_ítem_default, leer_archivos_default: típicamente asignados al usuario anónimo para expresar que se desea por defecto (esto es, a menos que una política más específica lo niegue) que los ítems o archivos puedan ser leídos.

- administrar: acceso irrestricto a las funciones administrativas de la colección.
- Comunidad
 - agregar o eliminar comunidades.

Supervisión y colaboración

Para facilitar, como objetivo principal, la posibilidad de que los autores de tesis sean supervisados en la preparación de la versión electrónica de las mismas, DSpace provee la posibilidad de vincular grupos de usuarios (supervisores) con un ítem en un espacio de trabajo ajeno. El grupo vinculado puede tener políticas de sistemas asociadas que permitan diferentes niveles de interacción con el ítem en el espacio de trabajo del autor; por defecto podrá asignarse:

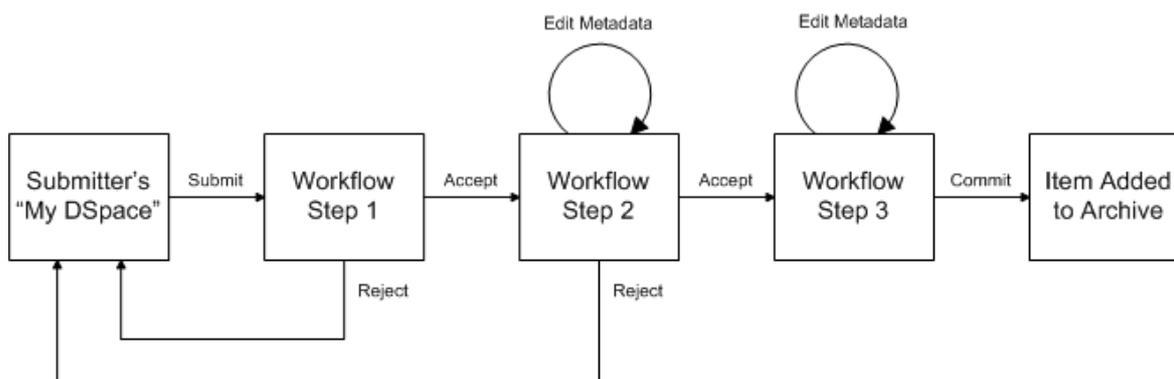
- Control editorial completo (permite editar metadatos y agregar nuevos campos).
- Visualización de contenidos del ítem.
- Ningún derecho.

Una vez asignado alguno de estos privilegios, el administrador puede modificarlos como lo haría con cualquier otro conjunto de políticas.

Esta funcionalidad puede usarse también en situaciones donde los investigadores desean colaborar en una carga ajena.

PROCESO DE AUTOARCHIVO Y PUBLICACIÓN

DSpace propone un proceso de depósito simple compuesto de a lo sumo 3 pasos, que son ejecutados en secuencia. Se puede definir para cada colección la utilización de cualquier combinación de los mismos.



Paso 1 (aceptar / rechazar): permite a un usuario simplemente aceptar o rechazar un ítem. Si lo rechazara, podría dar una razón la cual sería enviada por mail a quien hubiese hecho el depósito y el ítem aparecería nuevamente en el espacio de trabajo del depositante.

Paso 2 (aceptar / rechazar / editar metadatos): provee las mismas posibilidades que el paso 1 más la posibilidad de edición de metadatos.

Paso 3 (editar metadatos): permite solo editar los metadatos para corregirlos y/o mejorarlos.

Cada colección puede organizar su proceso de depósito independientemente de otras colecciones y cada paso se debe asociar a un grupo de revisión.

La definición tanto de los pasos del proceso de depósito como de los

usuarios o grupos de revisión para cada paso puede hacerse directamente a través de la interfase web, ya sea en el momento de crear la colección o más adelante. Esta tarea requiere privilegios administrativos.

Cuando un ítem ingresa en un paso del proceso de depósito el sistema envía un mail a los usuarios en el grupo de revisión asociado informándoles que hay una tarea pendiente. La tarea aparecerá en la lista de tareas de todos los usuarios del grupo de revisión hasta tanto alguno de ellos decida tomarla. Una vez tomada la tarea, el usuario podrá llevar a cabo las acciones correspondientes del paso del proceso, dejar la tarea pendiente pero bajo su responsabilidad o desligarse de la responsabilidad y retornar la tarea a la lista de tareas pendientes del grupo, volviendo así a aparecer en la lista de tareas de todos los usuarios del grupo.

PROCESO DE CARGA

La creación de un nuevo ítem comienza por la selección de la colección en la cual se depositará. Esto puede hacerse partiendo desde el espacio de trabajo del usuario y seleccionando la colección o directamente desde la página de la colección en la cual se desea depositar. Para proseguir, se debe contar con autorización de la comunidad para cargar ítems en la colección seleccionada.

El proceso de carga se divide en etapas:

- Etapa 1: descripción inicial del ítem. Se debe responder a ciertas preguntas respecto del ítem a depositar:
 - ¿Tiene más de un título?
 - ¿Ha sido publicado o distribuido previamente?
 - ¿Contiene más de un archivo?

DSpace personalizará los formularios en las siguientes etapas de acuerdo a las respuestas dadas.

- Etapa 2: ingreso de metadatos I. Completar campos de metadatos. Se debe de ingresar tanta información como sea posible. La presencia de algunos campos dependerá de las respuestas a las preguntas de la etapa 1. Deberán darse valores al menos para todos los campos definidos como requeridos para avanzar a la siguiente etapa.
- Etapa 3: ingreso de metadatos II. Completar más campos de metadatos. Típicamente agrupa palabras claves, *abstract*, sponsors y descripción. En esta etapa resulta de utilidad la función de vocabulario controlado (ver la sección del mismo nombre más adelante).
- Etapa 4: subir el(los) documento(s). Seleccionar y subir los archivos que componen el ítem. En el caso de que el ítem contenga más de

un archivo, podrá ingresarse una descripción para cada archivo incluido. Una vez seleccionado y subido el(los) archivo(s) DSpace permitirá definir el formato del mismo (en caso de que la detección de formato automática no haya sido posible o el usuario la considere incorrecta) y ofrecerá además la posibilidad de verificar que el archivo fue subido sin error u omisión mediante *checksum* (ver la sección del mismo nombre más adelante)

- Etapa 5: revisión de datos. Permite revisar la información ingresada durante las etapas anteriores y accesos directos a cualquiera de ellas para editar si correspondiese.
- Etapa 6: licencia. Se presenta la licencia de distribución y se ofrece a quien realiza la carga la posibilidad de aceptarla o rechazarla. La licencia es personalizable. En caso de aceptación se completa la carga y en caso de rechazo se cancela el proceso salvando el ítem en el espacio de trabajo del usuario.

En cualquier punto del proceso el usuario puede parar y salvar su trabajo para más adelante. Los datos que se hayan ingresado hasta entonces permanecerán almacenados hasta tanto se retome el trabajo. El usuario podrá ver y acceder sus cargas pendientes desde su espacio de trabajo.

Formatos de archivos soportados

La etapa 4 del proceso de carga permite seleccionar el formato de(los) archivo(s) a subir de entre una variedad de formatos que incluyen textos, presentaciones, imágenes, videos, archivos comprimidos y audio entre otros. Una vez seleccionado y subido el archivo, el sistema intentará detectar automáticamente el formato del mismo, pudiendo el usuario modificarlo si la detección no fuese correcta.

La selección del formato de archivo se hace solo a fin declarativo y de

soporte de preservación. DSpace no impone restricciones sobre formatos de archivos a subir.

DSpace permite al administrador mantener una lista de formatos soportados por el repositorio. De entre la información almacenada para cada formato es de especial utilidad para favorecer la preservación la definición de un nivel de soporte. El nivel de soporte indica la probabilidad de que la institución sea capaz de preservar el contenido en el formato correspondiente en el futuro. Existen tres posibles niveles de soporte que la institución puede asignar a un formato:

- Soportado: el formato es reconocido y la institución tiene alta probabilidad de ser capaz de hacer que archivos de este formato sean usables en el futuro, usando cualquier combinación de técnicas (migración, emulación u otras) que sea apropiada.
- Conocido: el formato es reconocido y la institución se compromete a preservar el archivo tal cual está y permitir su recuperación. La institución intentará reunir suficiente información para permitir que el nivel de soporte pueda elevarse a soportado.
- No soportado: formato no soportado. El único compromiso que asume la institución sobre archivos con formatos no soportados es el de preservarlos tal cual están y permitir su recuperación.

Los significados de cada nivel de soporte y el compromiso de la institución respecto de los mismos pueden ser personalizados. Los arriba presentados corresponden a la definición de MIT Libraries.

La administración de los formatos de archivos puede hacerse directamente desde la interfase web.

DSpace no impone límite alguno sobre el tamaño de los archivos que se

pueden subir, aunque da al administrador la posibilidad de hacerlo mediante edición del archivo de configuración principal.

Establecer fecha de embargo

DSpace no provee soporte nativo para esta funcionalidad. De todas maneras en la Wiki se encontraron dos casos de instituciones que la implementaron modificando el código y documentaron sus experiencias, ambas sobre versiones previas de DSpace:

Embargo on Bitstream (JSP)

<[http://wiki.dspace.org/index.php/Embargo_on_Bitstream_\(JSP\)](http://wiki.dspace.org/index.php/Embargo_on_Bitstream_(JSP))> [19 de Agosto 2009]

DSpace versión 1.3.2

Embargo on Bitstream v2 (JSP)

<http://wiki.dspace.org/index.php/Embargo_on_Bitstream_v2_%28JSP%29> [19 de Agosto 2009]

DSpace versión 1.4.2

Autocompletado

DSpace provee esta característica para ayudar a los usuarios a completar ciertos metadatos de manera consistente y evitar la duplicación en el repositorio.

La función esta disponible para los campos Autor y Titulo, en los cuales luego de ingresar unas letras se despliega una lista de los valores similares anteriormente almacenados en ese campo.

Otras facilidades de carga

DSpace genera una *checksum MD5* para cada archivo que almacena. La misma es usada internamente para verificar la integridad de los archivos en el tiempo. Los usuarios pueden utilizarla para asegurarse que el archivo que suben no haya sido corrompido durante la transmisión.

Ofrece la posibilidad de asociar campos con vocabularios controlados. Los vocabularios controlados son conjuntos de términos que forman un diccionario de descripciones de un tipo particular de contenido o materia, cuyo uso ayuda a estandarizar la categorización del material y por lo tanto favorece a la visibilidad del mismo. Por defecto el uso de vocabularios controlados está deshabilitado. El siguiente artículo describe como habilitar y poner en marcha esta funcionalidad:

Use controlled vocabularies

<http://wiki.dspace.org/index.php/Use_controlled_vocabularies_%28JSP%29> [19 de Agosto 2009].

Si, para una cierta colección, se deseara personalizar el formulario de recolección de metadatos en base al tipo del ítem, podría hacerse mediante la edición de código fuente. El siguiente artículo describe una implementación:

Document Type Based Submission

<http://wiki.dspace.org/index.php/Document_Type_Based_Submission> [31 de Agosto 2009].

NAVEGACIÓN DEL REPOSITORIO Y BÚSQUEDA DE OBJETOS

DSpace permite a los usuarios descubrir el contenido del repositorio de diversas maneras incluyendo:

- Búsqueda de una o más palabras en metadatos o texto completo extraído de un ítem.
- Navegación por título, autor, fecha de carga o índice de materias.

Métodos de búsqueda:

- Búsqueda avanzada: Permite ingresar una expresión de búsqueda formada por pares (metadato, valor) conectados mediante operadores booleanos (AND / OR / NOT). Permite buscar en todo el repositorio o limitar la búsqueda a una comunidad.
- Búsqueda simple: busca coincidencias con el(los) término(s) ingresados en los valores de varios metadatos de cada ítem. Si la búsqueda de texto completo está habilitada el texto completo de cada ítem también será considerado en la búsqueda.

Por defecto, la interfaz de búsqueda avanzada permite crear una expresión combinando a lo sumo tres términos de búsqueda. Se confirmó mediante consulta a la lista de correo que no existe opción configurable alguna para aumentar la cantidad de términos, sólo puede lograrse mediante edición del código de la interfaz y el código de búsqueda. Lo mismo sucede si se quisiese realizar una búsqueda no sobre el repositorio entero, tampoco sobre una comunidad individual sino sobre más de una comunidad.

El indexado de texto completo no se encuentra habilitado por defecto. El siguiente artículo describe como habilitarlo y configurarlo:

Configure full text indexing

<http://wiki.dspace.org/index.php/Configure_full_text_indexing> [20 de Agosto 2009]

Los módulos de indexado y búsqueda ofrecen una *API* que permite indexar nuevo contenido, regenerar los índices y llevar a cabo búsquedas en todo el repositorio o en comunidades o colecciones particulares. Detrás de la *API* está el motor de búsqueda Lucence. Los índices de búsqueda específicos de Lucence son configurables, permitiendo a cada institución personalizar qué campos de metadatos serán indexados.

Otro mecanismo importante para descubrimiento de contenido en DSpace es la navegación. Este es el proceso mediante el cual un usuario puede acceder a un índice particular, como el de títulos por ejemplo, y navegar a través del mismo en búsqueda de ítems de su interés. El subsistema de navegación provee una *API* que permite a su cliente especificar un índice y una subsección del mismo para seleccionar la porción adecuada del mismo. Los índices navegables por defecto son autor, título, fecha de carga y materia. Al igual que la búsqueda, la navegación puede limitarse a ítems en una comunidad o colección particular.

En ambos casos (búsqueda o navegación) los listados de resultados pueden personalizarse en base a diferentes parámetros como número de resultados por página u orden de los ítems.

Suscripción

En adición a los métodos de descubrimiento de contenido mencionados, DSpace ofrece a los usuarios la posibilidad de suscribirse a una o más colecciones de forma tal de recibir alertas cuando nuevos ítems aparezcan en esas colecciones. Cada día, los usuarios suscriptos recibirán un correo

electrónico con un resumen de los ítems que fueron agregados el día anterior a alguna de las colecciones a las cuales estén suscriptos.

PLATAFORMA DE INTEROPERATIVIDAD

DSpace provee dos scripts estándar para importar o exportar ítems. Ambos deben ser ejecutados desde la línea de comandos. Los datos a importar deben ser primero almacenados en una estructura de directorios respetando una forma de representación impuesta por DSpace. Consecuentemente, como resultado de una exportación se generara la misma estructura respetando la misma forma de representación.

La estructura a respetar se define en el manual como “formato simple de archivo DSpace”. El concepto básico es representar a un conjunto de ítems como un directorio con un subdirectorio por cada ítem. Cada directorio de ítem contendrá un archivo con los metadatos descriptivos más los archivos que lo compongan.

El proceso de importación de un ítem desde línea de comandos es el equivalente al proceso de carga desde la interfaz web por lo que una vez completado el mismo, el ítem deberá transitar los procesos definidos por el flujo de trabajo de la colección que le corresponda (ver *Proceso de autoarchivo y autopublicación* más arriba) de la misma manera que lo haría si fuese ingresado desde la interfaz web. El script de importación permite desactivar este comportamiento permitiendo que el ítem saltee los pasos intermedios de depósito y sea directamente archivado en la colección. La herramienta de importación permite no solo agregar ítems, sino además eliminarlos o reemplazarlos.

La herramienta de exportación puede exportar un único ítem o una colección de ítems, generando como salida una estructura que respeta el “formato simple de archivo DSpace”.

DSpace propone además de la carga y la importación un método alternativo de incorporación de ítems conocido como registración. El método

requiere, en vez de la ubicación de el(los) archivo(s) en una estructura de directorios definida, solo los metadatos más la ubicación de el(los) archivo(s). DSpace utiliza una variante de la herramienta de importación para llevar a cabo la registración.

Existe un tercer script provisto por defecto que permite exportar los metadatos de un ítem (junto con referencia a los archivos contenidos) en formato METS. Esta herramienta permite exportar ítems individuales, toda una colección o todo el contenido del repositorio.

Nativamente soporta:

Importación de registros con formato

“DSpace simple archive format”.

Exportación de registros DSpace hacia

“DSpace simple archive format”, METS.

DSpace ofrece además la posibilidad de extender las capacidades de exportación e importación mediante la creación de crosswalks (*plugins* de adaptación de representaciones externas de metadatos desde y hacia la representación interna de DSpace).

ADMINISTRACIÓN DE METADATOS

De un relevamiento junto con el personal de la Biblioteca de distintas opciones de conjuntos de metadatos a almacenar surgen las siguientes conclusiones:

- La calidad y cantidad de los metadatos tiene un fuerte impacto en la calidad de los servicios que pueden ser ofrecidos al usuario final.
- Los metadatos deben ser suficientes para permitir localizar los recursos de la manera más flexible y precisa posible.
- Demasiados metadatos, difíciles de identificar o ingresar tienden a frustrar a los autores en el proceso de auto-carga y por lo tanto atentan contra los objetivos del repositorio.
- Se debe encontrar un balance adecuado entre la cantidad y precisión de los metadatos y la facilidad de carga. Evitar requerir datos que no agreguen valor para el usuario final (entiéndase no solo personas sino también *harvesters*).
- La definición del conjunto exhaustivo de metadatos no es relevante en la etapa de evaluación de un software. Basta con considerar las capacidades provistas para personalizar el conjunto de metadatos, a saber, entre otras:
 - ¿Se pueden agregar, modificar, eliminar metadatos? ¿Cómo?
 - ¿Que tipos de metadatos provee?
 - ¿Se pueden integrar los metadatos no provistos por defecto a las vistas personalizadas?
 - ¿Se pueden generar búsquedas avanzadas usando como criterios metadatos no incluidos en el conjunto por defecto?

DSpace provee por defecto un esquema de metadatos Dublin Core

calificado, con los elementos y calificadores usados por MIT Libraries. Además, permite configurar múltiples esquemas y seleccionar campos de metadatos de entre una combinación de esquemas configurados para describir un ítem.

La información sobre los esquemas es mantenida en lo que DSpace llama registro de metadatos. El registro de metadatos puede ser administrado directamente desde la interfase web permitiendo editar esquemas agregando, quitando o modificando elementos.

El proceso de carga (ver sección del mismo nombre más arriba) puede ser personalizado permitiendo agregar metadatos a ser recolectados para un ítem. De esta forma se da la posibilidad de que cada colección recolecte diferentes conjuntos de metadatos para sus ítems. La personalización de los metadatos a recolectar y los formularios correspondientes debe hacerse editando archivos de configuración.

SOPORTE PARA PRESERVACIÓN

DSpace identifica 2 niveles de preservación digital:

- A nivel de *bit*: asegura que un archivo permanece exactamente igual a través del tiempo (ningún *bit* es alterado).
- Funcional: el archivo se mantiene inmediatamente usable (lo que sea que se podía hacer con el se puede seguir haciendo) a través del tiempo.

La preservación a nivel de bit es soportada mediante la recolección de metadatos técnicos respecto a los archivos almacenados (formato, *checksum MD5*, fecha de creación) y la disponibilidad de campos descriptivos adicionales para almacenar más información cuando fuese necesario. Con estos metadatos y procedimientos de producción adecuados (servidores y dispositivos de almacenamiento de calidad, buenas políticas de backup y recuperación) se disminuyen seriamente las probabilidades de alteración o pérdida de bits con el tiempo.

Con respecto a la preservación funcional, reconociendo la misma como una política a definir por la institución que implementa el servicio, DSpace ofrece la posibilidad de definir el nivel de soporte para diferentes formatos de archivos y hacer dicho compromiso visible al autor al momento de realizar una carga (ver subsección *Formatos de archivos soportados* en la sección *Proceso de carga*).

Identificadores persistentes

Un identificador persistente permite la identificación unívoca de un recurso de una manera que no cambiará incluso si el recurso fuese renombrado o reubicado y persistirá más allá del protocolo usado para accederlo. Esto significa que un recurso puede ser confiablemente referenciado en el futuro.

DSpace utiliza *handles* como medio para asignar identificadores globalmente únicos a los objetos que administra: comunidades, colecciones e ítems.

El proceso de configuración del sistema para el uso de *handles* implica registrarse en CNRI para usar un prefijo y configurar un servidor de resolución de *handles* local. Es importante considerar que los servicios de identificación y resolución de *handles* son pagos.

SOPORTE OAI-PMH

DSpace provee soporte para los protocolos de *harvesting OAI-PMH*.

Se dispone de documentación que explica exhaustivamente cómo configurar el soporte OAI para un repositorio DSpace. <http://wiki.dspace.org/index.php/Use_the_OAI-PMH_interface> [27 de Agosto 2009].

REUSABILIDAD DE REGISTROS ACTUALES

DSpace permite importar registros individuales o en bloque en el formato de importación soportado nativamente (descritos en la sección “*Plataforma de interoperatividad*” más arriba).

En caso que de contar con contenido digital que se desee depositar en DSpace pero cuya descripción de metadatos difiriese del formato soportado se puede obtener o programar lo que DSpace identifica como crosswalk para facilitar la importación en bloque de dicho contenido. Un crosswalk es un *plugin* que adapta representaciones de metadatos desde y hacia la representación interna de DSpace.

Mediante consultas a las listas de correo y búsquedas en Internet se identificaron algunos programas desarrollados por terceros con la intención de aliviar el esfuerzo de adaptación de registros para importación en bloque, entre ellos:

<<http://sites.google.com/a/ohiolink.edu/drmc/bulk-submission>> [31 de Agosto 2009].

<http://wiki.dspace.org/index.php/Google_Summer_of_Code_2008_Batch_1_import> [31 de Agosto 2009].

Queda a cargo del *implementador* la elección de alguna de las siguientes opciones, presentadas aquí en orden estimativo de preferencia:

- Adaptar los registros existentes al formato interno de representación de DSpace.
- Adaptar los registros existentes a alguno de los formatos de importación para los cuales se dispone de crosswalks.
- Desarrollar un crosswalk que permita importar registros en algún

formato al cual sea conveniente adaptar los registros de los que se dispone.

OBTENCIÓN DE ESTADÍSTICAS

Varios reportes sobre los contenidos y usos del sistema pueden ser automáticamente generados. Los reportes incluyen datos como:

- Un resumen general personalizable de las actividades del repositorio que por defecto incluye:
 - Número de visitas a ítems.
 - Número de visitas a colecciones.
 - Número de visitas a comunidades.
 - Número de pedidos *OAI-PMH* recibidos.
- Un resumen personalizable de los contenidos del repositorio.
- Listado categorizado de vistas de ítems.
- Errores y/o advertencias.
- Búsquedas ordenadas por popularidad.
- Inicio y fin de sesiones de usuarios.

Los resultados del análisis estadístico pueden agruparse por mes y están disponibles directamente desde la interfase web. Los reportes pueden configurarse como públicos o solo visibles para el administrador.

El siguiente artículo documenta los pasos para habilitar y configurar la obtención de estadísticas:

Use DSpace Statistics

http://wiki.dspace.org/index.php/Use_DSpace_Statistics_%28JSP%29 [27

de Agosto 2009]

Google Analytics

Google Analytics permite la recolección y presentación de otro conjunto diferente de estadísticas, que podrían resultar de interés tales como:

- Tiempo que los visitantes pasan en el sitio.
- Procedencia de los visitantes (desde donde llegan al repositorio).
- Términos que usaron en los motores de búsqueda que derivaron en el acceso al repositorio.
- Ubicación geográfica de los visitantes.
- Cuantas páginas acceden dentro del repositorio.
- En qué páginas comienzan y/o finalizan sus visitas.

DSpace provee soporte para integración con *Google Analytics* mediante edición de archivos de configuración o directamente del código fuente.

DSPACE 1.5.2

OTRAS CARACTERÍSTICAS DEL PRODUCTO

LONGEVIDAD

Primera versión lanzada en Noviembre de 2002. Versión actual: 1.5.2 de Abril de 2009. La versión 1.6 se encuentra en desarrollo y se espera este disponible para Noviembre de 2009. [31 de Agosto 2009].

448 repositorios registrados como usuarios de DSpace. <<http://www.dspace.org/index.php/DSpace-Instances/Repository-List.html>> [31 de Agosto 2009]

SOPORTE TÉCNICO

Comunidad de usuarios

De la experiencia propia de esta evaluación se observa la existencia de una nutrida comunidad de usuarios con buena predisposición para colaborar con el desarrollo y la popularización del producto.

La comunidad mantiene tres listas de correo:

DSpace general

<<http://mailman.mit.edu/mailman/listinfo/dspace-general>> [31 de Agosto 2009].

Preguntas generales, anuncios y discusiones sobre cuestiones no técnicas.

DSpace tech

<<https://lists.sourceforge.net/lists/listinfo/dspace-tech>> [31 de Agosto 2009].

Preguntas y discusiones técnicas.

DSpace devel

<<https://lists.sourceforge.net/lists/listinfo/dspace-devel>> [31 de Agosto 2009].

Lista para desarrolladores trabajando sobre el código de la aplicación.

Habiendo suscripto a todas las listas se puede observar que se encuentran en actividad y que las discusiones y preguntas no tardan en ser atendidas por la comunidad.

Soporte pago

Existen varias compañías que ofrecen diferentes servicios pagos para usuarios de DSpace tales como diseño, desarrollo, instalación, configuración, mantenimiento, soporte y entrenamiento.

Proveedores de servicios

<<http://www.dspace.org/index.php/Service-Providers/Service-Providers.html>>

[1 de Septiembre 2009].

Entrenamiento

Material de entrenamiento disponible en la web de DSpace, categorizado según nivel de complejidad y audiencia:

<<http://www.dspace.org/index.php/New-User/New-User-Training.html>> [1 de Septiembre 2009].

<<http://www.dspace.org/index.php/Training-Materials/DSpace-Training-Materials.html>> [1 de Septiembre 2009].

De la participación en las listas de correo y búsquedas en Internet se observó que con frecuencia se organizan cursos de entrenamiento presenciales. Se encontraron casos en que los cursos fueron anunciados como segmentos de eventos de mayor alcance sobre repositorios institucionales y casos en los que el entrenamiento fue anunciado como un evento exclusivo, siendo los últimos los

más intensivos aunque también los menos frecuentes.

Calendario de próximos eventos

<http://www.dspace.org/index.php/component?option=com_jcalpro/Itemid,314/

> [1 de Septiembre 2009].

DOCUMENTACIÓN

Disponible online en inglés.

DSpace System Documentation

<http://www.dspace.org/1_5_2Documentation/> [1 de Septiembre 2009]

Además en la wiki se listan varios manuales, preguntas frecuentes y tutoriales aunque en la mayoría de los casos referentes a versiones anteriores a la actual.

INSTALACIÓN

Prerequisitos

- Sistema operativo tipo Unix o Ms. Windows.
- Java JDK 5 o posterior.
- Apache Maven 2.0.8 o posterior.
- Apache Ant 1.6.2 o posterior.
- Motor de bases de datos relacional
 - PostgreSQL 7.3 o posterior.
 - Oracle 9 o posterior.
- Motor de servlets (Jakarta Tomcat 4.x o equivalentes).

DSpace System Documentation: Installation

<http://www.dspace.org/1_5_2Documentation/ch03.html> [1 de Septiembre 2009].

BACKUPS

El proceso de obtención y almacenamiento de una copia de seguridad del repositorio es descrito en el siguiente artículo:

BackupRestore

<<http://wiki.dspace.org/index.php/BackupRestore>> [1 de Septiembre 2009]

Los métodos analizados en el artículo son automatizables y no presentan mayor complejidad.

INTERFACES GRÁFICAS

DSpace provee soporte para dos interfaces de usuario: JSPUI y XMLUI.

JSPUI está implementado usando *Java Servlets* para manejo de lógica de negocio y *JavaServer Pages* para generación del código HTML enviado a los navegadores. Esta implementación de la interfaz de usuario permite realizar personalizaciones muy básicas rápidamente y con poca habilidad técnica (conocimiento básico de CSS y HTML). Una modificación avanzada del aspecto de la interfase requiere conocimientos en programación Java y puede llevar un tiempo considerable.

XMLUI es una implementación basada en el *framework* Apache Cocoon que persigue cuatro objetivos:

- Permitir que cada comunidad y colección representada en DSpace

mantenga un aspecto diferente.

- Aumentar y facilitar el soporte para internacionalización.
- Separar la lógica de negocio de los controles de estilo, incrementando la adaptabilidad.
- Proveer una alternativa a la implementación basada en JSP que no requiera cambios al núcleo de DSpace.

La interfase XMLUI esta formada por dos componentes: aspectos y temas. Los aspectos son extensiones de la interfase: componentes interactivos que modifican características existentes o proveen nuevas características. Los temas administran el estilo visual del repositorio, las comunidades o las colecciones. El administrador del repositorio puede definir y cambiar aspectos y temas mediante la edición de archivos de configuración.

DSpace XMLUI: Manakin Project

<[http://wiki.dspace.org/index.php/Manakin#DSpace_XMLUI: Manakin Project](http://wiki.dspace.org/index.php/Manakin#DSpace_XMLUI:_Manakin_Project)> [18 de Septiembre 2009].

CONCLUSIÓN

Los resultados de la presente evaluación permiten concluir que es factible implementar un Repositorio institucional que cumpla con los objetivos establecidos usando DSpace.

De entre las características consideradas se identifican como relevantes para la decisión entre este y otros softwares las siguientes:

- Organización de contenidos: jerarquía comunidad / colección. Posibilidad de sectorizar el repositorio con diferentes características, permisos y aspecto. Favorable si se desea descentralizar el repositorio implementando áreas heterogéneas.
- Existencia del concepto de grupo de usuarios permite mayor flexibilidad en la definición de permisos.
- Lista de formatos soportados favorece la preservación, permitiendo guiar a los usuarios sobre cómo ayudar para que su trabajo depositado perdure.
- El soporte para importación y exportación de contenido es limitado y restrictivo.
- Obtención y presentación de estadísticas funcionando en instalación por defecto.

Una visión general del producto permite destacar como característica favorable fundamental un gran trabajo en cuanto a flexibilidad del repositorio y la independencia entre comunidades, logrando la idea de varios repositorios en uno.

En base a la propia experiencia más una observación de experiencias ajenas se observa que la instalación, mantenimiento y personalización resultan de una complejidad superior a la media de herramientas competidoras.

La interfase gráfica y la organización del sitio ofrecida por defecto no resultan demasiado satisfactorias. Algunas de las facilidades del proceso de carga (autocompletado por caso) resultan bastante limitadas.

En conclusión, se recomienda la selección de este producto especialmente si se pretende hacer un aprovechamiento intensivo de su organización jerárquica. La misma requeriría un escenario con un fuerte compromiso de las diferentes áreas tanto en una etapa inicial de relevamiento de requerimientos como en la administración a futuro.

Es importante tener en cuenta como la decisión a favor de la flexibilidad representa un sacrificio de otras calidades.

BIBLIOGRAFÍA

DSpace 1.5.2 Manual. <http://www.dspace.org/1_5_2Documentation/> [11 de Agosto 2009].

DSpace Basic tutorial.

<http://www.dspace.org/index.php/Abstract_training/DSpace-Basic-Tutorial.html>

[11 de Agosto de 2009].

DSpace Wiki. <http://wiki.dspace.org/index.php/Main_Page> [22 de Septiembre 2009].

The DSpace Course. Stuart Lewis, Chris Yates.

<<http://cadair.aber.ac.uk/dspace/handle/2160/615>> [11 de Agosto de 2009].

OTRAS FUENTES DE INFORMACIÓN CONSULTADAS

Consultas particulares enviadas a:

- Lista de correo DSpace general
<<http://mailman.mit.edu/mailman/listinfo/dspace-general>> [31 de Agosto 2009].
- Lista de correo DSpace tech
<<https://lists.sourceforge.net/lists/listinfo/dspace-tech>> [31 de Agosto 2009].
- Lista de correo DSpace devel
<<https://lists.sourceforge.net/lists/listinfo/dspace-devel>> [31 de Agosto 2009].