

Proyecto Biblioteca Digital de la Universidad Nacional de Cuyo

Isabel Piñeiro, Horacio Degiorgi, Pilar Piñeyrua, Cecilia Pincolini, Juan
Groisman y Adrián Mendez.

Sistema Integrado de Documentación (SID), Universidad Nacional de Cuyo,
Centro Universitario, Parque "General San Martín", M5502JMA Mendoza,
Argentina. bdigital@uncu.edu.ar

Resumen: La Biblioteca Digital (repositorio institucional de la UNCuyo) es un proyecto innovador ya que administra varios formatos de documentos: audiovisuales, sonoros y textuales (libros, tesis, artículos, publicaciones periódicas). Para gestionarla, se evaluaron distintos softwares disponibles en el 2006. Se constató que ninguno permitía la administración del contenido con diversidad de formatos (video, audio y texto), carga descentralizada, relación jerárquica de objetos (revista, número y artículo). Se decidió entonces, comenzar un desarrollo propio basado en software libre. El software diseñado (BDuncu) permite el manejo y descripción de objetos digitales a través de plantillas de metadatos que respetan estándares internacionales (Dublin Core, Mods). Esto permite recuperar el contenido desde el propio sitio o desde otros metabuscadores que toman la información mediante el harvesting a través del protocolo OAI-PMH.

BDuncu fue desarrollado sobre plataforma Web en lenguaje PHP5, con motor de bases de datos Postgresql. Se han realizado múltiples actualizaciones y mejoras en las opciones de recuperación y presentación de la información. Dispone de un buscador avanzado y un lenguaje de plantillas propio para presentar los metadatos asociados a los objetos digitales. Utiliza librerías AJAX (Prototype y Scriptaculous) asegurando el soporte en distintos navegadores. Se está trabajando para ser usado en otras instituciones.

Abstract: The UNCuyo's institutional repository "Biblioteca Digital" -Digital Library- it's an innovative project for it supports several document formats: sound, video and text (books, theses, reviews, research journals articles). Different repository management softwares were evaluated in 2006 by our work group. After advertising that none of them allowed this video-audio-text handling, distributed administration and hierarchical object relationship (such as journal, number, article), we decided to start our own free software development. Bduncu, the result of our work, allows Digital Object handling and description through international standard (Dublin Core, MODS) metadata formats.

This allows the content recovery from both our internal Search Engine and external Metasearch Engines that access our metadata by harvesting through the OAI PMH protocol. Bduncu is a web-based platform in PHP5 language, with a Postgresql DataBase Engine. Several updates and upgrades in data recovery and visualization were performed. It has an advanced Search Engine and a self-developed template language for metadata visualization. Uses AJAX Libraries (Prototype y Scriptaculous) assuring navigation through different browsers. We are working for making this software usable in other institutions.

Palabras Claves: REPOSITORIO INSTITUCIONAL – BIBLIOTECA DIGITAL – OAI PMH – SOFTWARE LIBRE

1 Introducción

La necesidad de contar con un repositorio institucional surge como un objetivo de las universidades que tienden a generar un espacio común para alojar su producción científica y académica. Los repositorios institucionales brindan practicidad al proceso de descripción y almacenamiento de los documentos y a la vez proveen una interfaz de búsqueda y navegación para que el usuario pueda recuperarlos con facilidad.

La Biblioteca Digital de la Universidad Nacional de Cuyo, el repositorio institucional, es un proyecto innovador ya que está diseñada para administrar varios tipos de formato de documentos. De esta manera contiene objetos digitales de formatos audiovisual, sonoro y textual (libros, tesis, literatura gris, publicaciones periódicas, etc.)

El proyecto implica el trabajo coordinado de diversas áreas: una textual, una audiovisual/sonora y desarrollo informático e interfaz web.

Una vez formado el equipo y comenzados los primeros trabajos de digitalización y catalogación en 2006, surgió la necesidad de contar con un sistema que pudiera satisfacer nuestras necesidades como repositorio. Luego de probar distintas opciones de software disponible en ese momento, se constató que ninguno permitía la administración del contenido que se necesitaba: diversidad de formatos (video, audio y texto), carga descentralizada, relación jerárquica de objetos (revista, número y artículo). En ese momento se decidió comenzar un desarrollo propio basado en software libre.

El software diseñado permite el manejo de objetos digitales y su descripción a través de plantillas de metadatos que respetan los estándares internacionales (Dublin Core, Mods). Esto permite la recuperación del contenido desde el propio sitio o desde otros metabuscadores que toman la información de nuestra base de datos mediante el harvesting a través del protocolo OAI-PMH.

2 Situación-Problema u Oportunidad

1. No existía un repositorio institucional digital de la UNCuyo que pudiera albergar y difundir de manera sistemática toda la producción académico-científica producida por los miembros de la comunidad universitaria a través de sus diversas dependencias (centrales y de las facultades).

2. No existía un software libre de bibliotecas que sirviera para gestionar, realizar el tratamiento documental y visualizar las referencias bibliográficas y las manifestaciones completas de todos estos documentos producidos en diversos soportes (texto, audio y video). La Universidad contaba con material en distintos formatos: textos, audios y videos.
3. No era posible participar de repositorios internacionales para mostrar las investigaciones y la producción intelectual de la UNCuyo.

2.1 Oportunidad

1. Decisión política de contar con un repositorio institucional para mostrar la producción científico y académica de la UNCuyo, financiado a través del Fondo Universitario para el Desarrollo Nacional y Regional (FUNDAR), como un proyecto de I+D.
2. Surgimiento del movimiento Open Access a nivel internacional y adhesión de la UNCuyo a los conceptos involucrados en el libre acceso al conocimiento.
3. Existencia de un convenio con la Maison des Sciences de l'Homme para realizar acciones conjuntas con el objetivo de la difusión de la información científica y la generación de espacios comunes que permitieran un acceso universal.
4. Se contaba con material textual y audiovisual digitalizado y "nacido" digital en distintos formatos y repartido en las distintas Unidades Académicas de la UNCuyo.
5. Existía la demanda de autores institucionales por publicar y de usuarios por acceder al material, lo cual suponía un servicio no prestado por la Universidad y por el SID (Sistema Integrado de Documentación).
6. Se contaba con personal capacitado en lenguajes de desarrollo para generar un producto propio que resolviese las necesidades de software.
7. Existencia de personal capacitado en digitalización, gestión de ediciones electrónicas, edición de videos y uso de tecnologías web para la administración del proyecto.
8. Se contaba con personal con conocimientos en estándares internacionales bibliotecológicos.

3 Solución

Aprobación por parte del Consejo Superior (órgano máximo de la UNCuyo) del proyecto para desarrollar una Biblioteca Digital dentro de la estructura del Sistema Integrado de Documentación (Secretaría Académica, Rectorado).

La gestión de esta Biblioteca Digital ha incluido hasta ahora las siguientes actividades:

- Organización de un sistema para recibir los documentos desde las diferentes instancias productoras de contenido (Facultades, Secretarías, Direcciones, Editoriales, Institutos, Cátedras, Equipos de Investigación, etc.).

- Digitalización de las tesis, revistas, libros, informes y otros trabajos de investigación, imágenes, videos y audios, y/o recepción de documentos “nacidos digitales”.
- Tratamiento documental de cada objeto digital a través del Software desarrollado (carga de metadatos básicos, catalogación según reglas bibliotecológicas) para integrar otros sistemas y exportar registros a través del protocolo OAI.
- Publicación en Internet de una Biblioteca Digital a texto completo a través del software desarrollado (visualización que refleja la estructura jerárquica y la diversidad de documentos).
- Difusión de la Biblioteca Digital dentro y fuera de la UNCuyo, en circuitos académicos.

Para una mejor comprensión de lo descrito, ver gráfico de “Proceso de Obtención y Publicación de los Objetos Digitales”.

Se generó una plataforma con desarrollo propio que permitiese:

- Administración por diversos tipos de usuarios para la gestión, tratamiento documental y difusión de objetos digitales en distintos formatos respetando estándares internacionales. Esta plataforma permite no solamente albergar los objetos digitales, sino realizar el proceso documental (o bibliotecológico) necesario: generar un registro por cada objeto en lenguaje xml, con los metadatos que describen los distintos tipos de objeto de manera sistemática. Esto permite que otros buscadores, bibliotecas, colecciones y repositorios académicos internacionales puedan recuperar la información que constituye la Biblioteca Digital (“harvesting”).
- La estructura del repositorio se adecuó a la realidad de los documentos académicos. Por ejemplo: algunos de los objetos digitales están relacionados jerárquicamente en 2 o 3 niveles. Es el caso de la colección de una revista, sus números o volúmenes y sus artículos. O un documento audiovisual que se subdivide en bloques temáticos. O un Informe final de investigación con artículos y ponencias asociadas producidos por un equipo de investigadores.

Ver “*Descripción del software de la Biblioteca Digital de la UNCuyo*”.

4 Innovación e Inédito

Para la comunidad a la que sirve el proyecto supone una oportunidad de publicación, difusión y contactos inédita con 3 principales puntos a destacar:

1. Facilidad de la publicación: existe dentro del proyecto personal capacitado para procesar los materiales recibidos.
2. Gran difusión de lo publicado: la integración en repositorios internacionales nos permite asegurar una amplia difusión internacional.
3. Costo cero para los autores y editores de la Universidad para que su producción intelectual tenga visibilidad y se asegure su accesibilidad.

Es el primer y, hasta donde se ha averiguado, único software de biblioteca digital producido en la Argentina que permite cumplir con las siguientes funciones: protocolo OAI-PMH, publicación de objetos de audiovisuales y sonoros, integración con otros sistemas y exportación a XML de metadatos, carga realizada en plataforma Web (que permite trabajar en forma descentralizada)

Se ha trabajado en una interfaz de usuario para realizar las consultas utilizando un motor de búsqueda textual (sphinx-search) que permite ranqueo de resultados y búsquedas usando operadores booleanos y filtrado de resultados por tipo de objeto y tipo de documento, etc. Este motor de indexación full-text permite al usuario recuperar documentos en forma simple, como suele presentarse en un buscador de Internet.

5 Beneficiarios

Existen beneficiarios directos de la aplicación de la solución:

1. Universidad Nacional de Cuyo: cuenta desde el 2006 con un repositorio institucional donde publicar y difundir su producción científico-académica. Se integra con otras instituciones internacionales a través de los harvester (recolectores de registros OAI) que permiten la divulgación del conocimiento producido en la Universidad. A través de esta iniciativa la UNCuyo se suma al movimiento de Open Access, dándole visibilidad a su producción y contribuyendo a la democratización en el acceso a la información.
2. Los autores/editores/investigadores de la UNCuyo: disponen de un espacio donde publicar su producción intelectual y difundir sus investigaciones. Los editores pueden llegar a un público mayor sin la necesidad de invertir recursos propios.
3. Los estudiantes/investigadores/usuarios: poseen un repositorio local donde encontrar las investigaciones, revistas y lo producido en la Universidad. Cuentan además con el acceso libre a los documentos audiovisuales generados por la UNCuyo (programas de radio, conferencias, clases magistrales y programas de televisión).
4. Personal del equipo de la BDigital: formación continua en desarrollo de software, digitalización de objetos, definición de metadatos y administración de proyectos.

- Beneficiarios indirectos: comunidad científica nacional e internacional; organismos de investigación, agencias de financiamiento, otras instituciones que quieran aplicar el software o la experiencia obtenida del proyecto.

6 Efectividad

La efectividad de un repositorio institucional se mide por su *visibilidad, uso y referencia*. La presencia en portales y buscadores proveen *visibilidad* a los datos de nuestro repositorio. Esto se debe a una estructuración adecuada del código HTML optimizado para los motores de búsqueda más importantes de la red.

Nuestro sitio aparece en varios harvesters de nivel internacional:

- Único repositorio institucional argentino que está listado en el Ranking mundial de Repositorios Web, generado por el Laboratorio de Cibermetría del CSIC de España en la ubicación 378. (http://www.webometrics.info/index_es.html)
- Registrados en ROAR (<http://roar.eprints.org/>) a partir del 29 de febrero de 2009, junto a ocho repositorios argentinos y entre los seis repositorios argentinos en OpenDoar (<http://www.opendoar.org/>).

Las *estadísticas de uso* se obtienen a través de sistemas de medición como Google Analytics y Piwik (alternativa libre de estadísticas web) y el análisis propio de Logs del servidor. Se evalúa cantidad y duración de las visitas, descargas y toda la actividad realizada por los usuarios en el sitio web.

Mediante el uso de los sistemas de medición relevamos la cantidad de enlaces existentes en la web que *referencian o enlazan* hacia la biblioteca digital.

7 Facilidad de Reproducción

Se decidió el desarrollo como software de código abierto. Existe un convenio con la Universidad Nacional del Nordeste para implementarlo, por lo cual se prevé un rediseño para hacerlo autoinstalable.

El software está realizado de tal manera que permite generar y administrar plantillas, tipos de objetos y metadatos en forma flexible permitiendo en la misma instancia trabajar con distintos formatos si fuese necesario, con la posibilidad de adaptación de acuerdo a la realidad de la institución que lo implemente.

8 Ambiente de Hardware y Software

El sistema actualmente está corriendo sobre un único servidor de aplicaciones dual core con 1gb de ram y 320 gb de disco. Todo el sistema está basado en software de código abierto: Apache, Php, Postgresql, Sphinx y Bash.

Corre sobre un servidor Ubuntu aunque puede hacerlo sobre cualquier variante de Linux o Unix que permita correr las herramienta antes descriptas.

Para el desarrollo se utilizan también herramientas Open Source como Netbeans y Aptana.

El manejo de la carga de datos se realiza a través de formularios en base web Html y Php. La visibilidad de los datos -la interfaz de usuario- es a través de una web pública (html - css - php).

9 Descripción del software

9.1 Introducción

El software de la Biblioteca Digital fue desarrollado en 2006 sobre *plataforma Web* en lenguaje PHP5 y utilizando como motor de bases de datos Postgresql. Desde entonces se han realizado múltiples cambios y actualizaciones en base a los requerimientos de usuarios y mejoras en las opciones de recuperación y presentación de la información.

BDuncu permite administrar una biblioteca de objetos digitales realizando descripción de los mismos utilizando metadatos. Dispone de un buscador avanzado y un lenguaje de plantillas propio para presentar los metadatos asociados a los objetos digitales. El software integra opciones de seguridad en la administración de seguimiento de actividades de la misma.

Fue concebido siguiendo las necesidades propias de la Uncuyo y por aportes de los usuarios. Se está trabajando en su puesta a punto para ser usado en otras instituciones.

Está desarrollado utilizando librerías AJAX (Prototype y Scriptaculous) asegurando el soporte en distintos navegadores.

9.2 Módulos del sistema

El sistema consta de 6 módulos:

1.- Módulo: Administración de objetos

Un objeto digital es cualquier archivo subido a la Biblioteca Digital.

Los archivos pueden subirse a través de protocolo http o ftp quedando disponibles para su utilización. El límite de tamaño está definido por la configuración del servidor. Los archivos se mantienen almacenados en una estructura de directorios basada en el número de objeto.

Una vez subidos los archivos se inicia el proceso de carga de los metadatos descriptivos. El usuario debe seleccionar el documento y asociarlo a una plantilla de carga de metadatos previamente definida. Este módulo asigna al Objeto un número único (en adelante *IDObjeto*).

El proceso es el siguiente:

1. Se realiza una carga del objeto digital ya preparado para su publicación (pdf, mp3, flv, mpg, jpg y otros formatos de imágenes, etc.) y se almacena en una carpeta privada separada por tipos de archivos para facilitar la selección posterior.
2. Se asocia el objeto al tipo de plantilla ya definida con anterioridad, por ejemplo: un mp3 a la plantilla de *archivos sonoros*. El sistema asigna el IDObjeto y el

mismo queda disponible para cargar los metadatos descriptivos. La relación entre objeto físico y los metadatos se realiza en una tabla “objetos” y otra llamada “metadatos”.

3. El archivo seleccionado es movido al árbol de objetos donde se crea una carpeta usando el IDObjeto.
4. Se registra en el módulo de usuarios quién aportó el objeto e inició la carga de metadatos.

2.- Módulo: Definición de plantillas de carga

Antes de comenzar a cargar metadatos descriptivos para cada objeto debemos definir qué nombres de metadatos serán usados y a qué estándar adherimos. Además se deben definir las relaciones entre objetos y entre metadatos.

Para estas tareas el sistema dispone de un administrador de plantillas y de campos de ingreso asociado. Se pueden crear tantos campos de entrada como se necesiten y en el formato elegido (mods, dublin core, dublin core calificado, etc.). Los nombres de los metadatos y sus relaciones son totalmente libres y pueden ser modificados en cualquier momento por los administradores designados.

Esta definición se hará desde el módulo de plantillas de carga donde se declara:

1. Nombre de la plantilla.
2. Formato y tipo de objeto que alojará, los definidos hasta el momento son:
 - Revistas (jpg, tapa de revista).
 - Números de revistas (jpg, tapa de número).
 - Artículos de revistas (pdf, artículo completo).
 - Videos (jpg, fotograma de video).
 - Bloques de video (flv, bloque-hasta 10 megas- de video).
 - Audio (jpg, imagen).
 - Segmento de audio (mp3, hasta 10 megas).
 - Libros Electrónicos (pdf e imagen asociada).
 - Tesis (pdf, material complementario).
 - Producción científica académica (pdf).
 - Producción científica académica (pdf).
 - Microprograma audio (mp3).
 - Microprograma video (flv).
 - Informe de investigación (pdf).
 - Artículos relacionados -informes parciales, de avance- (pdf).
3. La relación con otras plantillas.
4. La plantilla HTML de publicación (plantilla pública que debe generarse usando el lenguaje de templates).
5. La plantilla XML de publicación OAI.
6. Se definen cuales son las etiquetas de metadatos que se muestran en la parte pública para título, autor y tipo de documento. Por ejemplo: en el metadato dc.creator pueden colocarse “productor” y no “autor” del objeto en el caso de los

videos.

El módulo permite obtener un listado de las plantillas definidas y de sus campos.

ID	Nombre	Tipo Objeto	Formato	Tipo
4	Revistas	Revistas	.pg	textuales
5	Video	Video	.pg	videos
6	Eloque de Video	Bloque_video	avi	videos
7	Número de Revista	Numeros	.pg	textuales
8	Artículo de Revista	Articulos	pdf	textuales
10	Informe de Investigación	Articulos	pdf	textuales
11	Producción Científico Académica	Articulos	pdf	textuales
14	Microprograma (Video)	Informe	avi	videos
16	Audic	Audio	.pg	audios
18	Segmento de Audio	Bloque_audio	mp3	audios
19	Libros Electrónicos	libro_electronico	pdf	textuales
21	Tesis	Tesis	pdf	textuales
22	Produccion Científica Independiente	Articulos	pdf	textuales
23	Colección CDH	Objeto Libre	nulo	textuales
24	Documento Fotográfico - CDH	Foto Historica	.pg	textuales

Fig. 1. Listado de plantillas actuales de la BDuncu

Para la creación de los metadatos de carga se cuenta con un administrador Web donde se define qué tipo de metadato aparecerá en la pantalla de carga y qué etiquetas, ayudas, listas desplegables, etc. aparecerán en formulario de entrada de datos.

Este editor de campos permite definir el orden en el que aparecerán los metadatos en el formulario de ingreso.

3.- Módulo: Administración de metadatos

Una vez definidas las plantillas de ingreso de datos el sistema genera los formularios automáticamente.

El proceso es el siguiente:

1. Se selecciona el objeto aportado (el cual posee un IDObjeto y el archivo digital correspondiente) y el sistema presenta la hoja de carga de acuerdo a la definición de los metadatos creados. Ver Anexo “Formulario de ingreso de metadatos”.
2. El operador completa la carga de los metadatos descriptivos utilizando todos los recursos presentados en el formulario de entrada de datos.
3. El objeto y sus metadatos asociados quedan almacenados y se marcan como registro en revisión. El registro se lista como disponible para los catalogadores de la Biblioteca Digital. En este estado ya se pueden ver como si estuviesen publicados y chequear la salida OAI correspondiente para verificar la correcta visualización.

4. El registro una vez validado por los catalogadores se publica y desde ese momento integra los índices de búsqueda y los listados públicos.
5. Todos los movimientos de carga, edición y corrección se registran en el sistema de usuarios, esto permite ver un registro de cambios de cada objeto y sus metadatos.

Se dispone de un árbol de plantillas, metadatos y valores ingresados desde donde se pueden hacer correcciones, verificaciones de los datos ingresados y cambios globales.

4.- Módulo: Visualización y recuperación de los objetos>

Este módulo integra las opciones de búsqueda y visualización de las fichas de los objetos.

Para facilitar la tarea del diseñador y permitir una separación en capas de la lógica, datos y visualización se desarrolló un sistema de plantillas de publicación (templates) que permite presentar los metadatos en diversas formas.

Las plantillas públicas (diseño de las fichas de los distintos tipos de objetos) se almacenan en el sistema como archivos HTML para facilitar el desarrollo desde herramientas como Aptana, Dreamweaver y otros editores de páginas webs.

Este lenguaje de plantillas permite al diseñador del sitio público una total libertad a la hora de mostrar los metadatos ingresados. Para ello se utilizan tags html para la obtención y transformación de los metadatos. El lenguaje se ha diseñado como funciones de transformación sobre los metadatos propios del objeto y de los objetos relacionados.

Ejemplos:

Metadatos	Occ	Función Aplicada	Resultado	Tipo
La historia argentina	1	{ \$dc.title }	La historia argentina	Presentación directa del metadato
Calderón, Lisandro Valpreda, Edda Claudia	2	{ \$dc.creator guiones }	Calderón, Lisandro– Valpreda, Edda Claudia	Presentación usando una función de transformación
arte historia argentina	3	{ \$dc.subject occ:2 }	historia	Presentación usando una función de transformación con parámetros (en este caso el número 2)

<p>Extracto de una plantilla pública <div >Título: {\$dc.title} </div> <div >Autores: {\$dc.creator i} </div></p>	<p>Salida Título: La historia argentina Autores: • Calderón, Lisandro • Valpreda, Edda Claudia</p>
---	---

Todas las plantillas de carga deben tener una plantilla de publicación para poder presentar la información a los usuarios externos. En caso de no definirse los metadatos se presentarán en formato XML.

Este lenguaje de plantillas también se utiliza para cualquier recuperación del sistema: actualmente se usa para la obtención de RSS, exportación XML a otras webs de la UNCuyo y para el sistema OAI-PMH del que haremos referencia más adelante.

Recuperación de la Información

El sistema de búsquedas de la biblioteca ha sido modificado recientemente para permitir:

1. Mayor velocidad en la respuesta de las búsquedas.
2. Ranking y ordenamiento de los resultados usando “pesos” en los distintos metadatos ingresados.
3. Filtrado por tipo de objeto, disciplinas, y otros filtros definibles desde el administrador de plantillas.
4. Facilitar la experiencia del usuario en la interpretación de los términos de búsqueda.

Todas las búsquedas del sistema, ya sean desde la interfaz pública o desde los webservices se realizan utilizando el motor *Sphinx*. El motor nos devuelve solo un set de registros que son luego cruzados con las bases de datos Postgresql para la obtención de los datos de los objetos o su trabajo posterior para presentarlo en una u otra forma.

Entre otras cosas Sphinx nos permite:

- Realizar múltiples búsquedas en la misma instancia utilizando distintos filtros, de esta manera se recuperan registros basados en el tipo principal, según el soporte (texto, audio o video).
- Permitir al usuario utilizar un lenguaje de búsqueda muy difundido. Los operadores booleanos que utiliza son los mismos de los grandes buscadores de Internet. Ej: para obtener información de arquitectura pero excluyendo el término Mendoza podemos realizar la siguiente consulta: arquitectura -Mendoza.
- Los términos ingresados se buscan por palabras raíces, y derivadas del *lexema* original. Con eso se mejora la búsqueda porque también busca las palabras en plural o singular, masculino o femenino, en diferentes tiempos verbales,

gerundio, etc. Por ejemplo: la búsqueda de “consumidores” nos traerá resultados ya sea para consumidor o para consumo. El ranking de los resultados asegura que los primeros resultados sean los adecuados para la palabra que estamos buscando.

- Todas las consultas realizadas quedan almacenadas en un log que permite realizar análisis de las búsquedas realizadas. Durante el año 2010 esta información está siendo analizada para ajustar los índices de búsqueda y las palabras que describen los objetos.

5.- Módulo: Usuarios

La administración de la BDuncu se realiza en un *entorno web protegido* con usuario y clave. El módulo permite:

1. Seguimiento de los operadores del sistema (registra todos los movimientos en objetos, cambios en los metadatos, cambios globales, modificación de listas desplegadas, etc.)
2. Permite activar o desactivar usuarios y asignarles permisos de gestión dentro del sistema. Posee 5 niveles de usuarios que determinan las tareas que se pueden realizar en otros módulos.

6.- Módulo: OAI-PMH

Desde 2009 se cuenta con un módulo de exportación y presentación de objetos utilizando el protocolo OAI-PMH (Open Archives Initiative – Protocol Metadata Harvesting).

Según la wikipedia: “El protocolo *OAI-PMH* (Open Archives Initiative-Protocol Metadata Harvesting), es una herramienta de interoperabilidad independiente de la aplicación que permite realizar el intercambio de información para que desde puntos (proveedores de servicio), se puedan hacer búsquedas que abarquen la información recopilada en distintos repositorios asociados (proveedores de datos)”

La implementación de la BDuncu permite utilizar el mismo sistema de plantillas públicas para la presentación de los metadatos descriptivos, de esta manera se cumple con el protocolo OAI y se permite realizar harvesting sobre la colección de objetos digitales.

Esto permite que la biblioteca sea listada en recolectores de registros de objetos digitales e integrar redes de repositorios internacionales.

ANEXOS

Metadatos: “*Los metadatos son datos asociados a un documento digital que recogen información fundamentalmente descriptiva (autor, título, etc.). También pueden incluir información de administración (creación del recurso, derechos, control de acceso ...), y preservación (tipo de formato, etc.).*”

En nuestro sistema los metadatos se administran mediante formularios webs que permiten su carga o modificación ya sea en forma individual o en forma global (mediante cambios globales).

Un metadato mantiene su relación con el objeto del que depende usando el número de Objeto asignado en forma secuencial por el sistema.

Todos los metadatos son repetibles (aceptan múltiples ocurrencias por registro), se mantiene el orden en el que son ingresados. Esto permite hacer pesar su posición en las búsquedas y la visualización.

Otras opciones de los metadatos: Un metadato puede definir el ordenamiento en listas de objetos. De esta manera, una fecha puede definir cómo se presentan ordenados los números de las revistas.

En la creación de metadatos se pueden definir si los valores van a generar entradas en el índice para la recuperación. Así podemos hacer que cualquier metadato se integre a la lista de títulos o listas de palabras claves y descriptores.

Listas de autoridades o de control: El sistema permite crear listas de control que sirven para facilitar la carga y evitar errores de tipeo.

Cada lista puede tener “n” valores y textos descriptivos que son usados como metadatos luego durante la descripción de los objetos.

Estas listas se presentan como selectores desplegados en la interfaz pública de consulta, además permiten totalizar la cantidad de objetos que contiene un valor definido. Ej: En la home del sitio “ARTE (83 objetos)”.

Formulario de ingreso de metadatos: Este formulario se genera en forma automática permitiendo el uso de campos, subcampos, listas desplegadas y otras opciones.

La modificación y presentación del formulario está restringido por el nivel de usuario y los permisos definidos en el módulo de usuarios. Desde el formulario de edición se puede acceder al histórico de cambios. Este muestra los usuarios y fechas de modificación del registro.

En las plantillas de carga se pueden usar distintos tipos de campos de entrada de datos.

1. Texto en una sola línea: Campo simple con un largo máximo de 255 caracteres. Se presenta en la hoja de carga como un campo `<< input type="text" >>`. Campo de tipo repetible que permite además seleccionar de los valores antes ingresados mediante el uso de una ventana flotante. Contempla el uso de subcampos que pueden ser usados para definir otros metadatos o atributos del metadato principal. Ejemplo: dc.title para el título del objeto, y un campo “language” como subcampo de éste donde se define el lenguaje en que está el título.
2. Texto multilínea: Texto en múltiples líneas que permite almacenar además el salto de página, acepta múltiples ocurrencias y permite el uso de subcampos. Ejemplo: dc.abstract o resumen.
3. Fecha: permite almacenar una fecha en formato ISO 8601 de la forma YYYY-MM-DD. Se presenta como un campo `<< input type="text" >>` con un calendario para facilitar la carga. Permite múltiples ocurrencias y además subcampos.
4. Asociado a una lista de autoridades o de control: este tipo de campo se presenta como un selector flotante o autocompletable (mediante el uso de peticiones Ajax) y no permite la creación de nuevos valores. La administración de las listas de control se realiza previamente a la carga y por los administradores del sistema.
5. Asociado a otro objeto: en este caso se permite la creación de relaciones entre

objetos mediante el IDObjeto. Esto permite por ejemplo asociar números de revistas a revistas y artículos a números de revistas produciendo un árbol jerárquico de objetos que facilita la navegación del usuario. Para la selección del objeto superior se usan ventanas flotantes donde se muestran los objetos disponibles para realizar la relación.

Nota: Submetadato o subcampo: un concepto muy importante que incorpora la BDuncu es la posibilidad de trabajar con subcampos. Esto permite por ejemplo mantener información estructurada para describir un autor que trabaja en dos universidades

- Metadato: dc.creator->Calderón, Lisandro
 - Submetadato: filiación->UNCuyo, Fac. de Filosofía y Letras.
 - Otra ocurrencia de submetadato: filiación-> UNC, Fac. De Artes.
 - Submetadato: email-> cliz@uncu.edu.ar

Además se permite que exista más de una ocurrencia para cada submetadato. Todos los tipos de campos anteriormente descriptos pueden ser usados como submetadatos.

Todos los campos se almacenan respetando el orden de ingreso. Esto permite su recuperación por número de ocurrencia, especialmente importante para marcar autores principales de secundarios o el peso de las palabras claves y descriptores basados en el orden de ingreso.



Fig. 2. Proceso de Obtención y Publicación de los Objetos Digitales

Bibliografía Consultada

Sphinx - Free open-source SQL full-text search engine. (s.d.). . Accedido el 25 Junio 25 de 2010, a partir de <http://www.sphinxsearch.com/docs/current.html>.

Hernández Pérez, T., Rodríguez Mateos, D., & Bueno De la Fuente, G. (s.d.). Open Access: el papel de las bibliotecas en los repositorios institucionales de acceso abierto. Artículo. Accedido el 25 de Junio de 2010, a partir de <http://digitum.um.es/jspui/handle/10201/4017>.