

"Objetos digitales : una experiencia de representación con metadatos Dublin Core"

Lic. Nélide Elba García
nellygarcia@nodoalem.com.ar
Bib. Susana Eunice Jaroszczuk
su_eunice@yahoo.com

Universidad Nacional de Misiones. Facultad de Humanidades y Ciencias Sociales.
Carrera de Bibliotecología
biblio@fhycs.unam.edu.ar
Tucumán 1946 – Posadas – Misiones Argentina

Resumen

En este trabajo se exponen algunas tendencias y estrategias vigentes acerca de la estructura y normalización de la información en entornos digitales recurriendo al modelo de los metadatos como una de las alternativas para la organización y conservación de objetos digitales en la Web.

Se describen estándares que intervienen en la organización de las bibliotecas y los repositorios digitales como los metadatos y el software libre, ostentando a los repositorios institucionales creados en entornos de “acceso abierto”, como espacios pertinentes para la conservación, búsqueda y acceso universal a la información digital académica.

Se presentan las políticas establecidas para la selección y descripción de los documentos digitales que integran el corpus; señalando las herramientas de descripción utilizadas para la conformación de los objetos digitales que protagonizan el repositorio de documentos como así también las herramientas tecnológicas que soportan la creación de este tipo de servicios de información digital.

Se describe la experiencia de construcción de un repositorio de documentos utilizando software Greenstone y los quince elementos que componen el modelo de metadatos Dublin Core en el marco del proyecto de investigación 16/H-213; FHyCS-UNaM: “Iniciativas de acceso abierto para la conformación de repositorios institucionales”.

Palabras clave:

Objetos digitales – Organización – Acceso Abierto – Repositorio de documentos —
Metadatos Dublin Core – Greenstone

1. Introducción

El crecimiento incontrolable de los recursos de información existentes en la web dieron lugar a la paradoja “acceso global a la información”, mientras en la realidad los usuarios se enfrentan a una red confusa carente del orden necesario para recuperar la información pertinente. En ese contexto, la tarea profesional vinculada con las bibliotecas y sus colecciones y los servicios de información se han visto modificadas en aspectos como la selección, adquisición, procesos de descripción y modo de acceso aplicados a los objetos de información; dando lugar a un nuevo reto acerca de cómo organizar la información en Internet para garantizar su localización y recuperación exitosa. El desarrollo de las tecnologías de la información y las comunicaciones ha hecho que todos puedan crear y publicar en la red, sin prever que la falta de tratamiento de ese acopio documental provocaría puntos débiles en la tarea de localización y recuperación que realizan los usuarios a través de los denominados buscadores de información.

El objetivo de este trabajo es reflexionar sobre las estrategias vigentes acerca de la estructura y normalización de la información en entornos digitales recurriendo al modelo de los metadatos y al software libre como una de las alternativas para la organización y conservación de objetos digitales en la web. Reseñando la experiencia llevada a cabo en el marco del Proyecto de Investigación 16/H-213 “Iniciativas de acceso Abierto para la conformación de repositorios institucionales”, ejecutado durante el período 2006-2007 desde la Secretaría de Investigación de la UNaM, cuyo uno de los objetivos más importantes ha sido demostrar que es posible construir un repositorio institucional donde resguardar y preservar la información académica local respondiendo a los parámetros mencionados.

2. Información digital: características, propiedades y alternativas de preservación

La información digital es aquella que está codificada en un formato que puede interpretar un ordenador. Y la información digital está en línea cuando es posible acceder a ella desde terminales u ordenadores remotos, a través de redes de área local, de área amplia o de combinaciones de ambas (Codina, 2001).

Características:

La **organización** de los recursos digitales no es centralizada al estilo de los catálogos tradicionales ni se organizan en tablas por atributos al estilo de los sistemas de gestión de bases de datos.

La **ubicación** o localización de los mismos es remota y un mismo recurso puede alojarse en varios sitios distintos.

Un mismo recurso puede existir en distintas **versiones** como Postscript, ASCII, etc. Son móviles e **inestables** comparados con los recursos de naturaleza fija; los archivos con sus datos sufren constantes traslados entre distintos servidores transformando en obsoleta la dirección original, URL. Las versiones viejas de documentos, obsoletas o no, permanecen sin eliminarse marcando **redundancia**.

La información en Internet es hipertextual; los documentos se relacionan mediante enlaces y generalmente es difícil determinar los límites de un documento o un sitio. Además, de esta **complejidad** un sitio en Internet puede consistir en un solo documento o en varios.

Propiedades:

Como propiedades la información digital se distingue por la **computabilidad**, es decir que puede ser procesada o “calculada” por un ordenador. Dada su propiedad de **virtual** la información digital no está sujeta a las limitaciones propias de la analógica. Y su **capacidad** admite la ausencia de limitaciones prácticas en cuanto al volumen de información al que puede tener acceso en línea mediante interfaces unificadas.

Las particularidades señaladas acerca de los recursos digitales frente a la naturaleza “indeleble” de los recursos físicos propios de la biblioteca tradicional hacen vulnerables los procedimientos y herramientas tradicionales de procesamiento, instalando en las organizaciones que se vinculan con su generación, recuperación y uso la reflexión y, cada vez más, la decisión de establecer normas que permitan estructurar su contenido y contribuir a facilitar el proceso de búsqueda y recuperación de la información en el entorno web pero sobre todo asegurar su durabilidad, es decir, su preservación.

Subestimar la necesidad de ese tratamiento documental implicaría correr *el riesgo de no poder acceder a mucha información valiosa y lo que es peor, desconocer totalmente su existencia* (Daudinot Founier, 2008).

La problemática socio-cultural que se ha generado en relación a la cantidad y variedad de recursos de información disponibles en la red, ha sido considerada por la UNESCO cuando publicó la “Carta para la preservación del patrimonio digital”, en cuyo Art. 1 establece el alcance de los objetos digitales:

El patrimonio digital consiste en recursos únicos que son fruto del saber o la expresión de los seres humanos. Comprende recursos de carácter cultural, educativo, científico o administrativo e información técnica, jurídica, médica y de otras clases, que se generan directamente en formato digital o se convierten a éste a partir de material analógico ya existente. Los productos “de origen digital” no existen en otro formato que el electrónico.

Los objetos digitales pueden ser textos, bases de datos, imágenes fijas o en movimiento, grabaciones sonoras, material gráfico, programas informáticos o páginas Web, entre otros muchos formatos posibles dentro de un vasto repertorio de diversidad creciente (Biblioteca Nacional de Australia. (2003).

Preservación:

Así como el desarrollo de las tecnologías de la información y las comunicaciones ha generado una serie de ventajas en relación con la información y su acceso, también ha instalado una fuerte preocupación en quienes tienen la misión de organizarla relacionada con la importancia de conservar los objetos de información en soporte digital. Es necesario elaborar procedimientos que aseguren la perdurabilidad del contenido de esos soportes digitales. Al respecto la UNESCO expresa en su Carta:

... Muchos de esos recursos revisten valor e importancia duraderos, y constituyen por ello un patrimonio digno de protección y conservación en beneficio de las generaciones actuales y futuras (Art.1).

Hay que preservar y poner a disposición de cualquier persona el patrimonio digital de todas las regiones, naciones y comunidades a fin de propiciar, con el tiempo, una representación de todos los pueblos, naciones, culturas e idiomas (Art.9).

La preservación digital puede definirse como un conjunto de actividades relacionadas con la preservación de documentos en formato digital, con el objetivo que perdure el mayor tiempo posible en su estado original. Lo que implicaría ocuparse no solo del mantenimiento del objeto sino también de su contenido informativo. Por lo tanto, “la planificación y el desarrollo teórico de las tareas de preservación es igual cuando se trata de información digital que de cualquier otro objeto de información” (Candas Romero, 2006).

Según la “Cornell University Library”, el objetivo de la preservación digital es “mantener la capacidad de visualizar, recuperar y utilizar colecciones digitales frente a las infraestructuras y elementos tecnológicos y de organización que cambian con mucha rapidez”.

Definición de la que Candas Romero rescata dos ideas fundamentales: por un lado, la necesidad de prestar atención a los elementos tecnológicos asociados a la información digital, y por otro, la rapidez con que se producen los cambios en este entorno.

La evolución de la tecnología puede afectar tanto los componentes físicos (hardware) como los elementos lógicos (software) de la información digital, siendo este último el problema más habitual de su preservación. Se modifica el formato de los archivos y/o los programas de lectura y escritura de documentos dificultando la lectura de los DLO (document like object = documentos como objetos) en su forma original. En este caso habrá que optar por métodos como la migración, emulación u otros como estrategias para preservar los soportes de almacenamiento e impedir con ello que el avance tecnológico torne inaccesible la información digital. No obstante, esa manipulación de la información digital debe realizarse conservando la originalidad (el documento primario), la autenticidad (autoría) y la integridad (estar completo) del objeto de información digital y de la información misma.

Significa que aún siendo un factor externo a la información y al propio soporte el entorno tecnológico es el que distingue la información encontrada en soporte tradicional, a la cual el acceso se realiza en forma tangible y directa, de la información digital a la que caracteriza y define.

Desde la última década del siglo pasado numerosos y significativos proyectos internacionales se han ocupado de adquirir experiencia concreta en la preservación de los documentos digitales fomentando en ese sentido el desarrollo de políticas institucionales (Serra Serra, 2001).

La tarea se caracteriza por ser global, continua y compleja involucrando factores físicos y lógicos de la información y exigiendo una adecuada tarea por parte de los responsables de centros de información y bibliotecas de sumar la preservación digital a sus proyectos cotidianos.

3. Información digital: herramientas de procesamiento

El nuevo modelo estructural de la información surgido con el avance de Internet ha dado lugar a una nueva definición de las funciones bibliotecarias en general. Aunque en particular, la problemática reside en el modo de organizar la información digital de manera que se pueda garantizar su localización y recuperación y, lo que es más importante aún, su preservación.

Metadatos

En este contexto de modificación sustantiva las propuestas de procesamiento y manejo de los recursos digitales tienen como pilar a los metadatos, cuyo uso normalizado se presenta como la solución a los gestores y creadores de recursos de información quienes tendrán que agregar a éstos descripciones susceptibles de procesarse por los buscadores de información.

Los metadatos se emplean para describir el continente, el contenido y otras características de los datos primarios posibilitando a una persona o máquina ubicar y entender la información. Son datos secundarios como pueden ser el autor, el título, las palabras clave, el resumen, la fecha, u otros que describen los datos primarios o recursos de información.

En este sentido, son comparables al tradicional trabajo de descripción catalográfica. Sin embargo, algunos teóricos consideran que *"el contexto general de producción y uso de esta información es substancialmente diferente y está impulsada para ir más allá del paradigma tradicional de la catalogación. Entonces, considerar el proceso de creación de metadatos como un tipo de catalogación simplificada, sería probablemente un error importante (Gradmann, 1998)"*.

A diferencia de las fichas catalográficas, una de las características más importantes de los metadatos, es su capacidad de relación o de establecer enlaces indizando y clasificando inmensas cantidades de información de diversos tipos. Por eso, se han hecho imprescindibles en la recuperación de la información en línea (San Segundo, 1998).

Más allá de su identificación terminológica y etimológica así como de su evolución y condición de documento secundario, lo destacable es que el concepto estricto de metadatos surge en la web; es amplio, permite muchas conceptualizaciones, interpretaciones y aplicaciones.

En cuanto a tipología, funciones y modelos los metadatos presentan múltiples alternativas que dependen de factores también diversos como: el tipo de información que describen, el nivel de estructuración de esta información, el lugar donde se encuentren, su ámbito de aplicación, el tipo de usuarios que los utiliza, sus finalidades, entre otros.

A propósito de la preservación digital, se define a los metadatos como "elementos o estructuras de organización de la información que, asignados a cada objeto de información electrónica la clasifican, categorizan y describen. En definitiva, es información estructurada sobre información distribuida" y su papel emergente es la actualización y la preservación de la información (Méndez Rodríguez, 2002 citado por Candás Romero).

Los metadatos se encuentran en fase de construcción técnica y por eso no existe un consenso generalizado en su conceptualización o sobre los tipos y categorías existentes. Lo

evidente es que entre sus funciones fundamentales está la de describir los objetos de información, informar a los usuarios la existencia, el contenido, la calidad y los objetivos de los recursos de información descritos, garantizando el punto de acceso a la información que los usuarios desean consultar.

En cuanto a los **modelos**, que también son varios, en este trabajo sólo se aborda el modelo Dublin Core por ser un modelo de metadatos elaborado y auspiciado por la DCMI (Dublin Core Metadata Initiative), una organización dedicada a fomentar la adopción extensa de los estándares interoperables de los metadatos y a promover el desarrollo de los vocabularios especializados de metadatos para describir recursos que permitan sistemas más inteligentes en cuanto a su localización.

Dublin Core es un sistema de 15 definiciones semánticas descriptivas que son opcionales, se pueden repetir o aparecer en cualquier orden. Diseñado para proporcionar un vocabulario de características "base", capaces de proporcionar la información descriptiva básica sobre cualquier recurso, sin que importe el formato de origen, el área de especialización o el origen cultural. El nombre tiene su origen en Dublin (Ohio, Estados Unidos), ciudad que en 1995 albergó la primera reunión a nivel mundial de muchos de los especialistas en metadatos y web de la época.

En general, los elementos que lo componen se clasifican en tres grupos que indican la clase o el ámbito de la información que se guarda en ellos: elementos relacionados con el **contenido**, la **propiedad intelectual** y la **instanciación** del recurso.

En cuanto al **uso** los metadatos Dublin Core presentan ventajas que lo hacen accesibles para describir recursos en un sistema de información como: simplicidad, flexibilidad, independencia sintáctica, interoperabilidad semántica, alto nivel de normalización formal, crecimiento y evolución del estándar a través de una institución formal consorciada: la DCMI, consenso internacional, modularidad y arquitectura de metadatos en la web.

En función del trabajo que se presenta, lo fundamental es que los metadatos también son útiles para preservar en forma documentada la información que poseen las instituciones, evitando que estas sean vulnerables a perder todo el conocimiento sobre sus datos, si por alguna razón desaparecen los responsables de su origen.

Software Libre

Actualmente la gama de aplicaciones de software libre orientadas al manejo de información digital en web y a la gestión bibliotecaria y de servicios de información es numerosa. Uno de los principales problemas que enfrenta la sociedad contemporánea al momento de implementar sistemas más justos y paritarios en materia de acceso a la información, se encuentra en las decisiones de carácter político y económico que conllevan la utilización de una determinada tecnología informacional (Zurita Sánchez, 2008).

Para cumplir con los objetivos de este movimiento conocido como Open Access se han desarrollado una serie de estándares y herramientas basadas en software libre que permiten levantar repositorios de recursos digitales (Open Archives) de manera rápida y con pocos medios. Este tipo de aplicaciones son numerosas y se encuentran gratuitamente en Internet generando una importante disminución en los costos de producción y mantenimiento de los servicios de información. Se trata de una alternativa sólida y viable para el diseño de servicios e información a bajo costo y con un alto grado de sentido social y comunitario, ya que el trabajo que se desarrolla utilizando software libre puede ser compartido y reutilizado por otros profesionales y colegas, sin temor a violar o transgredir licencias y derechos patrimoniales, como normalmente sucede con la utilización de software propietario (Siló, 2005).

Modelo de “acceso abierto” a la información

Como consecuencia de estos cambios sustantivos, en las comunidades académicas nacionales e internacionales, se vienen produciendo importantes reflexiones y prácticas con el objeto de preservar la información originada en soporte digital.

Se produce una *crisis del modelo tradicional de comunicación científica* frente a un nuevo modelo de gestión de los procesos de generación, conservación, intercambio y transferencia de la información fundadas en varias iniciativas internacionales que perfilan el compromiso de respaldar un paradigma que privilegie el acceso del usuario a aquellos recursos de información digital restringidos a la tarea administrativa o de investigación hacia el interior de las instituciones académicas.

Una de las iniciativas más destacadas es el movimiento que emerge bajo la denominación de “acceso abierto” (open access) y tiene por objeto facilitar el acceso a la literatura científica en forma libre y sin restricciones a través de Internet permitiendo a cualquier usuario su lectura, descarga, copia, impresión, distribución o cualquier otro uso

legal de la misma, sin ninguna barrera financiera, técnica o de cualquier tipo. La única restricción sobre la distribución y reproducción sería dar a los autores control sobre la integridad de su trabajo y el derecho a ser citado y reconocido adecuadamente (Budapest Open Access Initiative-BOAI, aprobada el 14 de Febrero de 2002).

La difusión de la filosofía “open access” ha ido evolucionando a la par del presente siglo a través de acuerdos y adhesiones emergentes de distintas organizaciones. Y con ella, la conciencia de que el patrimonio intelectual de las instituciones académicas debe difundirse a través de la red y a la luz de los parámetros de calidad internacionales permaneciendo allí en forma de publicaciones electrónicas a disposición de toda la sociedad que lo financia.

“Gran parte de la creación intelectual surge en centros académicos, financiada mayoritariamente con fondos públicos” (Dávila, 2006). De ahí el derecho y la necesidad de acceso libre al conocimiento de una sociedad que atestigua el desarrollo mediado por la tecnología en todas sus esferas.

Repositorios institucionales

Las universidades como gestoras de la información y el conocimiento son las que protagonizan los procesos de generación, conservación, intercambio, transferencia, aplicados a la solución de problemas y asociados al desarrollo. Por ello, es la universidad también la responsable de conservar el patrimonio intelectual de los pueblos. Razón por la cual es indispensable generar políticas, mecanismos e incentivos que permitan el acceso a ese conocimiento a la mayor cantidad de personas.

Actualmente, los Repositorios Institucionales (RI) constituyen una alternativa a los modelos de publicación, difusión y preservación de contenidos académicos. Estos se refieren a un conjunto de servicios de almacenamiento, gestión y disseminación de materiales digitales disponibles a los miembros de una determinada comunidad académica donde antes que la herramienta computacional asociada a su manejo, importan el contenido, la calidad, la seguridad, la actualización constante, la facilidad de acceso a su contenido y la facilidad de su difusión.

Implementar un RI implica resolver los problemas de tipo organizacional de la institución quien debe demostrar su capacidad de apropiación tecnológica descubriendo, generando y aplicando esquemas metodológicos para desarrollar, captar y difundir contenidos digitales con la convicción de que las TICs constituyen uno de los ejes centrales de la actividad académica al promover nuevas maneras para la producción, preservación y disseminación del conocimiento.

En este nuevo escenario los especialistas de la información tienen la oportunidad de demostrar su interés y capacidad de asumir el reto de la preservación de los documentos electrónicos, la creación y mantenimiento de repositorios institucionales, el establecimiento de normas y canales para la difusión eficaz de contenidos y la calidad garantizada de los metadatos.

4. Resultados de una experiencia con Dublin Core y Software Greenstone

En el marco del Proyecto de Investigación 16/H-213 “Iniciativas de acceso abierto para la conformación de repositorios institucionales”, ejecutado durante el período 2006-2007, uno de los objetivos más importantes ha sido demostrar que es posible construir un repositorio institucional donde resguardar y preservar la información académica local respondiendo a los parámetros mencionados.

La literatura revisada permitió visualizar que una institución puede incrementar su visibilidad y su prestigio demostrando asimismo su valor para las fuentes de financiación y subvenciones creando un repositorio con la finalidad de exponer su memoria académica. Por su parte, los investigadores pueden alcanzar mayor visibilidad de sus trabajos ampliando la difusión y el uso de los mismos. Para alcanzar estos objetivos se pueden desarrollar proyectos en diferentes categorías: **1)** la creación de repositorios de documentos; **2)** la creación de revistas de libre acceso y **3)** el acceso libre a publicaciones científicas.

Desde la perspectiva y realidad académico-administrativa en la que nos desenvolvemos el propósito ha sido abordar la tarea de alcanzar el primer objetivo: es decir, la prefiguración de un modelo de “acceso abierto” para la **creación de un repositorio de documentos relacionados con la labor del campo disciplinar de la biblioteconomía**, esto es la creación de un modelo propio adecuado a nuestras necesidades cuya demostración de implementación y funcionamiento pudiera realizarse usando un corpus acotado de documentos institucionales como demostración de la viabilidad de un proyecto propio de RI en la UNaM.

La toma de decisiones referida al producto diseñado y construido, bajo la modalidad de un demostrativo, implicó la consideración de componentes y características como la **selección del corpus documental**, la **propiedad intelectual**, la **política de normalización** y la **plataforma tecnológica** aplicada a la descripción de documentos utilizando los metadatos Dublin Core.

Selección del corpus documental

Se ha procedido a seleccionar diversos tipos de materiales documentales correspondientes a la temática bibliotecológica, generados en los últimos años “en” o “acerca de” la Carrera de Bibliotecología de la Facultad de Humanidades y Ciencias Sociales de la UNaM; documentos que por ser resultados de investigación, informes de proyectos académicos, productos de transferencia o extensión merecen ser visibles, accesibles y disponibles a través de esta acción de composición del repositorio institucional. Así es que la tipología de recursos que componen el corpus documental seleccionado está conformado por artículos científicos y de divulgación; ponencias a texto completo y resumidos; proyectos, informes, material didáctico; libros y partes de libros; informes de investigación, etc. También se han incorporado textos que constituyen material académico legal como ser plan de estudios e instrumentos resolutivos con el fin de exponer las diversas situaciones que plantea al respecto la propiedad intelectual, el derecho de autor y los permisos de publicación.

La propiedad intelectual

Por tratarse de una primera etapa de la investigación en curso, se ha procedido al recorrido teórico de la problemática de la propiedad intelectual quedando pendiente el proceso de gestión de las licencias correspondientes. Aunque se ha relevado información y realizado los primeros contactos para solicitar la licencia Creative Commons. Como proceso experimental se ha procedido al diseño de un formulario de autorización provisional en el que las partes (autor-institución) acuerdan: a) el autor a dar consentimiento para su tratamiento y difusión y b) el espacio institucional a garantizar su visibilidad y accesibilidad conforme a los ítems del acuerdo; con fines académicos y de investigación.

La política de normalización

En primer lugar, el análisis de los recursos que conforman el corpus documental seleccionado para la construcción del repositorio se ha llevado a cabo aplicando técnicas tradicionales de control bibliográfico en sus dos fases, formal y de contenido.

El análisis formal, ha respondido a técnicas de descripción en el orden que proponen las ISBDs (1993) y las AACR2R (2004); mientras que el análisis de contenido se ha realizado utilizando el Tesauro en línea de la UNESCO (2007) para establecer un mínimo de diez palabras claves que identifican el alcance temático de cada documento.

Ejemplo de cita construida desde ambos tipos de análisis, formal y de contenido

García, Nélica Elba ; Oria, Mónica y Miranda, Mirta Juana. *"Repositorios institucionales: una modalidad de construcción de memorias académicas"*. Ponencia presentada en: Jornadas Bibliotecológicas 2007 "Recursos humanos, preservación y memoria". Organizado por Universidad Nacional de Misiones, Facultad de Humanidades, Departamento de Bibliotecología, Secretaría de Extensión. Posadas, 12 al 15 de septiembre de 2007.

Archivo de Word en 11p.

Nombre del archivo: 009 Ponencia Día Bibliotecario UNaM 2007

Palabras claves: RECURSOS HUMANOS % CONSERVACION DE DOCUMENTOS % FUENTE DE INFORMACION % BIBLIOTECA % ACCESO A LA INFORMACION % BRECHA DIGITAL % INVESTIGACION % LIBRE CIRCULACION DE LA INFORMACION

Luego de establecer los criterios y utilizar las normas mencionadas en el proceso de construcción de los registros de citas formales y de contenido, se procedió a identificar cada elemento informativo de las mismas según el sistema de metadatos Dublin Core. Para ello, se han establecido guías de procedimiento y la ejemplificación correspondiente que aportaron claridad a la fase de etiquetado de las citas previo a su incorporación a la plataforma informática mencionada. Es decir, más allá de ser recopilados, su contenido informativo ha sido procesado a través de la aplicación de un conjunto de técnicas y tratamientos haciendo posible su análisis y descripción con el objeto de facilitar su búsqueda y su acceso dentro del repositorio. Los aspectos considerados fueron autoría y responsabilidad, descripción y contenido informativo-documental.

Etiquetas Dublin Core	Descripción de Documentos en Dublin Core
<Creador> Creador	Miranda, Mirta Juana; Kuna, Horacio Daniel; Prevosti, María Norma; García, Nélica Elba; Oria, Mónica.; Jaroszczuk, Susana Eunice
<Title> Título	La instalación del paradigma de acceso abierto en la UNaM: un proyecto de investigación en curso
<Subject> Materias	INVESTIGACION APLICADA % LIBRE CIRCULACION DE LA INFORMACION % BRECHA DIGITAL % PROPIEDAD INTELECTUAL % ACCESO A LA INFORMACION % BIBLIOTECA UNIVERSITARIA % RED INFORMATICA % BIBLIOTECARIO % SISTEMA DE INFORMACION EN LINEA % (TECNOLOGIA DE LA INFORMACION)
<Contributor> Colaborador	
<date> Fecha	2006-10-19

<Description> Descripción	Esta presentación pretende hacer conocer un proyecto de investigación institucional relacionado con <i>acceso abierto</i> y los <i>repositorios institucionales</i> , desarrollado en la Secretaría de Investigación y Posgrado de la FHyCS- a partir del año 2006. Su objetivo consiste en analizar las alternativas que ofrece el paradigma de “acceso abierto” para atender la problemática de la producción documental institucional -resguardo, preservación, registro, visibilidad y disponibilidad- considerando las distintas dimensiones implicadas: institucional, legal, metodológica, tecnológica e instrumental. Se analizan nuevas posibilidades de procedimiento normalizado de resguardo, preservación y proceso documental formal que aseguren: autoría, permanencia, visibilidad y disponibilidad de documentos.
<Publisher> Editor	Universidad Nacional de Cuyo (Mendoza, Argentina)
<Type> Tipo	Ponencia
<Format> Formato	Adobe Acrobat Document (pdf) - 52,7 KB (54.025 bytes)
<Coverage> Cobertura	“UNaM 2006-2007”
<Rights> Derechos	No se registra en esta etapa
<Relations> Relación	http://www.amicus.udesa.edu.ar/4bibliotecadigital/programa.html “Jornada sobre la Biblioteca Digital Universitaria - JBD 2006. “La Web y las bibliotecas digitales en el ámbito académico: acceso, interacción, visibilidad” (4ª : 19 y 20 de octubre de 2006 : Universidad Nacional de Cuyo Ciudad de Mendoza)
<Source> Fuente	No se registra en esta etapa
<Lenguaje> Idioma	spa
<Identifier> Identificador	www.amicus.udesa.edu.ar/4bibliotecadigital/ponencias/Miranda.htm (consultado: 04-12-07)

Ejemplo de un recurso descripto según las etiquetas Dublin Core

Este conjunto de procedimientos ha permitido un marco normalizado para el proyecto, aplicando instrumentos vigentes y aceptados internacionalmente en la tarea de control bibliográfico para entornos de “acceso abierto”.

La plataforma tecnológica

La construcción del repositorio modelo responde al recurso de código libre denominado Greenstone; conjunto de programas de software diseñado para crear y distribuir colecciones digitales, proporcionando así una nueva forma de organizar y publicar la información a través de Internet o en forma de CD-ROM. Es una aplicación que tiene como núcleo el motor de indización y recuperación de información textual. Es capaz de procesar e incorporar a las colecciones documentos en numerosos formatos.

Para el modelo de repositorio construido se ha logrado un producto que permite el acceso a los documentos ingresados y descriptos por categorías de autor, título y tema. El sistema también permite recuperar resultados de una búsqueda efectuada a texto completo en versión word y pdf.

Selección de Software Greenstone

Para la selección de este Software se han hecho los recorridos teóricos por espacios institucionales que ofrecen herramientas para la creación de repositorios comparando Dspace, Fedora, Content DM y Greenstone; visitado los sitios institucionales de universidades públicas de la argentina (UNLP; UNCu entre otros); y relevado información mediante una encuesta instalada en el ámbito de confluencia de los actores vinculados a la creación de bibliotecas digitales y repositorios institucionales, como la JBDU 2007.

Algunas de las conclusiones alcanzadas en esta tarea comparativa fueron:

- DSpace trabaja solo en idioma inglés, aspecto en el que difiere del resto de las herramientas que se pueden configurar en múltiples idiomas, entre ellos en español.
- Greenstone, DSpace y Fedora, al ser programas open source, gratis y con licencia pública, aunque Fedora se distribuye bajo Licencia de la Comunidad Educativa.
- CONTENTdm –al ser pago y con licencia privada– no es flexible con sus usuarios, que deben utilizarlo tal cual la empresa productora lo presenta.
- Todas permiten asignar metadatos según estándares ampliamente distribuidos, como Dublin Core, aunque también brindan la posibilidad de que los especialistas declaren otros.
- Content DM fue descartado por ser una herramienta paga y poco flexible.
- DSpace fue descartada fundamentalmente por solo tener el idioma inglés tanto su área de procesamiento como la de presentación a los usuarios.
- Fedora requiere instalarle otras aplicaciones para que funcione.
- Greenstone tiene mayor documentación que Fedora en castellano y no requiere otros componentes para funcionar.

El análisis de repositorios institucionales efectuados permitió seleccionar el software Greenstone por las siguientes razones:

- Es un software gratuito basado en la filosofía Open Source.
- Es de fácil instalación y operación.
- Tiene documentación y es operado en castellano.
- Su interfase es amigable.
- Es una herramienta flexible.

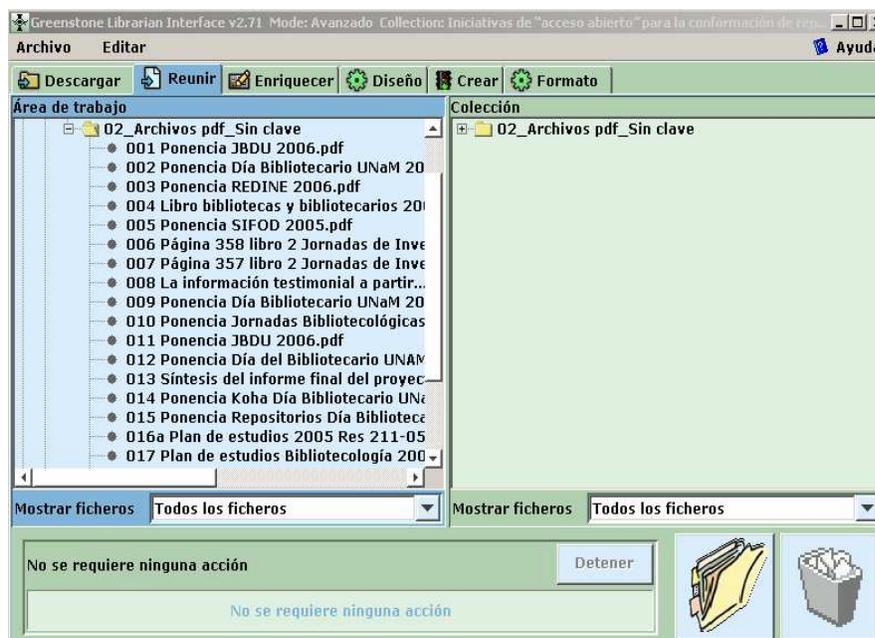
- La funcionalidad cubre las necesidades planteadas en el proyecto.
- Soporta los metadatos de Dublin Core y otros.
- Es el software de mayor utilización en nuestro país para la conformación de repositorios institucionales.

Aplicación del Software Greenstone y el conjunto de metadatos Dublin Core a la creación del repositorio institucional

El recurso electrónico logrado en esta experiencia es la demostración práctica de aplicar las alternativas del modelo de “acceso abierto” representado en la **creación de un repositorio de documentos relacionados con la labor del campo disciplinar de la Biblioteconomía**. Constituye un modelo propio, basado en componentes normalizados y software libre; adecuado a las necesidades del contexto institucional cuya demostración de implementación y funcionamiento ha sido posible con la aportación de un corpus acotado de documentos institucionales propios. Se reseñan instrucciones aplicadas en la experiencia de creación del repositorio *Iniciativas de “acceso abierto” para la conformación de repositorios institucionales16/H-21*:

Parte I: instalación del software y creación de la nueva colección

La gestión del software y su instalación ha sido realizada a partir de su localización en la página oficial <http://www.greenstone.org>. Una vez ingresado al Software Greenstone el primer paso ha sido la creación de una nueva colección denominada *Iniciativas de acceso abierto para la conformación de repositorios institucionales16/H-213*, realizando una breve descripción acerca del contenido de la colección.



Interfaz de Greenstone

La figura anterior presenta a Greenstone una vez instalado y en proceso de creación de la colección; organizado en menús y 6 solapas cada una con una función diferente que hacen posible la creación de un repositorio institucional.

Parte II: reunión de los documentos dentro de la colección (solapa reunir): en esta instancia se ha procedido a incorporar los documentos previamente seleccionados para la conformación del repositorio, transformados todos a documentos en formato .pdf con previa autorización de los autores.

Parte III: selección de metadatos y llenado de etiquetas utilizando el modelo de metadatos Dublin Core (solapa Enriquecer): teniendo en cuenta los fundamentos teóricos presentados anteriormente respecto a la relevancia de los metadatos Dublin Core y considerando que el Software Greenstone permite gestionar dicho modelo entre varios tipos¹ según las necesidades de cada colección, se ha establecido trabajar con los elementos de este conjunto de metadatos.

En esta primera etapa se aplican los elementos del modelo de metadatos Dublin Core de la siguiente manera:

Etiqueta: DC. Title, Título	Se ingresa el nombre del recurso establecido por el/los autor/es. Ejemplo: “La instalación del paradigma de acceso abierto en la UNaM: un proyecto de investigación en curso”.
Etiqueta: DC. Subject, Materias y palabras clave	Se ingresa el tema del contenido del recurso expresado en palabras clave, utilizando el vocabulario controlado: Tesauro de la UNESCO y estableciendo un máximo de cinco tópicos por documento.
Etiqueta: DC. Creator, Autor	Se describe el/los nombre/s o del/de los autor/es principal/es del documento según establecen las Reglas de Catalogación Angloamericanas 2a ed. Revisadas. Ejemplo: Miranda, Mirta Juana; Kuna, Horacio Daniel; Prevosti, María Norma; García, Nélide Elba; Oria, Mónica.; Jaroszczuk, Susana Eunice
Etiqueta: DC. Publisher, Editor	Se describe el nombre de la persona u organización responsable de que el recurso se encuentre disponible. Ejemplo: “Universidad Nacional de Misiones”
Etiqueta: DC. Date, Fecha	Se describe la fecha asociada a la creación o modificación del recurso siguiendo la notación AAAA-MM-DD indicada por la norma ISO 8601. Ejemplo: “2003-09”

Algunas etiquetas Dublin Core utilizadas en cada documento descrito en Software Greenstone

¹Extracted Greenstone Metadata 1.0 (ex); Australian Government Locator Service Metadata Element Set, Version 1.3 (agls); Development Library Subset Example Metadata (dls); Greenstone metadata set (gs), etc.

Es importante destacar que GLI recuerda los metadatos ya asignados, es decir, que una vez ingresado un metadato para determinado campo el Software Greenstone lo registra y cada vez que se selecciona ese mismo campo para el ingreso de un metadato a un nuevo documento hace posible la selección en la lista de valores.



Parte IV: diseño de la presentación de la colección (solapa Diseño): se ha procedido a explotar esta herramienta definiendo los plugins, es decir, definiendo los índices de búsqueda y los clasificadores de exploración:

- Los plugins (convertidores) de documentos: se han seleccionado de acuerdo al formato de los documentos que conforman el repositorio.
- Índice de búsqueda: se han especificado las partes de la colección por las que se pueden efectuar búsquedas (campos autor, título y palabras claves).
- Clasificadores de Exploración: se han determinado los campos (puntos de acceso) para la creación de listados.

Parte V: creación de la colección (solapa Crear): se ha utilizado esta herramienta para diseñar la vista *previa de la colección*, compilando el proyecto completo -esto es- creando el OPAC al finalizar la creación.

Parte VI: creación del CD/DVD de la colección: se ha efectuado la exportación de la colección creada a un recurso electrónico CD-ROM utilizando la opción Exportar en el menú Archivo.

En resumen, estos han sido los pasos para crear la colección en Software Greenstone:

- Creación y definición de la colección.
- Recopilación/reunión de los documentos que la forman.
- Definición del conjunto de metadatos.
- Selección de los plugins para procesar los documentos.
- Determinación del tipo de búsqueda y los índices que se van a crear.
- Definición de los clasificadores que se ofrecen al usuario.
- Creación de los formatos de presentación.
- Inicio del proceso de creación.
- Creación del CD-ROM.

5. Referencias

Biblioteca Nacional de Australia. (2003, marzo). Directrices para la preservación del patrimonio digital. Recuperado el 07 de julio de 2008, de <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>

Budapest Open Access Initiative (BOAI). 2003. La Iniciativa de Acceso Abierto de Budapest. GeoTrópico online, 1 (1), 2003. p. 98-100, versión PDF. Disponible en: http://www.geotropico.org/files/PDF-Boai_Espanol_1-1.pdf

Candás Romero, Jorge. El papel de los metadatos en la preservación digital. En: El profesional de la información, v. 15, n. 2, marzo-abril 2006.

Cervantes Cruz, A., & Zurita Sánchez, J. M., & Miranda Quevedo, P. (2007). *Propuesta de normalización bibliográfica para la creación de metadatos basados en Dublin Core en los Repositorios Universitarios del proyecto 3R*. Ponencia presentada en el II Encuentro Nacional de Catalogación "En los umbrales de un nuevo código de catalogación", Ciudad de México. Recuperado el 05 de julio de 2008, de <http://eprints.rclis.org/archive/00011794/>

Codina, L. (2001, diciembre). Las propiedades de la información digital. *El profesional de la información*. Vol. 10, nº 12. Recuperado el 05 de julio de 2008, de <http://www.elprofesionaldelainformacion.com/contenidos/2001/diciembre/5.pdf>

Daudinot Founier, I. (2006). Organización y recuperación de información en Internet : teoría de los metadatos. *ACIMED*. Vol. 14, nº 5. Recuperado el 05 de julio de 2008, de http://bvs.sld.cu/revistas/aci/vol14_5_06/aci06506.htm

Dávila, J. A., & Nuñez, L. A., & Sandía, B., & Torrén, R. (2006, enero). *www.saber.ula.ve: un ejemplo de repositorio institucional universitario*. *Interciencia*, Vol. 31, nº 1. Recuperado el 07 de

julio de 2008, de <http://www.saber.ula.ve/bitstream/123456789/16545/1/saber-experiencias-repositorio.pdf>

Gradmann, Stefan. Catalogación Versus Metadata: Vino Viejo en Odres Nuevos? En: 64th IFLA General Conference August 16 - August 21, 1998.
http://fahu.uncoma.edu.ar/academica/materias/Tratamiento_de_la_informacion_I/Links/catalogacion_versus_metadata.doc

ISBD (G) : Descripción Bibliográfica Internacional Normalizada General : texto anotado / preparado por el Grupo de Trabajo del Comité de Catalogación de la IFLA ; traducción al español de María Cristina Guillén Bermejo. – Edición rev. – [Madrid] : ANABAD : Arco/Libros, 1993.

Méndez Rodríguez, E. M. (2003). Catalogación/Organización de documentos digitales: estado de la cuestión, tendencias y perspectivas desde España. *BIBLIODOC : Anuario de biblioteconomía, documentación e información*. Recuperado el 05 de julio de 2008, de <http://eprints.rclis.org/archive/00008088/>

Méndez Rodríguez, Eva Ma (2002). Metadatos y recuperación de información : estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón : Trea.

Muñoz de Solano & Palacios, B. (2006) El profesional de la información digital : aspectos organizativos. *Revista Anales de Documentación*. Vol.9. Recuperado el 07 de julio de 2008, de <http://eprints.rclis.org/archive/00007474/fullmetadata.html>

Reglas de Catalogación Angloamericanas. — 2a ed., rev. 2002, act. 2003. — Santafé de Bogotá : Rojas Eberhart Editores, 2004.

Rodríguez Bravo, B. (2007) Los repositorios de información : guardianes de la memoria digital. *Anales de Documentación*. N° 10. Recuperado el 05 de julio de 2008, de <http://eprints.rclis.org/archive/00011517/>

San Segundo, María Rosa (1998). Organización del conocimiento en Internet : metadatos bibliotecarios DUBLIN CORE. Disponible en: http://fesabid98.florida-uni.es/Comunicaciones/r_sansegundo.htm

Senso, J. A., & Rosa Pinero, A. (2003). El concepto de metadato : algo más que descripción de recursos electrónicos. *Ciência da Informação*. Vol.32, no.2. Recuperado el 05 de julio de 2008, de <http://www.scielo.br/pdf/ci/v32n2/17038.pdf>

Serra Serra, J. (2001, septiembre). Gestión de los documentos digitales : estrategias para su conservación. *El profesional de la información*. Vol. 10, n° 9. Recuperado el 05 de julio de 2008, de <http://www.elprofesionaldelainformacion.com/contenidos/2001/septiembre/1.pdf>

Siló, T. (2005). "Fundamentos Tecnológicos del Acceso Abierto: Open Archives Initiative y Open Archival Information System", *El Profesional de la Información*, Vol. 14, No. 5, pp. 365-380.

Thesauro de la UNESCO [en línea]. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. Disponible en: <http://databases.unesco.org/thessp/> [Consultado el 14/12/2007]

Zurita Sánchez, Juan M. Software libre : una alternativa para la gestión de recursos de información en bibliotecas. Recuperado el 10 de octubre de 2008, de http://eprints.rclis.org/archive/00013770/01/software_libre_alternativa.pdf