

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

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

EPRINTS 3.1

**AUTORES: BELLADONNA, MARIANO ANDRÉS
MAROZZINI, CARINA PAOLA**

**Número de versión 02
14 de Julio 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.
- *Browser*: (*navegador o visor*) Lee la información almacenada y la presenta en el orden y agrupamiento con el que el browser fue definido en el desarrollo del sistema.
- *Buffer*: ubicación de la memoria en una computadora o en un instrumento digital reservada para el almacenamiento temporal de información digital, mientras que está esperando ser procesada.
- *Carga*: Conjunto de acciones realizadas para ingresar un objeto en el EPrints.
- *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 ofrecido por Google que genera estadísticas detalladas sobre los visitantes de un sitio web.
- *Google scholar*: buscador de Google especializado en artículos de revistas científicas, enfocado en el mundo académico, y soportado por una base de datos disponible libremente en Internet que almacena un amplio conjunto de trabajos de investigación científica de distintas disciplinas y en distintos formatos de publicación.

- *Harvester*: Recolector automático de metadatos.
- *Hosting*: (*alojamiento*) servicio que provee a los usuarios de Internet un sistema para poder almacenar información, imágenes, vídeo, o cualquier contenido accesible vía Web.
- *Implementador(es)*: Persona(s) responsable(s) de la puesta en marcha, personalización y/o el mantenimiento del Repositorio institucional.
- *Objeto*: metadatos + documentos en varios formatos.
- *Parche*: sección de código que se introduce a un programa. Dicho código puede tener varios objetivos; sustituir código erróneo, agregar funcionalidad al programa, aplicar una actualización, etc.
- *Perl*: lenguaje de programación diseñado por Larry Wall en 1987. Estructuralmente, Perl está basado en un estilo de bloques como los del C o AWK, y fue ampliamente adoptado por su destreza en el procesado de texto y no tener ninguna de las limitaciones de los otros lenguajes de script.
- *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.
- *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.

- *Workspace*: (*espacio de trabajo*) sección de un sistema de archivos donde los archivos de interés (para una dada tarea) están localizados. Actúa como un ambiente aislado donde trabajar mientras dura la tarea.

ÍNDICE DE ABREVIATURAS Y/O SIGLAS

- *CAB*: Centro Atómico Bariloche.
- *IB*: Instituto Balseiro.
- *IT*: (Information technology) estudio, diseño, desarrollo, implementación, soporte o manejo de sistemas de información basados en computadora.
- *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
- *URL*: (*Uniform Resource Locator*), localizador uniforme de recurso. Es una secuencia de caracteres, de acuerdo a un formato estándar, que se usa para nombrar recursos, como documentos e imágenes en Internet, por su localización.

ÍNDICE DE CONTENIDOS

Introducción	1
Objetivos del RI del <i>CAB-IB</i>	4
Funcionalidad del producto	6
Usuarios	7
Visitantes	7
Registrados	7
Editores	8
Administradores	8
Limitar privilegios de visitantes	8
Proceso de autoarchivo y publicación	9
Proceso de carga	11
Formatos de archivos soportados	12
Establecer fecha de embargo	13
Autocompletado	14
Otras facilidades de carga	14
Navegación del repositorio y búsqueda de objetos	15
Plataforma de interoperatividad	17
Administración de metadatos	18
Soporte para preservación	20
Exportación de objetos complejos	20
Módulo de historia	20
Declaración de derechos de preservación	21
Soporte OAI-PMH	22
Reusabilidad de registros actuales	23
Obtención de estadísticas	25
Herramientas externas para gestión de estadísticas	25
Otras características del producto	28
Longevidad	29

Soporte técnico	29
Comunidad de usuarios	29
Soporte pago	30
Entrenamiento	30
Documentación	30
Instalación	30
Backups	31
Conclusión	33
Anexo A: Establecimiento de fecha de embargo	35
Bibliografía	40
Otras fuentes de información consultadas	39

INTRODUCCIÓN

El presente documento describe el proceso de evaluación de EPrints (Versión 3.1) como posible software para la implementación de un Repositorio Institucional de la producción intelectual local, presentando los interrogantes que surgieron después de realizar una demostración del producto al personal de la Biblioteca Leo Falicov.

EPrints es un software desarrollado en el seno del Open Citation Project dirigido por Stevan Harnad en la Universidad de Southampton (UK). Está diseñado con el objetivo de ser fácil, rápido de instalación y gratuito. EPrints se distribuye bajo la licencia GNU, lo cual significa que el código fuente es accesible y modificable por cualquier programador, con la condición que las modificaciones se hagan también accesibles públicamente.

Un repositorio EPrints es un conjunto de servicios ofrecidos a todos los miembros de una institución para acceso abierto, preservación, enseñanza, e-learning, e-publishing, compartir datos, colaboración, publicaciones, investigaciones, publicidad, etc.

EPrints ofrece un sitio de demostración en línea para probar el producto en <<http://demoprints3.eprints.org/>> [10 de Junio 2009].

Para tener una visión más amplia de la funcionalidad del producto, se realizó una demostración del mismo utilizando el sitio de prueba provisto por los desarrolladores.

De esta forma, con el EPrints instalado se pudo hacer una simulación completa de una carga, desde el punto de vista del personal de la Biblioteca.

En la demostración se probó:

- El ciclo completo de carga de un objeto, donde se vieron los metadatos que se pueden almacenar, algunos tipos de documentos que se pueden cargar (luego se probó la carga de los demás tipos de documentos que EPrints permite almacenar).
- Las actividades que puede realizar un autor y un editor.
- Las consultas que se pueden realizar, tanto simples como avanzadas.
- Se navegó por los browsers que se instalan por defecto.

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

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 EPrints 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 EPrints 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.

EPRINTS 3.1

FUNCIONALIDAD DEL PRODUCTO

USUARIOS

EPrints distingue cuatro tipos de usuarios, por orden de privilegios: visitantes, usuarios registrados, editores y administradores. Un usuario en un nivel de privilegio dispone de todas las características de su nivel, más las disponibles en los niveles inferiores (Así, por ejemplo, un editor dispone de todas las facilidades de las que dispone un usuario registrado más algunas propias de su nivel de privilegio).

Visitantes

Son los usuarios con menos privilegios. Se les permite navegar el repositorio, realizar búsquedas, visualizar y descargar objetos públicos y subscribirse a las novedades del repositorio mediante RSS.

Usuarios registrados

Todo usuario nuevo que necesite más privilegios que un visitante debe crear una cuenta de usuario en el repositorio. El proceso de registración consiste en completar un formulario con datos personales y elección de credenciales de acceso y no finaliza hasta tanto no se active la cuenta mediante un mail automático enviado por el sistema a la casilla del usuario.

Para hacer valer sus privilegios un usuario registrado debe presentar sus credenciales al repositorio mediante la interfaz web. Además, EPrints permite a los usuarios reestablecer sus contraseñas en caso de olvidarlas.

Un usuario registrado puede realizar cargas de objetos (en ingles *deposit* – ver “Proceso de carga” más adelante) y salvar sus búsquedas. Las búsquedas se salvan dando un nombre descriptivo, la opción de recibir un mail con determinada frecuencia y la posibilidad de dejar la búsqueda públicamente visible o hacerla

privada.

Editores

Responsables de revisar los objetos cargados por los usuarios registrados, pudiendo editarlos, devolverlos al autor, eliminarlos o agregarlos al repositorio. La devolución o eliminación incluye una notificación para el autor.

Administradores

Responsables de la administración del repositorio. Dentro del alcance de este informe se evaluó de entre las posibles tareas administrativas solo la posibilidad de administrar metadatos mediante la interfaz web (ver “*Administración de metadatos*” más adelante).

Limitar privilegios de visitantes

El sistema permite deshabilitar la posibilidad de que los visitantes se registren como usuarios vía web, quedando así la responsabilidad de la creación de cuentas de usuarios en manos del administrador.

Se puede también cerrar completamente el repositorio para visitantes, quitando así la posibilidad de realizar acción alguna sobre el repositorio sin previa identificación como usuario registrado. <http://wiki.eprints.org/w/Login-Only_Repository> [17 de Junio 2009]

PROCESO DE AUTOARCHIVO Y PUBLICACIÓN

EPrints permite separar la visión externa del archivo de objetos de los procedimientos internos para agregarlos y/o modificarlos:

- Cada usuario almacena objetos incompletos en su *workspace*.
- Cuando el usuario considera que el objeto está listo, mueve el objeto finalizado al *buffer*.
- Una vez validado, un editor mueve el objeto al *archivo público de objetos*.
- En circunstancias excepcionales, un editor puede mover objetos desde el *archivo público de objetos* al buffer de eliminación.

Un repositorio EPrints aísla el *archivo público de objetos* de los nuevos objetos que están siendo cargados al asignar a cada usuario registrado su propio *workspace*. Este es el área donde los nuevos objetos son guardados temporalmente hasta que todos los metadatos han sido completados y el usuario decide llevar a cabo la carga.

Cuando el usuario ordena llevar a cabo la carga de un objeto, el mismo es movido desde el *workspace* del usuario al *buffer* del sistema. Esta es el área del archivo de objetos donde los objetos recientemente creados son almacenados en espera de revisión editorial. EPrints permite al administrador deshabilitar este *buffer*, de forma tal que los objetos sean cargados directamente por los autores en el *archivo público de objetos*. Luego de revisar un objeto, los editores pueden moverlo al *archivo público* o bien regresarlo al *workspace* del usuario para corrección.

Otro privilegio de un editor es la habilidad de mover un objeto al *buffer de eliminación*. Cuando un objeto es borrado, no desaparece completamente, sino

que es movido al *buffer de eliminación*, dejando así registro de su existencia. Una vez allí, no aparecerá en los resultados de las búsquedas, y, si se accede directamente a su página de referencia, se mostrará como “*borrado*”, junto con un puntero a cualquier versión posterior.

PROCESO DE CARGA

La carga de objetos esta disponible para usuarios registrados, editores y administradores. Se realiza mediante una interfaz web. EPrints propone dos métodos para realizar una carga: importar información desde fuentes externas a objetos EPrints o realizar una carga manual del objeto.

La importación puede hacerse copiando y pegando los registros a importar en un campo del formulario de importación o subiendo un archivo con el contenido de los registros a importar y seleccionado el formato de importación de entre los disponibles (ver "*Plataforma de interoperabilidad*" más adelante).

La creación de un nuevo objeto puede hacerse mediante una interfaz dividida en etapas:

- Etapa 1: seleccionar tipo de objeto.
- Etapa 2: subir el(los) documento(s). Aquí puede elegirse el formato del archivo (ver *formatos de archivos soportados* más adelante), entrar una descripción, definir visibilidad, seleccionar el tipo de licencia e ingresar una fecha de embargo (ver *establecer fecha de embargo* más adelante).
- Etapa 3: ingreso de metadatos. Se debe ingresar tantos datos sobre el objeto como sea posible. Los conjuntos de metadatos varían según el tipo de documento seleccionado en la etapa 1, y pueden ser personalizados (ver *Administración de metadatos* más adelante). Cada conjunto de metadatos contiene campos requeridos, sin los cuales el sistema no permite llevar a cabo la carga. En esta etapa resulta de especial utilidad la función de autocompletado (ver *autocompletado* más adelante).
- Etapa 4: seleccionar que temas trata el objeto. El conjunto de temas

ofrecidos para seleccionar es personalizable por el administrador.

- Etapa 5: Finalizar la carga. Permite completar el proceso de carga si no existen campos requeridos faltantes.

Formatos de archivos soportados

La etapa 2 del proceso de carga permite seleccionar el formato del(los) archivo(s) a subir de entre los siguientes:

Textos, presentaciones:

HTML, XML, PDF, Postscript, Plain Text, Rich Text (RTF), Microsoft PowerPoint, Microsoft Excel, Microsoft Word.

Imágenes:

JPEG, PNG, GIF, BMP, TIFF.

Video:

MPEG, QuickTime, AVI, mov, WMV, mp4.

Archivos:

BZ2, TGZ, ZIP.

Audio:

WAV, MP3, OGG, FLAC, WMA.

La selección del formato del archivo se hace solo a fin declarativo. EPrints no impone restricciones sobre el archivo a subir basándose en esta información. Nada impide al usuario subir un video y seleccionar como formato PDF por ejemplo. Así, EPrints permite en la práctica subir cualquier tipo de archivo, siendo el usuario responsable entonces de seleccionar la opción “otro” (en inglés “others”) de la lista de opciones Formato (en inglés “Format”) cuando el formato del archivo subido no se correspondiese con ninguna de las opciones provistas.

EPrints no impone límite alguno sobre el tamaño de los archivos que se pueden subir, quedando esto determinado entonces únicamente por el sistema de

archivos subyacente y/o el servidor web.

Establecer fecha de embargo

Un documento puede estar cargado en el repositorio pero ser inaccesible por los usuarios por un período de tiempo determinado por el(los) autor(es). Una vez finalizado este período, el sistema libera el documento para poder ser accedido. También se puede poner un período interminable para que el documento siempre tenga que ser solicitado a el(los) autor(es). Para liberar estos documentos, EPrints posee un *script* que en la demo no se está ejecutando, esto se consultó al equipo de desarrollo directamente.

“... lift_embargos: This script should be run once a day. It removes security on documents which embargo date has passed...”

Mientras la publicación no está habilitada para ser leída o descargada libremente, se habilita un botón de pedido del objeto, que al ser presionado le envía un mail al autor con la solicitud del documento. Luego el autor decide si se lo envía al solicitante o no.

Esta opción se utiliza, por ejemplo, cuando el autor publica en una revista que por un período de tiempo no le permite publicar el documento en otro lado. También se puede utilizar si sólo se desea brindar el documento a pedido explícito y así conocer quien lo solicita.

En el caso en el que el autor publicó en una revista que no le permite liberar nunca el documento, es mejor subir al repositorio sólo el *abstract* o un capítulo, según lo que se le permita.

El procedimiento de uso de esta funcionalidad se describe en el Anexo A.

Autocompletado

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

El primer campo que soporta esta característica es Título (Title), luego de ingresar unas letras en el campo correspondiente, se despliega una lista de objetos en el repositorio con títulos similares. Se puede hacer click en los títulos de los objetos en la lista para ver la página de resumen del objeto correspondiente y asegurarnos de no estar cargando un objeto ya existente.

Esta función esta disponible también para el campo Autor (Creador), cuando algunas letras del nombre del autor han sido ingresadas, se despliega una lista de posibles autores. Un click sobre los datos de un autor en la lista llenará los datos del campo con los datos contenidos en el repositorio. De la misma forma, el autocompletado es usado para entrar datos de una publicación, al hacer click sobre una publicación de la lista desplegable se completa el resto del título, el ISSN y la editora.

Otras facilidades de carga

Permite subir actualizaciones de un documento en el mismo conjunto de metadatos y se crea un árbol de versiones del documento permitiendo el acceso a cualquiera de ellos.

Los metadatos ingresados para un documento se pueden reutilizar para la carga rápida de otro documento con iguales características. Se utiliza la operación de copia o la de clonación que ofrece el software.

Para evitar duplicación de objetos el sistema alerta al usuario en caso de que un objeto existente se vea similar al que se esta ingresando.

NAVEGACIÓN DEL REPOSITORIO Y BÚSQUEDA DE OBJETOS

EPrints provee diversas maneras en las que un repositorio puede ser dividido en diferentes colecciones o ser navegado según diferentes perspectivas. Por defecto, provee dos vistas para que los usuarios naveguen el contenido del repositorio: por año de publicación o por tema que trata el objeto (ver Etapa 4 de “*Proceso de carga*” más arriba). El software permite definir vistas personalizadas por cualquier metadato.

A su vez, provee una variedad de métodos de búsqueda:

- **Búsqueda avanzada:** Se puede ingresar términos de búsqueda para cada metadato disponible. Permite por cada término definir si la frase ingresada debe coincidir en su totalidad o solo en parte con el resultado.
- **Búsqueda simple:** permite ingresar uno o más términos de búsqueda. También permite elegir si se desea buscar coincidencias exactas o parciales. Permite seleccionar además el criterio para ordenar los resultados.

Al igual que para la navegación, el repositorio puede ser configurado para ofrecer búsquedas según cualquier combinación de metadatos.

EPrints permite además, a usuarios registrados, guardar los perfiles de búsqueda para mantenerse actualizados respecto de sus resultados. Ofrece incluso la posibilidad de suscribirse para ser notificado cuando un nuevo objeto que coincide con los criterios de la búsqueda es cargado en el repositorio.

Los resultados de las búsquedas pueden ser exportados en una variedad de formatos. Los formatos soportados por defecto son ASCII Citations, BibTeX, Dublin Core, EP3 XML, EndNote, HTML Citations, METS, MODS, OpenURL

ContextObject, Refer and Reference Manager. Soporte para otros formatos o servicios web puede agregarse mediante *plugins*.

PLATAFORMA DE INTEROPERATIVIDAD

De los desarrolladores sobre la interoperabilidad: “...El sistema fue diseñado para ser extensible, por lo que, en caso de necesitar importar (exportar) datos desde (hacia) algún formato no soportado nativamente se puede extender la funcionalidad del sistema desarrollando un plugin o instalando alguno de los existentes...”.

EPrints provee un registro de *plugins* en su repositorio para desarrolladores <<http://files.eprints.org/>> [10 de Junio 2009].

Nativamente soporta:

Importación de registros con formato

XML, BibTeX, PubMed ID, PubMed XML, DOI (via CrossRef).

Exportación de registros EPrints hacia

ASCII Citation, BibTeX, Dublin Core, DIDL, EP3 XML, EndNote, EPrints Application Profile, HTML Citation, JSON, METS, MODS, OpenURL ContextObject, OpenURL ContextObject in Span, Refer, Reference Manager, Simple metadata.

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?

Con respecto a estas cuestiones EPrints ofrece:

De los desarrolladores sobre la administración de metadatos en EPrints: “... en la versión full se pueden crear, modificar o eliminar campos para metadatos excepto los del núcleo que son requeridos por el sistema...”.

La interface web provee a usuarios con privilegios administrativos la posibilidad de agregar metadatos. La recolección de los metadatos agregados por esta vía se incluye en el proceso de carga (ver “*Proceso de carga*” más arriba) de una manera predefinida (se agrega una nueva etapa titulada “Misc” entre las etapas 4 y 5).

Del capítulo titulado “*Configuring the archive metadata*” de la documentación oficial se observó como los metadatos e incluso las interfaces para ingresarlos pueden ser altamente personalizados a las necesidades del repositorio mediante la edición de archivos de configuración. Se puede emplear cualquier esquema de metadatos, no solo para los objetos sino también para los autores.

Cabe mencionar que los objetos individuales poseen 56 líneas de metadatos ocultas para *Google Scholar*.

SOPORTE PARA PRESERVACIÓN

Las características de soporte de preservación actualmente provistas por EPrints han sido desarrolladas en conjunto con el proyecto Preserv <<http://preserv.eprints.org/>> [11 de Junio 2009].

Las funciones están diseñadas para permitir a un repositorio EPrints proveer soporte para preservación a través de un proveedor de servicio de preservación. Las funciones claves a tal fin son:

Exportación de objetos complejos

Existen varias maneras de diseminar objetos almacenados en un repositorio EPrints (ver *“Plataforma de interoperatividad”* más atrás), dependiendo de si el pedido proviene de una persona o una maquina (e.g. un robot de un motor de búsqueda) y de que servicio se solicita.

Si el pedido de exportar un objeto desde un repositorio proviene de un proveedor de servicio de preservación, necesitará obtener todos los archivos asociados con el objeto. Para soportar más eficientemente la transferencia de objetos complejos 2 formatos que pueden ser usados para diseminar objetos para preservación digital son METS y DIDL.

EPrints cuenta con plugins para exportar en METS y DIDL que permiten al proveedor de servicio descargar todos los archivos y metadatos que forman un objeto.

Módulo de historia

EPrints provee una función “historia” que documenta todos los cambios hechos a un registro desde su creación. Actualmente esta función es usada para

proveer un registro de auditoria, pero a medida que los servicios de preservación vayan siendo desarrollados necesitarán eventualmente modificar el contenido de los repositorios (e.g. migrando formatos de archivos por ejemplo) y tales decisiones deben ser registradas para dar soporte a decisiones de preservación futuras.

Declaración de derechos de preservación

El proveedor de servicio de preservación necesita conocer los derechos que tiene para copiar y/o actuar sobre los contenidos de un objeto. Para este fin, se dispone en la etapa 2 del proceso de carga (ver “*Proceso de carga*” más arriba) de la lista desplegable Licencia (License) que permite al usuario explicitar la licencia de uso del objeto cargado. Si una licencia explícita fuese necesaria para preservación (e.g. dar derecho a un tercero de guardar y actuar sobre el material, por ejemplo, migrando formato) puede ser agregada fácilmente.

Para la próxima versión del software (3.2) se anuncia la inclusión de un *framework* sobre el cual servicios de preservación activos tales como clasificación de archivos, análisis de riesgo y herramientas de migración puedan ser aplicados. A diferencia de la versión actual, no solo se pensarán escenarios dejando la implementación al usuario, sino que se construirá un *framework* para la implementación de servicios existentes provistos por el proyecto Preserv2 <<http://preserv.org.uk/>> [11 de Junio 2009]. Más información al respecto puede encontrarse en la *wiki* de Preserv2 en el artículo titulado EPrintsPreservation <<http://wiki.preserv.org.uk/index.php/EPrintsPreservation>> [11 de Junio 2009].

SOPORTE OAI-PMH

EPrints provee soporte para los protocolos de *harvesting* OAI-PMH. El mismo debe ser configurado mediante la edición de archivos *perl*.

Se dispone de documentación que explica exhaustivamente cómo configurar el soporte OAI para un repositorio EPrints. <<http://wiki.eprints.org/w/OAI>> [12 de Junio 2009].

REUSABILIDAD DE REGISTROS ACTUALES

EPrints permite importar registros individuales o en bloque en cualquiera de los formatos de importación soportados nativamente (descritos en la sección “*Plataforma de interoperatividad*” más arriba) e incluso extender el conjunto de formatos de importación mediante plugins.

Se probó la importación desde un archivo de un bloque de 992 registros en formato BibTeX. El proceso resultó en la creación de un objeto EPrint por cada registro BibTeX, y la ubicación de los mismos en el *workspace* del usuario que realizó la importación (para más detalles sobre como EPrints maneja el proceso de autoarchivo, publicación y carga referirse a las secciones correspondientes más arriba). Para completar la simulación de lo que implicaría dejar los registros disponibles en el repositorio, usando un usuario editor, se realizó una búsqueda que arrojó los 992 registros recién importados (Admin. - “Search ítems” en la interfaz de usuario) y se aprovechó la facilidad de edición en bloque “Batch Edit” para cambiar el estado de los ítems importados de “User Workarea” a “Live Archive”.

De la prueba se rescata además como una característica útil la posibilidad de analizar el bloque de registros a importar previo a llevar a cabo la carga (botón “*Test run*” de la interfaz de importación). De esta forma, EPrints alerta sobre problemas en el conjunto de datos (errores o advertencias) que no le permitirían cargar algunos registros o campos y permite así depurar el bloque hasta que conforme con el formato esperado. También muestra errores y/o advertencias, si las hubiese, al momento de hacer efectiva la carga.

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

- Adaptar los registros existentes a alguno de los formatos de importación soportados nativamente.

- Adaptar los registros existentes a alguno de los formatos de importación para los cuales se dispone de *plugins*. Requiere además instalar y poner en marcha el *plugin* correspondiente.
- Desarrollar un *plugin* que permita importar registros en algún formato al cual sea conveniente adaptar los registros de los que se dispone.

OBTENCIÓN DE ESTADÍSTICAS

EPrints almacena todos los eventos que se realizan sobre los objetos pero aún no provee nativamente ningún proceso de recuperación y presentación de esta información.

A la fecha, se encontraron los siguientes proyectos de terceros para procesamiento de estadísticas sobre repositorios EPrints:

University of Tasmania statistics package.

<<http://eprints.utas.edu.au/262/>> [23 de Junio de 2009]

IRStats: Interoperable Repository Statistics.

<<http://trac.eprints.org/projects/irstats/wiki>> [23 de Junio 2009]

Simple usage statistics (using analog).

<<http://files.eprints.org/259/>> [23 de Junio de 2009]

Herramientas externas para gestión de estadísticas

Durante la presentación de las conclusiones de este informe al personal de la biblioteca se recogió, entre otras, la sugerencia de evaluar la posibilidad de gestionar el manejo de estadísticas a través de alguna herramienta externa al repositorio, en particular *Google Analytics (GA)*.

A continuación se listan las diferentes vías usadas para investigar al respecto, seguidas de los comentarios correspondientes a cada una:

- Consulta a la lista de correo *eprints-tech*:

Se consultó a los participantes de la lista respecto de la posibilidad de

integración de GA con EPrints y se pidió a aquellos que hubiesen experimentado con esta y/u otras herramientas externas que compartan su experiencia. Aquí la versión web del correo enviado <www.eprints.org/tech.php/thread-11313.html>.

En base a las respuestas se concluye que es posible extraer estadísticas del repositorio mediante GA: algunos repositorios actualmente lo hacen e incluso uno de los participantes acompañó su respuesta con un gráfico que muestra los tipos de estadísticas que están extrayendo y de que forma pueden presentarse.

Alguna de las respuestas subrayan como limitaciones de GA en comparación con la utilización de *plugins* la imposibilidad de generar estadísticas que tengan que ver con la organización interna del sitio, tales como agrupamiento por categorías. Esto se debe a que GA no tiene forma de conocer la organización jerárquica del sitio sobre el cual recolecta estadísticas.

- Búsqueda en la web de recursos sobre integración de RI y GA

Se analizaron diferentes opiniones de la comunidad de usuarios e implementadores de RI respecto de la posibilidad de utilización de GA como recolector de estadísticas llegando a las siguientes conclusiones:

La decisión entre utilizar una herramienta interna o una externa para gestión de estadísticas debe tomarse en función del tipo de estadísticas que se desean recolectar.

GA es una buena opción para determinar cómo los usuarios están accediendo al sitio, desde donde y qué páginas están visitando (información frecuentemente de mayor interés para los administradores del repositorio que para los investigadores que depositan sus trabajos), pero no es fácil ni práctico para llevar un registro de descargas de archivos (lo cual según experiencias ajenas suele ser muy requerido por los autores).

Otra capacidad que suele ser frecuentemente requerida es la posibilidad de agrupar estadísticas según la jerarquía interna del repositorio (10 archivos más descargados de una dada materia por ejemplo). Tal posibilidad no está al alcance de GA por los motivos que ya se mencionaron más arriba cuando se analizaron las respuestas a la consulta a la lista de correo.

En conclusión, las herramientas de gestión de estadísticas externas al repositorio deben ser tenidas en cuenta como una opción que puede utilizarse en repositorios que no provean esta capacidad o incluso en combinación con las capacidades provistas. Dependiendo de las estadísticas que se desee obtener, podría incluso usarse una u otra opción de manera exclusiva.

EPRINTS 3.1

OTRAS CARACTERÍSTICAS DEL PRODUCTO

LONGEVIDAD

Primera versión: 1.0 de Noviembre de 2000. Versión actual: 3.1 de Septiembre de 2008. La versión 3.2 se encuentra en desarrollo pero aún no hay fechas estimadas de lanzamiento publicadas. [12 de Junio 2009].

269 repositorios registrados como usuarios de EPrints. <<http://www.eprints.org/software/archives/>> [12 de junio 2009]

SOPORTE TÉCNICO

Comunidad de usuarios

De una revisión de documentación de experiencias de terceros implementando repositorios con EPrints se recogen en todos los casos buenos comentarios respecto del soporte y la posibilidad de despejar dudas por diversas vías: contacto directo con los desarrolladores, documentación de ayuda y/o consultas a la comunidad de usuarios.

En los casos en que para nuestra propia evaluación se requirió consultar al equipo de soporte técnico, se recibieron respuestas con no más de un día de demora.

Producto en desarrollo activo, *Parches* y actualizaciones disponibles regularmente. La comunidad de usuarios remarca que pueden ser aplicados fácilmente.

Dispone además de una lista de correo para discusiones técnicas sobre el uso y desarrollo del producto. Actualmente cuenta con 350 participantes y la suscripción es gratuita. <<http://mailman.ecs.soton.ac.uk/mailman/roster/eprints-tech>> [12 de junio 2009].

Soporte pago

EPrints ofrece servicios de soporte pagos incluyendo *hosting*, entrenamiento y consultoría, provistos por el mismo equipo de desarrollo de EPrints. <<http://www.eprints.org/services/>> [12 de Junio 2009].

Entrenamiento

EPrints reúne en su sitio oficial una colección de materiales de entrenamiento y tutoriales dirigidos a diferentes audiencias, desde usuarios finales hasta programadores y administradores con la intención de servir como guía para las tareas que cada uno deba llevar a cabo. <<http://www.eprints.org/software/training/>> [16 de Junio 2009].

Además, tiene un equipo dedicado a entrenamiento (EPrints training) que organiza cursos de entrenamiento presenciales. A la fecha, el último curso oficial registrado fue dado el 23 de Septiembre de 2008. <<http://www.eprints.org/services/training/>> [16 de Junio 2009].

DOCUMENTACIÓN

Disponible online en inglés.

Documentación técnica centralizada en

<http://wiki.eprints.org/w/Main_Page> [12 de Junio 2009].

Desde allí puede accederse a manuales de usuarios, preguntas frecuentes, tutoriales, etc.

INSTALACIÓN

EPrints fue desarrollado en Linux RedHat (tanto Fedora Core como Enterprise), pero puede ser usado y ha sido probado en un gran número de distribuciones Linux y otros sistemas tipo Unix incluyendo OS-X. También corre en Windows Vista y XP.

A una consulta sobre la plataforma más conveniente el equipo técnico contestó:

"...We have a port of EPrints 3.0 to Windows, but it's currently unfunded so we can't provide very much support for it.... "

Plataformas de instalación recomendadas:

- <http://wiki.eprints.org/w/Recommended_Platforms> [17 de Junio 2009]

Software adicional requerido:

- <http://wiki.eprints.org/w/Required_software> [17 de Junio de 2009]

Configuraciones / personalizaciones post instalación:

- <http://wiki.eprints.org/w/Web_Interface_Style_and_Images> [17 de Junio 2009]
- <<http://www.eprints.org/software/training/users/trieste/TriesteDay3-Config.pdf>> [25 de Junio de 2009]
- <<http://sdu.ictp.it/eya/openaccess08.php>> [25 de Junio de 2009]

BACKUPS

El proceso de obtención y almacenamiento de una copia de seguridad del repositorio según descrito en <<http://wiki.eprints.org/w/Backups>> [17 de Junio

2009] no resulta de mayor complejidad.

CONCLUSIÓN

EPrints reúne características suficientes para cumplir con los objetivos del repositorio institucional del *CAB-IB*:

Las capacidades de gestión de cuentas de usuarios cumplen las expectativas y cubren las necesidades expuestas.

El modelo de autoarchivo y publicación resulta adecuado para el flujo de trabajo pretendido.

De los procesos de navegación del repositorio, carga y recuperación de objetos se destaca: flexibilidad, simplicidad, tipos de objetos soportados.

El soporte para interoperabilidad y preservación a largo plazo es satisfactorio, y promete mejoras para futuras versiones.

Los registros con los que actualmente se cuenta pueden ser integrados al repositorio de varias maneras.

En general, se destaca como positivo la flexibilidad en la configuración. El repositorio es altamente configurable, propiciando la personalización inicial y la adaptabilidad a futuros requerimientos que pudieran surgir.

Durante el relevamiento se observaron además otras características, que si bien no se corresponden directamente con alguno de los objetivos expuestos, resultan beneficiosas: gran comunidad de usuarios, puesta en funcionamiento relativamente rápida, desarrollo en continua evolución, buena disponibilidad de soporte técnico entre otras.

Experiencias de terceros en el uso de EPrints destacan en general la

simplicidad, estabilidad, y la escasa necesidad de mantenimiento. Está ampliamente visto como una solución ideal para una implementación inicial en una institución con recursos financieros y soporte *IT* limitado.

EPRINTS 3.1

ANEXO A: ESTABLECIMIENTO DE FECHA DE EMBARGO

DETALLE DE PRUEBA DE FUNCIONALIDAD: ESTABLECIMIENTO DE FECHA DE EMBARGO

Se documenta a continuación la secuencia de pasos llevados a cabo para evaluar el funcionamiento del establecimiento de fechas de embargo sobre objetos cargados. Para más detalle sobre las características y utilidad de esta funcionalidad ver *“Establecer fecha de embargo”* en la sección *“Proceso de carga”* del documento al cual este anexo acompaña.

Si bien en general el objetivo de esta evaluación no consiste en documentar los procesos de prueba en si mismos sino principalmente sus resultados, se hace una excepción en este caso debido a que la prueba de esta característica llevo a relevar información que al momento de realizarla no se encontró disponible en ninguna de las fuentes de consulta.

1)

En el demo site <<http://demoprints3.eprints.org/>> [10 de Junio 2009] se creó un nuevo objeto conteniendo un documento en formato .rtf, estableciendo fecha de embargo *“2008-Oct-28”* y seleccionando en la opción *“Visible To:”* el valor *“Anyone”*.

2)

Al avanzar de Etapa se obtuvo el siguiente error:

“Rich Text (RTF) (Prueba de fecha de expiración y rtf): An embargo date has been specified, but the document is visible to anyone. Check the document visibility settings.”

3)

Se consultó la ayuda provista por la interfaz web para el campo Fecha de embargo (*“Embargo expiry date”*):

“The date that a publisher- or sponsor-imposed embargo expires. On and after this date, this document will be made publicly accessible.

Example: 2006 - all documents made publicly accessible on 1st January 2006”

La misma ayuda para el campo Visibilidad (“Visible to”) indica:

“Please indicate the required "security level" of this document. Who is allowed to download it? If you set this to anything other than "anyone" then consider adding a contact email for this eprint (on the next page), this will allow users to use a web form to request a copy of the document, and you can decide on a case-by-case basis.”

4)

Se modificó el valor de la opción *“Visible to”* por *“Registered users only”*, y en el formulario siguiente se ingresó una dirección de mail para que los que quieran descargar el documento llenen el formulario y llegue allí el pedido para poder autorizar la descarga.

Se logró publicar el nuevo objeto.

5)

Al día siguiente (debido a que el sitio de prueba demora un día en actualizar sus datos con los nuevos objetos) se puede hacer la búsqueda de lo registrado en días anteriores:

Desde la demo y sin haberse identificado como usuario del repositorio, muestra el título del documento de la siguiente forma:

Rich Text (RTF) (Prueba visibilidad y fecha exp) - Published Version
Restricted to Repository staff only until 28 October 2008.2600b

Presentando un botón “*Request a copy*” que lleva a un formulario donde hay que completar una dirección de mail a donde se desea que se envíe la copia (obligatorio) y un motivo por el cual se pide la misma (no obligatorio).

Se completó el formulario con el pedido.

Se recibió la transcripción del pedido en la casilla de mail que se estableció en 4).

El creador del documento recibió un mail con asunto: *Request for “título del documento”* conteniendo:

1. El motivo del pedido expresado en el formulario de pedido.
2. La opción para aceptar el pedido.
3. La opción para rechazar el pedido.

Al aceptar el pedido redirige a la página de *Login* del sitio de prueba.

6)

Al identificarse como creador del documento puede verse lo siguiente:

“Confirm response

You have chosen to **accept** the request for the following item:

Carina, P [Prueba visibilidad y expiración.](#) .

[Rich Text \(RTF\) \(Prueba visibilidad y fecha exp\)](#) - Published Version

Restricted to Repository staff only until 28 October 2008.

2600b

An acceptance notice will be emailed to the requester with the requested

document attached.

Tick here to make this document publicly available from now on.”

7)

Presionando el botón “*Confirm*” se obtiene:

“Thank you. Your response has been sent.”

8)

En la dirección de mail que se ingresó en el pedido del documento, se recibió un mail con el documento adjuntado.

9)

Se realizó un nuevo pedido del documento para probar la opción de rechazarlo.

10)

Llegaron las mismas notificaciones, una al autor del documento y otra al usuario que hace el pedido.

11)

El autor rechaza el pedido, pudiendo ingresar un motivo de rechazo

12)

El usuario que hace el pedido recibe la notificación con el aviso de rechazo.

BIBLIOGRAFÍA

EPrints 3 Documentation. <http://wiki.eprints.org/w/EPrints_Manual> [18 de Junio 2009].

EPrints Community. <<http://www.eprints.org/community/>> [18 de Junio 2009].

EPrints Files. <<http://files.eprints.org/>> [18 de Junio de 2009].

EPrints Handbook. <<http://www.eprints.org/documentation/handbook/>> [18 de Junio 2009].

EPrints Training Collection. <<http://www.eprints.org/software/training/>> [18 de Junio 2009].

EPrints Wiki. <http://wiki.eprints.org/w/Main_Page> [12 de Junio 2009].

INFORMAL comparison of some institutional repository solutions. Neil Godfrey. <<http://metallogger.wordpress.com/2008/10/19/comparing-some-institutional-repository-solutions/>> [18 de Junio de 2009].

Open Source Maturity Model. <<http://www.navicasoft.com/pages/osmm.htm>> [18 de Junio 2009].

Preserv Project. <<http://preserv.eprints.org/>> [11 de Junio 2009].

Repositories Support Project. <<http://www.rsp.ac.uk/index>> [18 de Junio 2009].

OTRAS FUENTES DE INFORMACIÓN CONSULTADAS

Workshop on Using Open Access Models for Science Dissemination, 7- 16 July 2008 - ICTP Miramare, Trieste, Italia. <<http://sdu.ictp.it/openaccess/>> [18 de Junio 2009]

Consultas particulares enviadas a:

- Christopher Gutteridge, Web Projects Manager, School of Electronics and Computer Science, University of Southampton. <<http://www.ecs.soton.ac.uk/people/cjg>> [18 de Junio 2009].
- Leslie Carr, EPrints Services Director, EPrints Open Source Technical Director, Repository Manager, Web Science Researcher, University of Southampton. <<http://www.ecs.soton.ac.uk/people/lac>> [18 de Junio 2009].
- Lista de correo "*EPrints-tech*": <<http://mailman.ecs.soton.ac.uk/mailman/roster/eprints-tech>> [12 de junio 2009].