

Universidad de las Ciencias Informáticas  
Facultad No.1



*Título:*

*“Análisis de tecnologías para la automatización de catálogos y creación de bibliotecas digitales.”*

Trabajo de diploma para optar por el título de  
Ingeniero en Ciencias Informáticas

Autor:

Georgina Reyes Riquene

Tutores:

Ing. Maikel Manuel Fernández Fernández

Ing. Yanedi Abreu Bartomeo

La Habana, de Junio del 2011  
Año 53 de la Revolución



*“No debemos subestimar la utilidad de planear antes de actuar”  
Fidel Castro.*

## Declaración de autoría

---

Declaro ser autora de la presente tesis y reconozco a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

**Georgina Reyes Riquene**

\_\_\_\_\_

Firma del Autor

**Ing. Maikel Manuel Fernández Fernández**

\_\_\_\_\_

Firma del Tutor

**Ing. Yanedi Abreu Bartomeo**

\_\_\_\_\_

Firma del Tutor

## Datos de contacto

---

---

Ing. Maikel Manuel Fernández Fernández: Graduado de Ingeniería Informática en la Universidad de Holguín, año 2005. Profesor instructor, imparte las asignatura de Matemática III y Matemática IV. Su labor productiva e investigativa se centra en el campo de la Gestión de Contenidos Web. Ha participado en eventos y cuenta con publicaciones sobre este tema.

Ing. Yanedi Abreu Bartomeo: Graduado de Ingeniería en Ciencias Informáticas en la Universidad de Ciencias Informáticas (UCI), año 2007. Profesor instructor, imparte las asignatura de Programación III. Su labor productiva e investigativa se centra en el campo de la Gestión de Contenidos Web. Ha participado en eventos y cuenta con publicaciones sobre este tema.

## Agradecimientos

---

---

*Les agradezco con todo mi corazón a las personas más importantes en mi vida, los que desde muy chiquita me enseñaron a no rendirme ante nada por difícil que pareciera y a luchar por ser cada día mejor. Los que me han apoyado siempre y me han guiado hasta aquí: mis padres. Por enseñarme todo lo que soy y seré, por su sacrificio y dedicación y por hacer posible que yo llegara aquí.*

*A mi madre, tu sabes que te amo, gracias por estar siempre a mi lado, por ser mi fortaleza, por ayudarme a levantarme ante cada caída y enseñarme que en la vida siempre hay una salida y por confiar en mí. Por todos los sacrificios que has hecho para que yo este hoy aquí, por dedicarme tu vida entera. Por ser mi amiga, mi confidente, por tus consejos, en fin, porque tener una madre mejor, sería imposible. Gracias mamita.*

*A mi papá, quien siempre ha sido un ejemplo para mí y aunque no convivimos juntos siempre me ha apoyado y querido. Y el cual sé que está muy orgulloso de mí. Te quiero papito.*

*A mi abuelita Isabel, por todo su amor, cariño, comprensión, consejos. Por estar a mi lado desde chiquitica y haberme guiado siempre por el mejor camino, por apoyarme y confiar plenamente en mí, por ser una madre más para mí. Gracias por existir. Te quiero mucho.*

*A mi abuelo Rolando, que aunque no llevamos la misma sangre siempre lo he querido y me ha querido como una nieta. Por sus consejos, apoyo, comprensión y guía. Por estar a mi lado siempre que lo necesite y nunca dejar de confiar en mí.*

*A mi abuelita Cuca, que aunque en estos momentos no se encuentra con nosotros sé que ahora estaría muy orgullosa de mí. Gracias por haber existido y haberme dado tu apoyo y amor.*

*A mi hermanita Lili, que ha sido como una hijita para mí, por su amor, cariño y comprensión. Para la cual quiero y espero siempre ser un ejemplo. Y quiero que sepa que siempre puede contar conmigo. Te quiero.*

*A mis hermanos, Adrián y Miguel, por formar parte de mi vida y estar siempre conmigo aun cuando no convivamos juntos.*

*A mi padrastro Román y madrastra Isabel quienes me han apoyado y querido mucho en todo este tiempo. Gracias por su apoyo.*

*A mi tíos: Dania, Miguelito, Rolando, Martha quienes siempre me han apoyado y querido como una hija más. Por sus consejos, cariño y comprensión. Por haber estado a mi lado en las buenas y en las malas.*

*A mis primos que han sido unas hermanos para mí y a quienes quiero con la vida.*

*A toda mi familia en general. Gracias por ayudarme en todo momento, por confiar en mí y darme su cariño. Todo se lo debo a ustedes.*

*A mi novio Yunier, por estar a mi lado, por todo lo que me ha enseñado, por incitarme a superarme y estudiar cada día más, y sobre todo por quererme tanto. Por darme fuerzas para seguir, sin él no hubiera podido vencer esta batalla. Agradecer también a su familia que desde que me conocieron fueron muy cariñosos conmigo.*

*A Yani y Mailitza, mis amigas de la primaria, secundaria y Pre, que más que amigas siempre fueron unas hermanas para mí. Por confiar en mí, por siempre darme fuerzas para llegar a este día, porque a pesar del*

## Agradecimientos

---

---

*tiempo que nos pasamos sin vernos sé que hoy están felices por mí. Y a las hijitas de Yani, quienes son mis sobrinitas. Las quiero mucho.*

*A Lisandra que durante estos cinco años, más que una amiga ha sido una hermana y una madre para mí. Gracias por tu apoyo, comprensión y por haberme aguantado todo este tiempo.*

*A todos mis amigos y compañeros de aula por haber compartido estos 5 años de estudio, por enseñarme el valor de la amistad. En especial a Yailin, Delvis, Sachy, Diana, Yeni, Yurita, Mylen, Lisandra, Rolando, Julio, en fin a todos y cada uno de ustedes que estuvieron a mi lado todo este tiempo.*

*A todos los chicos del 150 304, por haberme soportado en este último año y siempre brindarme su apoyo.*

*A mis tutores por todo el apoyo y la preocupación. Por todo el tiempo que me dedicaron siempre que los necesite. Muchas Gracias.*

*A todos mis profesores, quienes han hecho de mí la profesional que soy y seré siempre. Gracias por su dedicación durante todos estos años. En especial a Kelvis que más que un profesor fue para mí un padre. Gracias a todos.*

*En fin, a todos los que de una forma u otra me han ayudado en el transcurso de estos 5 años y han contribuido al desarrollo de este trabajo.*

*A la Revolución y al comandante en jefe Fidel Castro Ruz por brindarme la posibilidad de ampliar mis conocimientos en esta gran universidad que es la Universidad de las Ciencias Informáticas (UCI).*

## Dedicatoria

---

---

*Esta tesis va dedicada a las personas que más quiero. En especial a mis padres por su confianza, amor, apoyo y cariño incondicional en todo momento. A los cuales les debo todo lo que fui, soy y seré.*

*A mis hermanos por estar siempre conmigo, en las buenas y las malas. Y a los cuales espero servirle de ejemplo.*

*A mis abuelos que siempre estuvieron en todos los momentos dispuestos a tenderme una mano, y a guiarme siempre, depositando en mí toda su confianza.*

*A mis tíos por sus consejos y dedicación, que a cada paso que daba siempre tienen una palabra o un consejo que darme.*

*A mis primos quienes siempre han sido unos hermanos más para mí.*

## Resumen

---

---

La Web, prácticamente como una entidad en desarrollo, crece día a día excediendo la capacidad humana de poder organizarla. La recuperación de la información se vuelve compleja y difícil de manejar, si no se accede a resultados de búsquedas de información más precisas y acotadas.

En el contexto anterior nacen las bibliotecas digitales, las cuales ya no son un paradigma ni un desafío, sino que simplemente son una de las principales realidades en las que organizaciones y empresas deben poner su atención para resguardar y difundir el capital intelectual propio de las mismas.

El desarrollo de una biblioteca digital requiere de un ambiente con una serie de recursos que se ponen a disposición de los desarrolladores para llevar a cabo su tarea. Este entorno de desarrollo debe contar tanto con los recursos materiales como herramientas de software adecuadas a ser utilizadas durante el proceso de elaboración.

En este trabajo se hace un estudio del estado del arte de los repositorios y bibliotecas digitales en el mundo actualmente y de las estrategias para su implementación. También se realiza un análisis y pruebas a las principales herramientas que existen a nivel mundial para el desarrollo de los repositorios y bibliotecas digitales. Se tienen en cuenta diferentes experiencias de proyectos en el desarrollo e implementación de los mismos, de los cuales se tomaron algunas ideas relacionadas con la forma en que efectúan estos procesos.

Como resultado de la investigación se propone el uso de las herramientas analizadas para cada escenario tecnológico, así como la integración de al menos una de ellas con otros tipos de herramientas.

### **PALABRAS CLAVE**

**Bibliotecas Digitales, Escenario, Herramientas, Información, Repositorios Digitales.**



# Índice de contenidos

---

---

Introducción.....	1
Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales .....	5
1.1 Introducción.....	5
1.2. Biblioteca Digital.....	5
1.2.1. Conceptos.....	5
1.2.2. Antecedentes .....	6
1.2.3. Definición de Biblioteca Digital .....	7
1.2.4. Impacto de la biblioteca digital en la gestión y el acceso a la información.....	9
1.2.5. Componentes de la Biblioteca Digital .....	10
1.3. Repositorios Digitales .....	11
1.3.1. Evolución y desarrollo de los Repositorios Digitales .....	12
1.3.2. Tipos de Repositorios Digitales .....	12
1.3.2.1. Repositorios Institucionales .....	12
1.3.2.2. Repositorios Temáticos.....	14
1.3.2.3. Repositorios de Objetos de Aprendizaje .....	14
1.3.3. Análisis sobre la existencia de Repositorios a nivel mundial.....	17
1.3.3.1. África.....	17
1.3.3.2. Asia .....	17
1.3.3.3. Australia.....	18
1.3.3.4. Europa .....	18
1.3.3.5. América del Sur .....	20
1.3.3.6. América del Norte .....	20
1.3.3.6. Cuba .....	21
1.4. Gestión de documentos y colecciones.....	21
1.4.1. Protocolo OAI-PMH (Open Archives Initiative - Protocol Metadata Harvesting) .....	22
1.5. Herramientas para el desarrollo de catálogos y bibliotecas digitales.....	24
1.5.1. Dspace.....	24

# Índice de contenidos

---

---

1.5.2. Eprints.....	24
1.5.3. Greenstone.....	25
1.5.4. Drupal.....	25
1.6 Ejemplos de repositorios y bibliotecas digitales desarrolladas con estas herramientas.....	25
1.6.1 Repositorio Institucional de la Universidad Andina Simón Bolívar.....	25
1.6.2 Biblioteca Nacional de Cuba José Martí.....	26
1.6.3 Red de Bibliotecas Virtuales de Ciencias Sociales de CLACSO.....	27
1.6.4 ECURED.....	28
1.6.5 E-Prints Complutense.....	29
1.6.6 Repositorio de Tesis de Doctorado en Ciencias Biomédicas y de la Salud de Cuba.....	30
1.7 Conclusiones del capítulo.....	31
Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales.....	32
2.1 Introducción.....	32
2.2 Características de las herramientas.....	32
2.2.1 DSpace.....	32
2.2.2 E-Prints.....	37
2.2.3 Greenstone.....	41
2.2.4 Drupal.....	44
2.5 Comparación de las herramientas.....	48
2.6 Definición de posibles escenarios.....	50
2.6.1 Escenario en Java.....	50
2.6.2 Escenario de un repositorio.....	52
2.6.3 Escenario de una biblioteca digital con iteración de usuarios.....	55
2.6.4 Escenario en PHP.....	58
2.7 Conclusiones del capítulo.....	59
Conclusiones Generales.....	61
Recomendaciones.....	61

# Índice de contenidos

---

---

Referencias Bibliográficas .....	63
Glosario de Términos.....	67
Anexos .....	<b>¡Error! Marcador no definido.</b>

## Introducción

Durante la mayor parte de su historia, los sistemas de información han existido en el mundo de lo impreso. Todos los sistemas de gestión se basaban en un mundo en el que la información era estática y lineal. Estaba pensada para ser utilizada por las personas en lugares concretos. (ESCOLAR, 1990)

Desde las tabletas de arcillas usadas por la civilización mesopotámica hasta nuestros días, la información ha sido organizada y preservada por la sociedad, pues se necesitó consignarla y registrarla para el mantenimiento del estado, la conservación y comunicación de las creencias religiosas, la educación y la transmisión de la cultura a las nuevas generaciones (BEITRA, 2008). A partir de esto surgió la necesidad de crear un sitio donde se pudiera almacenar y preservar toda esta información y que a su vez sirviera como un sistema de información para todos. Como respuesta a esto surgieron las bibliotecas.

La historia del libro y las bibliotecas es la historia de la humanidad. Hablar de una biblioteca es referirse a uno de los instrumentos más antiguos del hombre para “*salvaguardar los testimonios de su paso por el mundo*”. La Real Academia de la Lengua define una biblioteca, como:

*“Colección de libros o tratados análogos o semejantes entre sí, ya por las materias de que tratan, ya por la época y nación o autores a que pertenecen.”*

Las tareas de una biblioteca se realizaron desde siempre manualmente, pero con la aparición de las nuevas tecnologías de la información y la comunicación (TICs), y en particular de Internet, surgió la posibilidad de automatizar sus acervos.

En pocos años, la información digital ha pasado de tener una presencia marginal a ocupar un lugar privilegiado en el proceso informativo.

A partir de la figura de una biblioteca tradicional han surgido en el transcurso de los años otros tipos de bibliotecas que hacen uso de la tecnología, cada una con sus características. Las bibliotecas automatizadas fueron las primeras en aparecer bajo estos nuevos lineamientos tecnológicos, luego las bibliotecas electrónicas, más tarde surgieron las bibliotecas virtuales y las bibliotecas híbridas. Actualmente se ve el crecimiento y la evolución de las bibliotecas digitales.

Las bibliotecas digitales o repositorios deben estar contruidos sobre sistemas y plataformas convencionales que permitan la compatibilidad e integración con otros estándares y procedimientos integrados. Para estas instituciones es necesario y de vital importancia contar con un software a la medida

## Introducción

---

de sus actividades, que les permita potencializar los servicios e impulsar a la biblioteca hacia esta realidad social, la sociedad de la información o del conocimiento.

El desarrollo de sistemas de información tipo bibliotecas trae consigo una serie de elementos teóricos y conceptuales que deben ser dominados para enfrentar este tipo de proyectos. Existe un grupo de procesos que tienen lugar dentro de estos sistemas que también se requiere de su interpretación para poder desarrollarlos. Entre estos procesos se encuentran los procesos de importación y conservación de la información. La disponibilidad de tecnologías que faciliten este tipo de servicios es una de las necesidades más acuciantes para conseguir extender los servicios y la información disponible a todos los usuarios.

Para el desarrollo de catálogos, así como para la gestión de documentos existen un grupo de herramientas informáticas ya desarrolladas, cada una con sus características. La mejor solución siempre estará condicionada por el entorno y muchas veces no es el empleo de una de estas herramientas por sí sola, sino, la integración con otros tipos de sistemas, como pueden ser los sistemas de gestión de contenidos. La planeación, la organización y otras tareas que plantean la construcción y desarrollo de una colección de documentos digitales superan largamente los problemas relacionados con la mera digitalización de fondos, esto obliga a probar y analizar directamente las diferentes aplicaciones existentes teniendo en cuenta el entorno donde se va a desarrollar. La tecnología, la naturaleza del problema que se enfrenta, los costos, el tiempo y las personas son factores que inciden a la hora de seleccionar cuales sistemas utilizar. La decisión es una tarea complicada, decisiva y de obligado cumplimiento, que condicionará el éxito o fracaso del proyecto.

Por lo anteriormente expuesto se plantea el siguiente **problema científico**: ¿Cómo lograr un entorno tecnológico efectivo para desarrollar bibliotecas digitales? A partir de esta problemática se definió como **objeto de estudio** de la investigación: Los sistemas de información. El cual se enmarca en el **Campo de acción**: Los sistemas de información tipo bibliotecas. Quedando definido como **objetivo general**: Proponer entornos tecnológicos por escenarios para el desarrollo de bibliotecas digitales. Derivándose como **objetivos específicos** para esta investigación:

- Realizar el estado del arte de los sistemas de información tipo bibliotecas.
- Realizar prueba de herramientas para la gestión de repositorios y bibliotecas digitales.
- Integrar herramientas de gestión de repositorios y bibliotecas digitales con un sistema de gestión de contenidos.
- Proponer una solución mediante la integración de sistemas.

# Introducción

---

---

Con el propósito de dar cumplimiento a los objetivos anteriormente planteados se definen las siguientes **tareas de investigación:**

- Revisar la bibliografía sobre los temas relacionados con el desarrollo de sistemas de información tipo bibliotecas.
- Definir los repositorios digitales como herramientas para la creación de colección de documentos digitales.
- Identificar las principales tecnologías que se pueden emplear para el desarrollo de repositorios y bibliotecas digitales.
- Caracterizar las tecnologías identificadas.
- Realizar pruebas prácticas sobre las tecnologías identificadas.
- Caracterizar los principales procesos que tienen lugar dentro de una biblioteca digital.
- Definir posibles escenarios.
- Proponer las tecnologías y sus relaciones para cada escenario y proceso.

La investigación se sustenta en la **idea a defender** que: una correcta selección de entornos tecnológicos propiciará eficiencia en la creación de bibliotecas digitales.

Los métodos de investigación que se han aplicado en la realización de este trabajo son:

## **Métodos teóricos:**

- Analítico-sintético: Este método se usa para el análisis por partes de la información recopilada en los medios bibliográficos y la integración de los resultados para la solución.
- Histórico-lógico: Se emplea para el estudio y profundización de la evolución de las tecnologías para el desarrollo de los repositorios y las bibliotecas digitales.

## **Métodos empíricos:**

- Observación: Este método se usa porque permite investigar los repositorios y las bibliotecas digitales existentes para llegar a la esencia de los mismos, lo cual sirve de ayuda para realizar un seguimiento del comportamiento de estos, sirviendo como guía para conocer si se retrocede o se avanza hacia el objetivo final.

El presente trabajo de diploma está estructurado en dos capítulos:

# Introducción

---

---

**Capítulo 1. Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales:** Brindar un acercamiento a los principales conceptos asociados al dominio del problema y que son abordados a lo largo del trabajo, así como el estado del arte de los sistemas de información tipo bibliotecas.

**Capítulo 2. Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales:** En este capítulo se presenta las propuestas de las tecnologías y sus relaciones para cada escenario y proceso para la automatización de catálogos y desarrollo de bibliotecas digitales.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

---

## **Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales**

### **1.1 Introducción**

El presente capítulo tiene como objetivo, tratar los principales conceptos y aspectos más significativos relacionados con las principales temáticas abordadas en diferentes fuentes bibliográficas acerca del estado del arte. Se proporciona una panorámica general sobre las bibliotecas digitales, sus antecedentes, el impacto causado por ellas en la gestión y el acceso a la información, así como, sus componentes y características. Además, menciona el concepto de repositorios digitales, los tipos de repositorios digitales, su existencia y desarrollo en el mundo.

### **1.2. Biblioteca Digital**

#### **1.2.1. Conceptos**

Para hablar de las bibliotecas digitales antes se debe hacer mención de algunos conceptos y términos importantes, los cuales ayudarán a lograr un mejor entendimiento de las mismas. Lo primero es aclarar el concepto de sistema de información, documento, documentación, comunicación, información y biblioteca, ateniéndose a las definiciones que da la Norma UNE 50 113 (PÉREZ, 1995):

- Sistema de información: Sistema de comunicación que permite comunicar y tratar información.
- Documento: Información de cualquier índole, registrada en un soporte y que puede considerarse como una unidad en un proceso de documentación.
- Documentación: Se refiere a la recogida y tratamiento de información registrada, de forma continua y sistemática y que permite su almacenamiento, recuperación, utilización y transmisión.
- Comunicación: Transferencia de significado, mediante la transmisión de señales.
- Información: Hechos que se comunican. Mensaje utilizado para representar un hecho o un concepto en un proceso de comunicación, con el fin de aumentar los conocimientos.
- Biblioteca: Cualquier colección organizada de libros y publicaciones en serie impresos, u otros tipos de documentos gráficos o audiovisuales, disponibles para préstamo o consulta.



# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

## 1.2.2. Antecedentes de las bibliotecas digitales

Diversos factores, relacionados más bien con necesidades prácticas y objetivas, como el incremento sostenido de los volúmenes de información, los límites humanos para su procesamiento, la diversidad de formatos portadores de contenidos valiosos, así como las dificultades para su recuperación; llevaron a que investigadores, especialistas y técnicos del ámbito bibliotecario, la computación y las telecomunicaciones, asumieran el reto de representar en el medio electrónico los sistemas desarrollados tradicionalmente por las bibliotecas con el propósito de satisfacer las necesidades de información de los usuarios.

Desde la década de los sesenta surgió un proceso paulatino de automatizar o transformar a versión electrónica/digital diversos procesos, colecciones y servicios en las bibliotecas, lo que ha dado como resultado un devenir histórico-social en el que han surgido diversos conceptos tales como biblioteca automatizada, biblioteca electrónica, bibliotecas virtual, biblioteca híbrida y biblioteca digital.

Primeramente, se automatizaron los catálogos bibliográficos, con el propósito de agilizar los procesos técnicos. Luego se automatizaron los procesos de préstamo de material. Así nacieron las llamadas bibliotecas automatizadas.

Posteriormente surgió el término de biblioteca electrónica, con énfasis en el acceso a fuentes de información electrónica y la evolución de los documentos hacia un formato electrónico. Por lo tanto, la biblioteca electrónica es una entidad individual que tiene una colección documental física y electrónica, cuenta con sistemas automatizados para ofrecer servicios ágiles y eficientes, así como sistemas de acceso en red.

Luego surgió la idea de biblioteca virtual, concebida como un sistema que hace uso de la realidad virtual mediante interfaces y simula el ambiente de una biblioteca real. Emplea tecnología multimedia de vanguardia y puede guiar al usuario a través de diferentes sistemas para encontrar colecciones en diferentes sitios, conectados a través de sistemas de cómputo y telecomunicaciones. (CABRAL, RAMÍREZ Y FIGUEROA, 2007)

Después surge el término de biblioteca híbrida que combina lo mejor de ambos mundos, el físico y el digital. El reto de la biblioteca híbrida consiste en integrar el acceso a toda clase de recursos, impresos, analógicos y digitales. De esta forma, una biblioteca híbrida proporcionaría un entorno y unos servicios, en partes físicos y en parte virtuales. En los años 90, el mundo de las bibliotecas da nacimiento al concepto de bibliotecas digitales.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

## 1.2.3. Definición de Biblioteca Digital

Son varias las definiciones que existen de biblioteca digital que pueden ayudar a comprender mejor el término como la formulada por la Federación de Bibliotecas Digitales (DLF draft strategy and business plan, 2001), quien la define como:

*“Organizaciones que proveen los recursos, incluyendo personal especializado, para seleccionar, estructurar, ofrecer acceso intelectual, interpretar, distribuir, preservar la integridad y asegurar la persistencia a lo largo del tiempo, de colecciones de obras digitales, de tal forma que estén oportuna y económicamente disponibles para su uso por parte de una comunidad definida o conjunto de comunidades.”*

Para el doctor Juan Voutssás Márquez (2006) una biblioteca digital es *“una biblioteca que ofrece documentos y servicios electrónicos”*, aunque sugiere una segunda reflexión que califica él mismo como simple y concisa, que es la del Taller de Planeación de Santa Fe Acerca de Medio Ambiente de Trabajo de Conocimiento Distribuido:

*“El concepto de biblioteca digital no es simplemente el equivalente de colecciones digitalizadas con herramientas de manejo de información. Es más bien un ambiente digital para integrar colecciones, servicios y personas en apoyo a un ciclo vital de creación, diseminación, uso y preservación de datos, información y conocimiento.”*

Christine Borgman las describe como *“un conjunto de recursos y de capacidades técnicas para crear, buscar y utilizar la información”*, pero subraya también que *“las bibliotecas digitales están concebidas y organizadas para (y por) una comunidad de usuarios, y sus medios son utilizados para responder a las necesidades de información y a los usos de esta comunidad”*. (HERVÉ LE CROSNIER, 2006)

Otro enfoque es el propuesto por el Virginia Tech Institute en 1998: *“una biblioteca digital no se resume en una colección digitalizada y en herramientas de gestión de la información. Es también un conjunto de actividades que ponen en relación las colecciones, los servicios y los usuarios a todo lo largo del ciclo de creación, difusión, uso y conservación de datos, información y conocimientos”* (HERVÉ LE CROSNIER, 2006)

Ahora bien, el término biblioteca digital se entiende de diversas maneras, y en el transcurso de los años han ido surgiendo múltiples definiciones, sin embargo, se puede identificar cinco características que son comunes a todas ellas (ORTIZ Y MOSCOSO, 2006):

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

- La biblioteca digital requiere hacer uso de las más nuevas tecnologías con el fin de unir e integrar los recursos de muchas bibliotecas.
- Los enlaces entre las bibliotecas digitales y los servicios de información son transparentes para el usuario final.
- El objetivo fundamental de una biblioteca digital es el acceso universal a la información.
- Las colecciones de las bibliotecas digitales no se limitan a documentos o sustitutos de los mismos, sino que se extienden a objetos digitales que no pueden representarse o distribuirse en formatos impresos.
- La biblioteca digital tiene que proporcionar un acceso coherente a un repositorio de información y conocimiento.

Las bibliotecas digitales tienen componentes peculiares muy interesantes, ya que han sido el punto de atención de diversas disciplinas y áreas de investigación. Al respecto, Nürnberg en 1995 explica algunos enfoques:

- Desde el punto de vista computacional y de la recuperación de la información, es una gran base de datos.
- Para los individuos que trabajan en tecnología de hipervínculos, es una aplicación particular de métodos de hipertexto.
- Para aquellos que trabajan en entrega de información a través de redes, es una aplicación de la web.
- Para los programadores y desarrolladores de aplicaciones de cómputo, son colecciones de algoritmos computacionales o piezas de programación, reunidas en un solo sitio.
- Para vendedores de bases de datos, revistas y documentos electrónicos, sus productos son bibliotecas digitales.
- Para las grandes corporaciones, una biblioteca digital es el sistema de administración de documentos que controla la información de sus negocios en formato electrónico.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

- Para la bibliotecología, es otro paso en la constante automatización de las bibliotecas, que comenzó hace más de veinticinco años.

La mayor ventaja que se ha visto de las bibliotecas digitales sobre las bibliotecas tradicionales es que permiten el acceso universal a los recursos, sin limitantes de tiempo ni espacio, dado que principalmente se hace uso de medios como Internet y la Web para interactuar con los usuarios y otros sistemas con los que se pueda compartir la información.

El uso de las bibliotecas digitales brinda una serie de beneficios (ORTIZ Y MOSCOSO, 2006): 1) rapidez en la obtención de resultados, 2) almacenamiento de grandes volúmenes de información, y 3) facilidad para encontrar información adecuada y/o actualizada por parte de científicos, investigadores, profesionales y estudiantes.

## **1.2.4. Impacto de la biblioteca digital en la gestión y el acceso a la información**

En el modelo de biblioteca digital el concepto de acceso adquiere una relevancia fundamental, ya que, en cierto sentido, se convierte en el pilar fundamental para su creación.

Si se piensa en el modelo de biblioteca tradicional, su colección se determina, esencialmente, por la propiedad y el depósito físico de los fondos. Por el contrario, en las bibliotecas digitales los recursos que forman la colección no están físicamente presentes, lo que brinda la posibilidad de acceder, tanto técnica como económicamente, a ellos.

Su sistema de adquisiciones se basa en un conjunto de criterios cuya finalidad es seleccionar objetos digitales de un universo de información muy amplio, objetos digitales cuya selección debe regirse por los criterios de unicidad, coherencia, utilidad y accesibilidad.

Así, los objetos digitales deben ser susceptibles de identificarse con independencia del formato o del medio. Del mismo modo, el contenido y volumen de la información que contienen deben guardar una relación de coherencia y utilidad con respecto a los objetivos, fines y usos de la biblioteca digital. Por último, estos objetos deben ser accesibles a través de sistemas bibliotecarios, lo que implica un proceso de descripción y análisis de los mismos.

Una biblioteca digital se distinguirá no sólo por el volumen de su colección, sino también por el tipo de tecnologías implementadas y la diversidad de servicios para acceder a los documentos existentes en las redes y a fuentes de información remotas. Tecnologías y servicios determinarán la agilidad y facilidad con la que el usuario puede llevar a cabo la actividad que requiere y la búsqueda de información que precisa.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

## 1.2.5. Componentes de la Biblioteca Digital

La biblioteca digital va más allá de una simple compilación de documentos e información subida a la red. Son sistemas automatizados que brindan una serie de servicios para la compartición de información u otros elementos de interés al usuario. Para ello utilizan una serie de componentes necesarios para su correcto funcionamiento, entre los cuales se pueden citar (CABRAL, RAMÍREZ Y FIGUEROA, 2007):

- Contenido: Conformar colecciones de alta calidad, es decir, documentos digitales rigurosamente revisados y seleccionados.
- Preservación: Introducir tecnología digital a la biblioteca, implica el desarrollo de infraestructura para hacer accesibles los materiales que desean preservarse y así implementar el acceso digital a obras de gran importancia.
- Flexibilidad: Los elementos digitales permiten la maleabilidad y la integración, por lo que se garantiza la adaptación a los constantes cambios provocados por la rápida evolución del entorno digital.
- Trabajo interdisciplinario: De éste depende en gran medida el éxito de las bibliotecas digitales, por la importancia que posee para el crecimiento y el buen funcionamiento de la biblioteca. La labor profesional de los bibliotecólogos, ingenieros, especialistas en cómputo, administradores; es esencial para crear y adaptar las nuevas tecnologías de manera inteligente.
- Formatos: Para diferentes tipos de objetos digitales existen diversos formatos pero es fundamental elegir los más idóneos y frecuentemente empleados a nivel mundial. Otro aspecto a considerar es que sean interoperables con otros formatos y objetos digitales.
- Metadatos: Debido a la necesidad de identificar los contenidos de las colecciones digitales para facilitar la búsqueda y recuperación efectiva de la información surgieron los metadatos, que no son otra cosa que datos acerca de los datos, es decir, datos que describen cierto tipo de información, tales como autor, título, fecha y otros.
- Servicios de valor añadido: Productos y facilidades creados para dar valor al contenido de la colección, adecuados a las necesidades y a los requisitos de sus usuarios. El acceso simple no se considera un servicio.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

- Personalización: Funcionalidad para que el usuario (o institución) pueda definir su espacio de interacción con la biblioteca digital y seleccionar en listas propias los elementos de la colección.
- Ciclo de vida de la información: Los contenidos digitales pueden tener fases diferentes en sus diversas etapas y debe llevarse un seguimiento del ciclo de cada recurso.

Además, debe realizar cinco funcionalidades básicas: digitalización, implementación de grandes depósitos (repositorios), transferencia rápida de datos, privilegios de conexión y gestión.

El desarrollo de la colección de documentos digitales es uno de los aspectos más importantes que se llevan a cabo en las bibliotecas digitales, para lo cual se utiliza los repositorios digitales.

## 1.3. Repositorios Digitales

Los repositorios digitales son herramientas eficaces para almacenar, organizar y hacer uso eficiente de la información y el conocimiento. Constituyen un sitio centralizado donde se almacena y mantiene información digital, habitualmente bases de datos o archivos informáticos. Los repositorios están preparados para distribuirse habitualmente sirviéndose de una red informática como Internet. Y pueden ser de acceso público, o pueden estar protegidos y necesitar de una autenticación previa. Los repositorios más conocidos son los de carácter académico y los institucionales. (MUÑOZ, 2008)

Son el resultado de la evolución de los criterios y esfuerzos de organizaciones tradicionales. Es una herramienta de gestión de contenidos que tiene la calidad de brindar servicios para la gestión de esa producción depositada.

Por su parte, López Medina en el 2007 define repositorio digital como un sistema en red formado por hardware, software, datos y procedimientos, con las siguientes características:

- Contiene objetos digitales.
- Contiene metadatos.
- Asegura la identificación persistente del objeto mediante un identificador único persistente.
- Ofrece funciones de gestión, archivo y preservación de los objetos.
- Proporciona un acceso fácil, controlado y estandarizado a los objetos.
- Ofrece los sistemas adecuados de seguridad para los objetos y los metadatos.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

- Es sostenible en el tiempo.

Los beneficios y ventajas que estos depósitos abiertos aportan al mundo científico, hacen que se manifiesten como una gran revolución en el campo de la investigación y su existencia esté avalada y apoyada por multitud de instituciones de todos los países. Permiten la recuperación, reutilización y preservación de los resultados de la investigación, además de favorecer la difusión y la visibilidad de la producción científica, garantizando de forma efectiva el avance de la ciencia.

## 1.3.1. Evolución y desarrollo de los Repositorios Digitales

El primer repositorio en crearse fue *ArXiv*, fundado por Paul Ginsparg en 1991 en Los Álamos, Estados Unidos, para la Física de Altas Energías, las Matemáticas y las Ciencias de la Computación. Actualmente, contiene alrededor de 300 000 trabajos y se utiliza ampliamente por investigadores de todos los continentes. En estos momentos, se administra desde la Universidad de Cornell. Su éxito lo ubica como el modelo de difusión científica más efectivo en el Movimiento de Acceso Abierto.

En 1996, se creó REPEN (*Research Papers for Economics*), una iniciativa para crear una base de datos de acceso público en economía y disciplinas relacionadas; en 1997, *CogPrints*, fue desarrollado por *Steven Harnad* en la Universidad de Southampton, Reino Unido, en el área de psicología, neurociencias y lingüística. En el área de las ciencias biomédicas, se encuentra *PubMed Central*, creado en el 2000 a raíz de la iniciativa de *Harold Varmus*. Hasta mayo de 2007, el *OpenDOAR* recogía 881 repositorios en todo el mundo; el 80 % de ellos se clasifican como institucionales. (FLORRES Y SÁNCHEZ, 2007)

## 1.3.2. Tipos de Repositorios Digitales

### 1.3.2.1. Repositorios Institucionales

Un repositorio institucional (RI) es el conjunto de servicios ofrecidos por una institución con el objetivo de gestionar, difundir, y facilitar el acceso a cualquier persona con conexión a Internet, a los documentos electrónicos creados por dicha institución y por las personas que forman parte de ella. Actualmente, se centran en una organización (universidad, departamento, instituto, sociedades científicas). Es posible definir políticas para que los miembros añadan contenidos. En esta clasificación, también se incluyen los repositorios de tesis doctorales.

Lynch en el 2003 define repositorio institucional como un conjunto de servicios que una institución ofrece a su comunidad para la gestión y difusión de los contenidos digitales generados por los miembros de esa

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

comunidad. Es, en su nivel más básico, un compromiso organizativo para el control de esos materiales digitales, incluyendo su preservación, su organización, acceso y distribución.

López Medina en el 2007 otorga a los repositorios institucionales las siguientes funciones:

- Es una herramienta común de gestión de contenidos digitales de la institución y de apoyo a la investigación y el aprendizaje.
- Es un vehículo proactivo del “Open Access”, coincidiendo aquí con Suber.
- Es un lugar de almacenamiento y preservación.

Por su parte Melero en el 2005, les confiere funciones como la de servir de perfil, sello de calidad y marca de identidad de la institución, así como la de favorecer la difusión, la visibilidad, la preservación y aumentar el impacto de la producción científica.

Otra cualidad técnica importante de los repositorios institucionales, es la interoperabilidad o capacidad de intercambiar datos, a pesar de que cada uno esté creado bajo un programa informático u otro y presente características formales diferentes.

Las características que debe tener un repositorio institucional son:

- Que sea de naturaleza institucional.
- Debe tener carácter científico acumulativo y perpetuo.
- Debe ser de carácter abierto e interoperable con otros sistemas.

Los beneficios que se obtienen a partir de un repositorio institucional son:

- Permitir que se eleve la visibilidad de las investigaciones de la institución.
- Las organizaciones científicas de todo el mundo obtienen acceso fácil y rápido a investigaciones realizadas en otras instituciones.
- Aumenta la proyección del autor.
- Favorecen la preservación de investigaciones producidas



# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

## 1.3.2.2. Repositorios Temáticos

Fueron los primeros repositorios en aparecer. Se crean en torno a una disciplina; por lo que se dice que se especializan en torno a una disciplina y pueden contener la más variada tipología de documentos. El acceso al repositorio es directo. No hay forma de obligar a los autores a remitir sus trabajos. Los repositorios temáticos almacenan y proporcionan acceso a la producción académica de un área temática particular.

Los repositorios temáticos recogen documentos científicos y/o académicos de una o varias disciplinas científicas específicas y son los investigadores de diversas instituciones quienes lo contribuyen autoarchivando sus trabajos. Como ejemplos se pueden mencionar ArXiv (Física, Matemática, Computación y ciencias afines), CogPrints (Psicología), REPEC (Economía), E-Lis (Bibliotecología y Ciencias de la Información), entre otros.

## 1.3.2.3. Repositorios de Objetos de Aprendizaje

Los Repositorios de Objetos de Aprendizaje(ROA) surgen de la necesidad de compartir Objetos de Aprendizaje(OA) que provienen de distintas fuentes y organizar su almacenamiento con el objetivo final de potenciar su reutilización, sin embargo, no siempre estos contienen objetos de aprendizaje propiamente dichos, sino que también incluyen otro tipo de contenido.

El programa Canarie dice que los ROA “*son un catálogo electrónico/digital que facilita las búsquedas en Internet de objetos digitales para el aprendizaje*”. Daniel, a partir de los términos repositorio digital, objeto de aprendizaje y metadato dice que “*los repositorios de objetos de aprendizaje son bases de datos con búsquedas que alojan recursos digitales y/o metadatos que pueden ser utilizados para el aprendizaje mediado*”. El JORUM+ Project adopta la siguiente definición: “*Un ROA es una colección de OA que tienen información (metadatos) detallada que es accesible vía Internet. Además de alojar los OA los ROA pueden almacenar las ubicaciones de aquellos objetos almacenados en otros sitios, tanto en línea como en ubicaciones locales*”. (LÓPEZ Y GARCÍA, 2006)

Las definiciones, en su sentido general no difieren mucho entre sí y dejan ver claramente que estos repositorios, sean bases de datos o catálogos, están creados para ser utilizados en un proceso de enseñanza, lo cual lleva a que los ROA se vean como facilitadores claves para incrementar el valor de los recursos de aprendizaje dando la oportunidad de reutilizar, reorientar y hacer reingeniería para cubrir las necesidades del usuario final.

## Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

Por la forma en la que se concentran los recursos, principalmente se identifican dos tipos de ROA (LÓPEZ Y GARCÍA, 2006):

- Los que contienen los objetos de aprendizaje y sus metadatos, en éstos los objetos y sus descriptores se encuentran dentro de un mismo sistema e incluso dentro de un mismo servidor.
- Los que contienen sólo los metadatos, en este caso el repositorio contiene sólo los descriptores y se accede al objeto a través de una referencia a su ubicación física que se encuentra en otro sistema o repositorio de objetos.

También es común encontrar repositorios mixtos, en los que se hace una combinación de estos dos tipos mencionados.

Los ROA más conocidos comúnmente funcionan de forma independiente (*stand-alone*). Son aplicaciones con una interfaz web, un mecanismo de búsqueda y listados con algún tipo de clasificación. Otras clases de ROA operan sólo como módulos adicionales a otros productos que utilizan los contenidos de forma exclusiva y sin que el usuario tenga acceso directo al repositorio. Lo deseable es que los ROA tengan ambas capacidades, tanto ofrecer una interfaz web, para que los usuarios puedan acceder a la colección, así como la capacidad de comunicarse directamente con las plataformas de aprendizaje y hacer posible la interoperabilidad entre sistemas de diferente naturaleza.

ADL (Advanced Distributed Learning) (LOOMS Y CHRISTENSEN, 2002) propone un conjunto básico de funciones que los repositorios deben proveer a fin de dar acceso a los objetos de aprendizaje en un ambiente seguro. Estas funciones son:

- Buscar/encontrar: Es la habilidad para localizar un objeto de aprendizaje apropiado. Esto incluye la habilidad para su despliegue.
- Pedir: Un objeto de aprendizaje que ha sido localizado.
- Recuperar: Recibir un objeto de aprendizaje que ha sido pedido. Enviar: Entregar a un repositorio un objeto de aprendizaje para ser almacenado.
- Almacenar: Poner dentro de un registro de datos un objeto, con un identificador único que le permita ser localizado.
- Colectar: Obtener metadatos de los objetos de otros repositorios por búsquedas federadas.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

- Publicar: Proveer metadatos a otros repositorios.

Los ROA deben cubrir los siguientes elementos: colección, servicios de valor añadido, personalización y ciclo de vida (los mismos que una biblioteca digital, puesto que se basan en estas) y deben tener las siguientes características deseables (LÓPEZ Y GARCÍA, 2006):

- Herramientas de búsqueda: Considera la búsqueda a través de palabras clave u otros metadatos, la posibilidad de que el usuario pueda realizar exploraciones en listados predefinidos en alguna categorización o clasificación, así como la capacidad del sistema para notificar a los usuarios sobre eventos determinados en el repositorio y la sindicación de los OA.
- Herramientas de recopilación: Creación de bookmarks de recursos o colecciones personales y posibilidad de creación de paquetes con varios recursos.
- Colectividad y evaluación: Posibilidad de que los usuarios puedan evaluar formal o informalmente un OA, mecanismos para registrar los diferentes contextos en los que el OA ha sido utilizado, y listas de OA que el usuario desearía se incluyeran o se modificaran.
- Metaetiquetado: Herramienta de etiquetado, soporte de estándares y/o varios esquemas, importación y exportación de metadatos, mecanismo de identificación única de los recursos especialmente importante en colecciones federadas.
- Administración de contenidos: Seguimiento del flujo de creación y publicación de un OA, control de versiones y funciones de almacenamiento, herramientas de tutoría.
- Administración y cumplimiento de derechos digitales de autor: Registro, transmisión, interpretación y hacer cumplir los derechos de autor, así como un sistema de pago cuando sea necesario.
- Presentación y salidas de consorcio: Accesibilidad, salidas en múltiples formatos para diferentes dispositivos, cambios de apariencia de la interfaz, soporte de caracteres de diferentes idiomas, habilidad para servir como puerta de entrada para varias colecciones, transformación de formatos.
- Integridad e interoperabilidad: Federación y búsqueda de otros repositorios, integración con un administrador de cursos, soporte de servicios web y de aplicaciones API (Application Programming Interface) que puedan extraer información de actividades dentro del repositorio.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

- Consideraciones técnicas: Autenticación, autorización y personalización, informe de uso, soporte para diferentes sistemas operativos, especificaciones de la base de datos requerida por el repositorio, escalabilidad, arquitectura del modelo software, soporte, requisitos técnicos y humanos para su puesta en marcha, cliente del navegador.

## 1.3.3. Análisis sobre la existencia de repositorios a nivel mundial

En todas las regiones del mundo existen repositorios implementados en menor o mayor medida, según los recursos y habilidades de cada institución o país. (FLORRES Y SÁNCHEZ, 2007)

### 1.3.3.1. África

En África sólo existen 8 repositorios institucionales. De ellos, 7 pertenecen a Sudáfrica, por lo que se decidió analizar sólo los de este país.

#### Sudáfrica

Sudáfrica es uno de los países subdesarrollados en los cuales es posible observar un creciente interés de investigadores y profesionales de la información hacia el Movimiento de Acceso Abierto y el tema de las políticas para la implementación de sus estrategias.

Tipología de documentos: En los casos estudiados en esta región, los documentos más frecuentes son las publicaciones, conferencias, tesis y libros y, en menor medida, puede encontrarse multimedia, así como documentos inéditos y postprints.

Plataforma de software: Se han utilizado EPrints, Dspace y ETD-db.

### 1.3.3.2. Asia

En Asia existen 50 repositorios, de los cuales 24 pertenecen a Japón y 18 a la India. La India está a la vanguardia del resto de los países subdesarrollados, tanto en términos de crecimiento económico como de productividad científica, y lidera el Movimiento Acceso Abierto y la creación de repositorios institucionales.

Por su parte, Japón se encuentra entre los países que lideran la producción científica mundial, porque junto a Estados Unidos, el Reino Unido y Alemania producen casi el 85 % de las publicaciones más citadas del mundo.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

Tipología de documentos: En la India, los documentos más depositados son tesis y disertaciones, publicaciones, documentos inéditos, conferencias, referencias, preprints, libros y patentes y, en menor medida, multimedia, objetos de aprendizaje y postprints. En cambio en Japón, los documentos más comunes son objetos de aprendizaje, libros, documentos inéditos, tesis, conferencias, publicaciones y, en una escala menos significativa, materiales especiales y postprints.

Plataforma de software: En la India, los únicos utilizados son versiones de EPrints y DSpace. En Japón, el software más utilizado es DSpace y en un caso se emplea Bepress.

### 1.3.3.3. Australia

En esta región existen 61 repositorios, de ellos pertenecen 52 a Australia y 9 a Nueva Zelanda.

#### Australia

Australia probablemente es uno de los principales países en términos de organización nacional y política de archivos Eprints.

Tipología de documentos: En Australia, el recurso más depositado son las tesis; también son frecuentes las publicaciones, referencias, conferencias, documentos inéditos y libros y, en menor cantidad, banco de datos, software y multimedia. En Nueva Zelanda, las fuentes más depositadas son las tesis, publicaciones y conferencias.

Plataforma de software: En Australia, en la mayoría de los repositorios no se especifica el software utilizado, pero predomina EPrints, DSpace y Fedora. En Nueva Zelanda el software más utilizado es EPrints y en algunos casos se usa DSpace y Bepress.

### 1.3.3.4. Europa

Hasta el 24 de mayo de 2007, el OpenDOAR registraba 434 repositorios en Europa: Alemania 112, Reino Unido 99, Holanda 44, Francia 32, Suecia 31, Italia 24, Bélgica 15 y España 14. El resto de los países europeos presenta una cantidad menor de repositorios.

#### Alemania

Es uno de los países de mayor producción científica a nivel mundial y líder en las iniciativas gubernamentales con respecto al Movimiento de Acceso Abierto. Una de las organizaciones que se destacan es Max Planck Gesellschaft (MPG).

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

## **Reino Unido**

Reino Unido es uno de los países que desarrolla iniciativas gubernamentales para apoyar el Movimiento de Acceso Abierto, al considerarse uno de los mayores productores de ciencia con alrededor de 130 000 artículos por año.

## **Holanda**

El programa DARE (Educación para Resistir el Abuso de Drogas), creado por científicos de las principales universidades holandesas con 47 000 documentos digitales de investigación, reúne publicaciones de 16 repositorios institucionales de las universidades del país, y recolecta todo el material digital disponible en los repositorios locales de sus instituciones afiliadas ofreciéndolos en forma integrada y completa en una única base de datos. Los contenidos no sólo son textuales, sino también sonoros, visuales y audiovisuales.

## **Francia**

Este es uno de los países europeos que muestra mayor interés en proporcionar acceso en línea a la información científica sobre la base de dos estrategias: la digitalización de sus revistas y la implementación de repositorios de acceso abierto.

## **Suecia**

Según un estudio realizado, este es uno de los países con mayor cantidad de artículos publicados bajo la modalidad de acceso abierto, aunque su producción científica global es pequeña.

Cabe destacar que el Directorio de Revistas De Acceso Abierto (DOAJ), que mantiene la Universidad de Lund, en Suecia, lista alrededor de 2 200 revistas, 633 de ellas recuperables hasta el nivel de artículos. Todas las revistas registradas en DOAJ presentan acceso abierto.

## **Bélgica**

Según un estudio desarrollado, Bélgica se encontraba en el primer lugar de países que publicaban artículos con acceso abierto; sin embargo, no es uno de los países que posee mayor cantidad de repositorios.

Tipología de documentos: En Europa, los recursos más comunes son tesis, publicaciones, documentos inéditos, referencias, objetos de aprendizaje y libros y los menos comunes multimedia y postprints.

## Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

Plataforma de software: En Francia y Bélgica, se utilizan plataformas de software como DOKS, DSpace, ETD-db, EPrints y HAL. En Alemania, además, se utilizan UBKA, Milless, EDoc, MONARCH, MyCoRe y OPUS. En España, el más utilizado es DSpace y existe un sitio soportado en OAICat. En Italia, los más comunes son DSpace y EPrints y en algunos casos CDSWare. En Reino Unido y Suecia se utilizan más Dspace y EPrints.

### 1.3.3.5. América del Sur

En América del Sur, existen 38 repositorios, la mayor cantidad de ellos pertenecen a Brasil, con 25; le continúa Chile, con 5.

Tipología de documentos: En estos países, el recurso más utilizado son las tesis y disertaciones, aunque también aparecen publicaciones, multimedia y objetos de aprendizaje.

Plataforma de software: En Chile se utiliza DSpace y en Brasil, DSpace, EPrints y Maxwell.

### 1.3.3.6. América del Norte

En América del Norte, de 284 repositorios existentes, 254 pertenecen a los Estados Unidos y el resto a Canadá.

#### Estados Unidos

Los repositorios institucionales en este país son clara y ampliamente reconocidos como infraestructuras esenciales para la comunicación científica en el mundo digital; además, constituyen infraestructuras de propósito general en el contexto de una práctica académica cambiante; las bibliotecas lideran esta actividad a nivel de formulación de políticas y de implementación operacional.

#### Canadá

Estudios realizados demuestran que es uno de los países con más de un 30 % de sus artículos con acceso abierto.

Tipología de documentos: En Estados Unidos y Canadá, existe un amplio depósito de recursos como tesis, publicaciones, documentos inéditos, conferencias, preprints, postprints, objetos de aprendizaje, referencias, entre otros.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

Plataforma de software: En estos países, predomina la utilización de DSpace y EPrints. (FLORRES Y SÁNCHEZ, 2007)

## 1.3.3.6. Cuba

En Cuba este es un campo poco explorado, se reconoce un sólo caso de repositorio institucional, administrado por INFOMED, aunque existen otros proyectos que se acercan en algunas características de los mismos, establecidos en varias de las instituciones con más desarrollo en el contexto informacional. Algunas instituciones poseen sitios Web institucionales cuyos propósitos se acercan a los de un repositorio institucional.

El Instituto de Información Científica y Tecnológica (IDICT) propuso, en el marco de la Semana Mundial del Acceso Abierto, celebrada del 19 al 23 de octubre de 2009, las bases para el desarrollo de una red de repositorios digitales de acceso abierto en Cuba con el objetivo de incrementar el acceso, la calidad y la visibilidad de los resultados generados por las entidades científicas y académicas del país. Es una red abierta y descentralizada de repositorios digitales basada en el uso de estándares y pautas que garanticen su interoperabilidad y la integración con otros proyectos similares a nivel regional e internacional.

## 1.4. Gestión de documentos y colecciones

De la creación de contenidos accesibles en la web a partir de la información conservada en archivos, se derivan las limitaciones que desde el punto de vista funcional presentan los buscadores para encontrar recursos digitales en general y documentales en particular.

Hasta hace algún tiempo, la utilidad de los buscadores se circunscribía a páginas estáticas HTML (HyperText Markup Language) aunque últimamente se ha incorporado la búsqueda en el contenido de documentos en PDF, Word y otro tipo de formatos. *“Desde luego, las bases de datos permanecen inaccesibles para ellos [los buscadores], lo cual representa un enorme reto a la hora de aumentar la visibilidad de los recursos digitales que produzcan los archivos.”* (ACUÑA Y AGENJO, 2005)

A partir de estos inconvenientes surge la necesidad de establecer entre los sistemas de información, principalmente en los Sistemas de Información Documental (SID) la interoperabilidad, con el fin de posibilitar la transferencia e intercambio de información entre sistemas.

La interoperabilidad es *“la capacidad de un sistema de información de comunicarse y compartir datos, información, documentos y objetos digitales de forma efectiva (con una mínima o nula pérdida de su valor y funcionalidad), con uno o varios sistemas de información (siendo generalmente estos sistemas*



# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

*completamente heterogéneos, distribuidos y geográficamente distantes), mediante una interconexión libre, automática y transparente, sin dejar de utilizar en ningún momento la interfaz del sistema propio".* (GÓMEZ, 2007)

En los SID, esta capacidad de interoperar entre ellos está dada por el uso de protocolos de intercambio, transferencia y recuperación de información bibliográfica.

Son numerosos los protocolos desarrollados a través de los años con el fin de permitir la interoperabilidad entre Sistemas de Información Documental (SID):

- Z39.50.
- Simple Digital Library Interoperability Protocol (SDLIP).
- OAI (Open Archives Protocol).
- Guildford protocol.
- Dienst protocol.
- ZING SRU/W.
- Starts (Stanford Protocol Proposal for Internet Retrieval and Search).
- Dienst protocol.
- Lyceum Protocol.
- Harvest: A Distributed Search System.

Sin embargo, sólo algunos de éstos se han consolidado internacionalmente como estándares aceptados por los profesionales de la información, son usados de manera masiva y su desarrollo y mantención se encuentra acreditada por iniciativas tanto públicas como privadas, entre los que se encuentra el Protocolo OAI-PMH.

## **1.4.1. Protocolo OAI-PMH (Open Archives Initiative - Protocol Metadata Harvesting).**

El protocolo OAI-PMH (Open Archives Initiative - Protocol Metadata Harvesting) es un protocolo que ordena el intercambio de información entre diferentes fuentes bajo el principio del libre acceso, pero

## Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

respetando los derechos de autor. El protocolo genera estándares de interoperabilidad que facilitan la difusión y el intercambio de la información (PÉREZ Y SILVA, 2007). Entre sus ventajas se tiene la posibilidad de contar con una única interfaz de búsqueda que provea registros de disímiles fuentes, además de la propia ventaja del acceso libre. “*Se ha puesto de manifiesto que el libre acceso a través de internet a los artículos científicos aumenta el número de citas que reciben frente a los que su acceso es restringido*” (MELERO, 2005).

OAI-PMH utiliza transacciones HTTP (HyperText Transfer Protocol) para emitir preguntas y obtener respuestas entre un servidor o archivo y un cliente o servicio recolector de metadatos. El segundo puede pedir al primero que le envíe metadatos según determinados criterios como por ejemplo la fecha de creación de los datos. En respuesta el primero devuelve un conjunto de registros en formato XML (Extensible Markup Language), incluyendo identificadores de los objetos descritos en cada registro, por ejemplo URLs (Uniform Resource Locator).

Esta estructura de interoperabilidad corresponde al enfoque de harvesting (recopilación), en la que los datos y metadatos son transferidos desde la fuente remota al destino en el cuál se realizarán los servicios de búsqueda.

El protocolo soporta múltiples formatos para expresar los metadatos, no obstante requiere que todos los servidores ofrezcan los registros utilizando Dublin Core no calificado, codificado en XML. Además de éste formato cada servidor es libre de ofrecer los registros en otros formatos adicionales. Un cliente puede pedir que los registros se le sirvan en cualquiera de los formatos soportados por el servidor.

La arquitectura de OAI-PMH se basa en clientes y servidores; los primeros son los archivos que proporcionan la información y los segundos son los recolectores o servicios que toman los datos, y los presentan a los usuarios finales.

OAI-PMH es en esencia una plataforma simple para acceder a la información bibliográfica disponible en un archivo o repositorio. De esta forma, cualquier institución puede realizar una implementación del protocolo para poner a disposición su propia base de datos o catálogo, y a la vez consultar otras.

Si bien OAI-PMH surgió en el seno de la comunidad académica y científica para la búsqueda y recuperación de textos electrónicos, es perfectamente aplicable en cualquier contexto documental.

Además, las características de flexibilidad del OAI-PMH para adaptarse a cualquier ámbito, pudiendo ofrecer información sobre todo tipo de recurso, tanto físico como digital, permiten la diseminación de

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

información a través de la generación de metabuscadores mediante su uso y por tanto la formación de repositorios de información y de bibliotecas digitales.

## **1.5. Herramientas para el desarrollo de catálogos y bibliotecas digitales.**

Entre los principales sistemas de software relacionados con la creación y desarrollo de catálogos y bibliotecas digitales se encuentran:

### **1.5.1. Dspace**

Es un sistema de información con arquitectura de repositorio digital que captura, almacena y ordena material de investigación digital con el propósito de garantizar que se preserve y distribuya toda la producción intelectual generado al interior de las instituciones que hacen uso de este. Es un proyecto conjunto de las bibliotecas del MIT (Massachussets Intitute of Technology) y Hewlet-Packard.

Provee soporte para captura y almacenamiento de datos en formatos digital, indexado para su posterior recuperación, distribución de contenido a través de la web y preservación a largo plazo. El sistema se encuentra disponible bajo licencia BSD, lo que significa que puede ser usado, modificado e incluso integrado a una aplicación comercial sin cargos en concepto de licencia. Este software fue desarrollado utilizando las normas y estándares existentes lo que le permite integrarse fácilmente a otros sistemas de información.

### **1.5.2. Eprints**

Es un software desarrollado en el seno del Open Citation Project dirigido por Stevan Harnad en la Universidad de Southampton (UK). Es un programa fácil operar y es gratuito. Puede funcionar en cualquier computadora que emplee sistema operativo GNU/Linux, y sus principales características son las siguientes:

- ✓ Permite almacenar documentos en cualquier formato, así como almacenar un mismo documento en varios formatos.
- ✓ Posibilita utilizar cualquier formato para almacenar la información bibliográfica sobre los documentos.
- ✓ Permite que los usuarios se registren como lectores o como autores para obtener un mayor aprovechamiento de sus funciones.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

---

## 1.5.3. Greenstone

Es un software para la creación de bibliotecas digitales. Permite organizar y publicar documentos electrónicos en cualquier formato. Greenstone es producido por el proyecto de Biblioteca Digital de la Universidad de Waikato, Nueva Zelanda, y desarrollado y distribuido en cooperación con la UNESCO (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura) y la ONG Human Info (Organización Non gubernamental belga). Es de código abierto, emitido conforme a los términos de la GNU General Public License. La interfaz completa de Greenstone y toda la documentación está disponible en inglés, francés, español, ruso y kazajo. Greenstone también tiene interfaces en muchos otros idiomas.

## 1.5.4. Drupal

Es un sistema de administración y gestión webs. Permite publicar contenidos textuales, imágenes y cualquier otro tipo de material. Incluye las funciones básicas de la comunidad como blogs, foros y formularios de contacto, y se puede extender fácilmente mediante la descarga de otros módulos de terceros y los temas. Es un programa de código abierto, con licencia GNU/GPL, escrito en PHP, desarrollado y mantenido por una activa comunidad de usuarios. Esta herramienta está siendo implementada en bibliotecas ya que incluye módulos específicos para bibliotecas (OPAC 2.0).

## 1.6 Ejemplos de repositorios y bibliotecas digitales desarrolladas con estas herramientas

### 1.6.1 Repositorio Institucional de la Universidad Andina Simón Bolívar

Es un repositorio de acceso abierto a la producción intelectual de la comunidad universitaria de la Universidad Andina Simón Bolívar (UASB) en su labor de docencia e investigación, conformado por libros, tesis, revistas, proyectos de investigación, artículos y otros documentos a texto completo. Esta desarrollado sobre el software DSpace.

El Repositorio Institucional de la UASB almacena, gestiona, difunde y ofrece acceso a los documentos digitales concernientes a los campos de interés académico de la universidad. Además, permite a los miembros de la comunidad académica auto-archivar sus documentos. (Ver Figura 1)

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales



Figura 1. Repositorio Institucional de la Universidad Andina Simón Bolívar (UASB)

## 1.6.2 Biblioteca Nacional de Cuba José Martí

La Biblioteca Nacional de Cuba José Martí es depositaria del tesoro patrimonial documental, bibliográfico, artístico y sonoro del país, así como de lo más representativo de la cultura universal. Para el desarrollo de esta biblioteca se utilizó el software DSpace.

A través de este sitio, la Biblioteca Nacional de Cuba José Martí, brinda la posibilidad de interactuar con la cultura, la ciencia, la historia y el arte. Al interactuar con este sitio, investigadores, estudiantes, profesores y lectores en general, se relacionen con la creación cultural, científica y artística cubana a partir de los documentos expuestos allí. (Ver Figura 2)

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

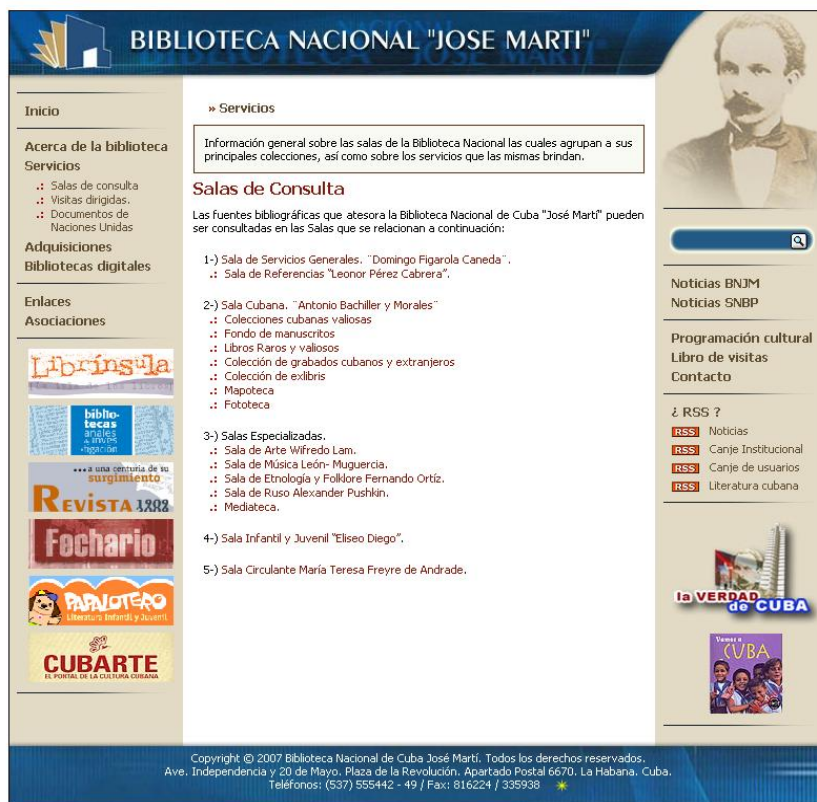


Figura 2. Biblioteca Nacional de Cuba José Martí (BNJM)

## 1.6.3 Red de Bibliotecas Virtuales de Ciencias Sociales de CLACSO

Con el objetivo de promover y facilitar el acceso a los resultados de las investigaciones de los Centros Miembros de CLACSO, se desarrolló esta biblioteca virtual con el software libre Greenstone, que permite acceder gratis desde cualquier lugar del mundo a los contenidos de la misma. La Red de Bibliotecas Virtuales de Ciencias Sociales de CLACSO está constituida por las bibliotecas de los Centros Miembros del Consejo Latinoamericano de Ciencias Sociales.

Los servicios que esta ofrece incluye el acceso libre y gratuito a la Sala de Lectura con textos completos de libros, artículos, ponencias y documentos de trabajo publicados por la red CLACSO. Además, ofrece una página de enlaces a otras bibliotecas, libros en línea, revistas de América Latina y el Caribe que puedan ser de interés para los investigadores de la red CLACSO. Por otro lado, ha desarrollado un portal multimedia, que incluye enlaces a producciones audiovisuales, radios en línea, archivos de audio y colecciones fotográficas realizadas por centros miembros de CLACSO. (Ver Figura 3)

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

The screenshot displays the homepage of the Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe de la Red CLACSO. The page features a search bar at the top with a dropdown menu set to 'Texto Completo'. Below the search bar, there are several navigation and information sections:

- Red de Bibliotecas Virtuales:** Includes links for 'Quiénes somos', 'Ingreso de información', 'Recursos para bibliotecas virtuales', and 'Novedades'.
- Portal de revistas Red CLACSO:** A prominent orange button labeled 'INGRESO'.
- Noticias:** A section for news from the network's members.
- Estadísticas:** A section with a 'Creative Commons' logo and links to 'Resumen Mensual', 'Temas más utilizados', 'Temas más buscados', and 'Estadísticas por texto'.
- ¿Tiene dificultades en esta Biblioteca Virtual?:** A section with a link 'Estamos para ayudarle'.
- Centros Miembros:** A section with a map of Latin America and the Caribbean.
- ¿Tiene sugerencia para mejorar nuestro servicio?:** A section for user feedback.
- Cooperación Interinstitucional:** A section listing partners like Scholar Google, University of Texas Latin America Network, etc.
- Ingresar:** A section for registered users.

The main content area features a large banner for 'Internacional Semana de Acceso Abierto' (Open Access Week) from October 24 to 30, 2011, with the CLACSO logo and the website 'openaccessweek.org'. Below the banner, there are three columns for 'Sala de Lectura', 'Portal de Revistas CLACSO', and 'Portal CLACSO REDALyC'. A central message encourages users to click on 'novedades' for updates from May 2005 onwards. A footer section contains contact information for the Consejo Latinoamericano de Ciencias Sociales and a list of participating member centers.

Figura 3. Red de Bibliotecas Virtuales de Ciencias Sociales de CLACSO (CLACSO)

## 1.6.4 ECURED

La ECURED es una biblioteca cubana que busca su proyección y participación universal. Ofrece un espacio interactivo para la creación y socialización de contenidos académicos. Tiene como objetivo facilitar el acceso a una selección de documentos académicos digitales organizados en tres repositorios de acceso abierto: artículos científicos, libros académicos y bibliografías especializadas. Esta biblioteca funciona con la herramienta Greenstone.

Los artículos científicos han sido seleccionados, en su mayoría, de revistas cubanas. La colección de libros académicos es una selección de las publicaciones de la Editorial Universitaria Cubana. La colección de bibliografías especializadas además de ser un repositorio, es un servicio de orientación, referencias y asistencia bibliográfica. (Ver Figura 4)

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

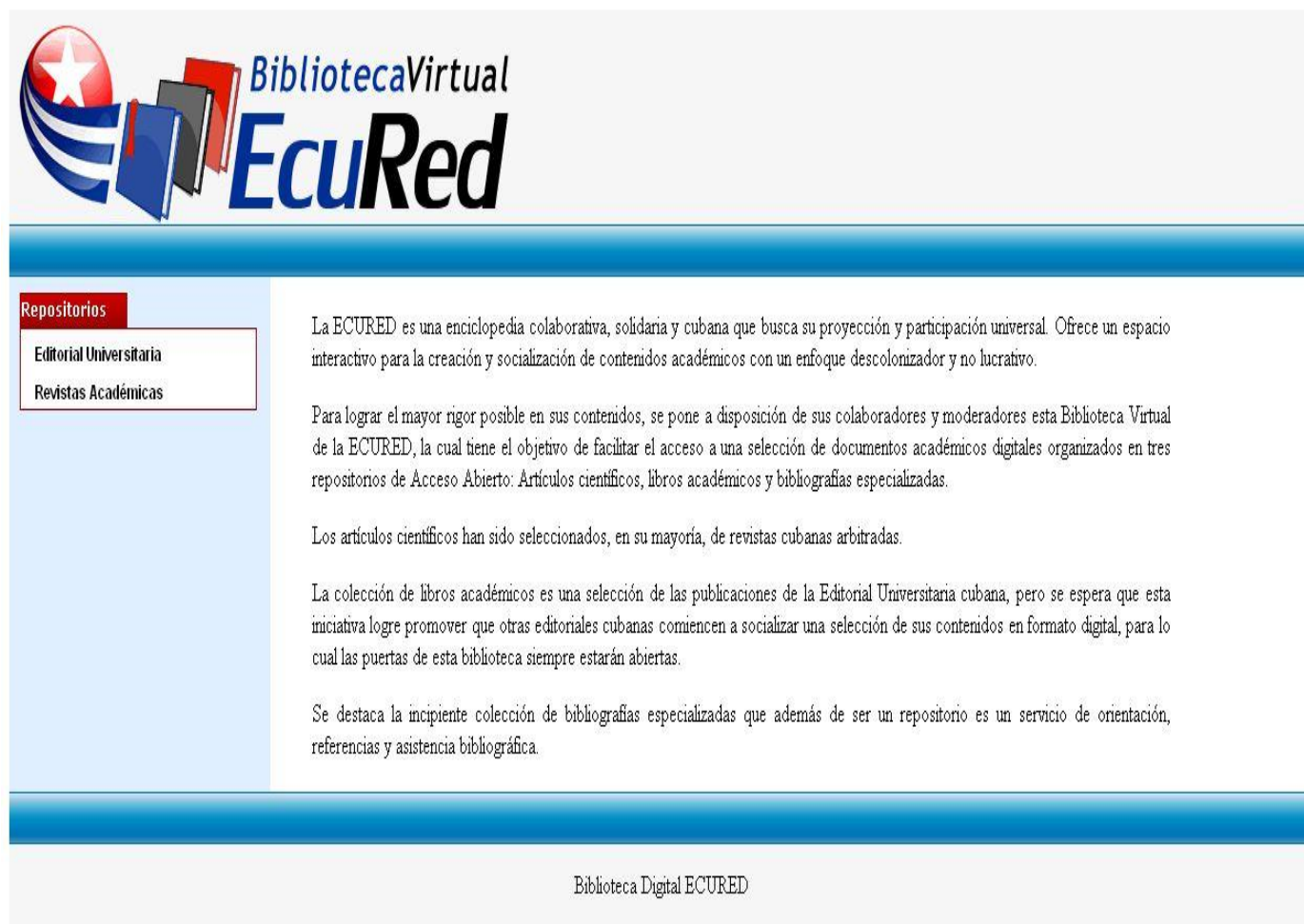


Figura 4. Biblioteca Virtual EcuRed (EcuRed)

## 1.6.5 E-Prints Complutense

E-Prints Complutense es el archivo institucional en acceso abierto desarrollado por la Biblioteca de la Universidad Complutense de Madrid (UCM), para gestionar la documentación digital, fruto de la actividad de sus docentes, investigadores y los grupos de investigación validados por la UCM. Con este archivo institucional, la biblioteca se propone difundir y preservar la investigación, así como mejorar su organización, acceso y distribución. Como su nombre lo dice es un repositorio que funciona con la plataforma EPrints.

Los contenidos abarcan todas las materias científicas que se imparten en la universidad como: tesis doctorales, documentos de trabajo, artículos de revistas, actas de congresos, libros, capítulos de libros y otros. (Ver Figura 5)



# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

The screenshot shows the homepage of the E-Prints Complutense repository. At the top left is the logo of the Biblioteca Universidad Complutense. The main title 'E-Prints Complutense' is centered in a large blue font. Below the title is a navigation menu with links: Inicio, English, Información, Navegación, Estadísticas, Contacto, and Incidencias. Underneath the menu are links for 'Depositar documentos', 'Cómo depositar', and 'Registrarse'. A message states: 'En este archivo los docentes e investigadores de la UCM depositan sus documentos en acceso abierto (open access)'. There are RSS feeds for Atom, RSS 1.0, and RSS 2.0. A yellow warning icon is followed by three bullet points: 'E-Prints Complutense implementa las directrices OpenAire', 'Si desea publicar su tesis doctoral en E-Prints Complutense, póngase en contacto con el Servicio de Publicaciones.', and 'La UCM se ha sumado al mandato de la Comunidad de Madrid para que los grupos de investigación publiquen en abierto sus resultados en este Repositorio.'. The main content area is divided into three columns. The left column has a search box with the text 'En los campos: título, autor, resumen y fecha.' and a 'Buscar' button. Below it is a checkbox for 'Buscar también en texto completo:' and a link for 'Búsqueda avanzada'. The middle column is titled 'Navegación' and lists: 'Por materias.', 'Por año.', 'Por autor.', 'Por tipo de documento.', 'Por Facultad, Departamento e Institutos.', 'Documentos de trabajo de las Facultades.', and 'Últimos documentos depositados.'. The right column is titled 'Área de usuarios' and lists: 'Depositar documentos.', 'Cambiar contraseña.', and 'Formulario de registro.'. Below that is a section 'Acerca de E-Prints Complutense' with links: '¿Qué es E-Prints?', '¿Tiene dudas sobre los derechos de autor de sus artículos?', and 'Software Eprints.'. At the bottom, there is a section 'Otros repositorios Open Access' with logos for 'Revistas Científicas Complutenses.', 'ciencia.mad E-Ciencia.', 'RECOLECTA Recolecta.', 'Oaister', and 'Semana de acceso abierto.'. At the very bottom, there are two small logos: 'SHERPA/RoMEO' with a link to 'Políticas de las editoriales de las publicaciones científicas respecto al copyright y auto-archivo' and 'eprints' with a link to 'El software de creación de archivos GNU EPRINTS está disponible de modo gratuito.'

Figura 5. E-Prints Complutense (E-Prints C)

## 1.6.6 Repositorio de Tesis de Doctorado en Ciencias Biomédicas y de la Salud de Cuba

El Repositorio de Tesis de Doctorado en Ciencias Biomédicas y de la Salud de Cuba forma parte de la estrategia regional de la Biblioteca Virtual en Salud para la divulgación de la producción científica cubana. Tiene como misión difundir y preservar digitalizado el patrimonio científico e intelectual de los profesionales del Sistema Nacional de Salud en Tesis de doctorado, e incrementar su visibilidad en un espacio abierto. Está coordinado por el Centro Nacional de Información de Ciencias Médicas, los Centros Provinciales de Información de Ciencias Médicas y la red de bibliotecas del Sistema Nacional de Información de Ciencias Médicas. Funciona con la plataforma EPrints.

Las tesis de doctorado se incluyen de acuerdo a tres criterios definidos en el repositorio.

1. Últimas tesis aprobadas por dictamen con resolución de la Comisión Nacional de Grados Científicos.

# Capítulo I: Tendencias actuales en entornos tecnológicos para el desarrollo de bibliotecas digitales

2. Tesis anteriores con versión digital.

3. Tesis pendientes de digitalizar. (Ver Figura 6)

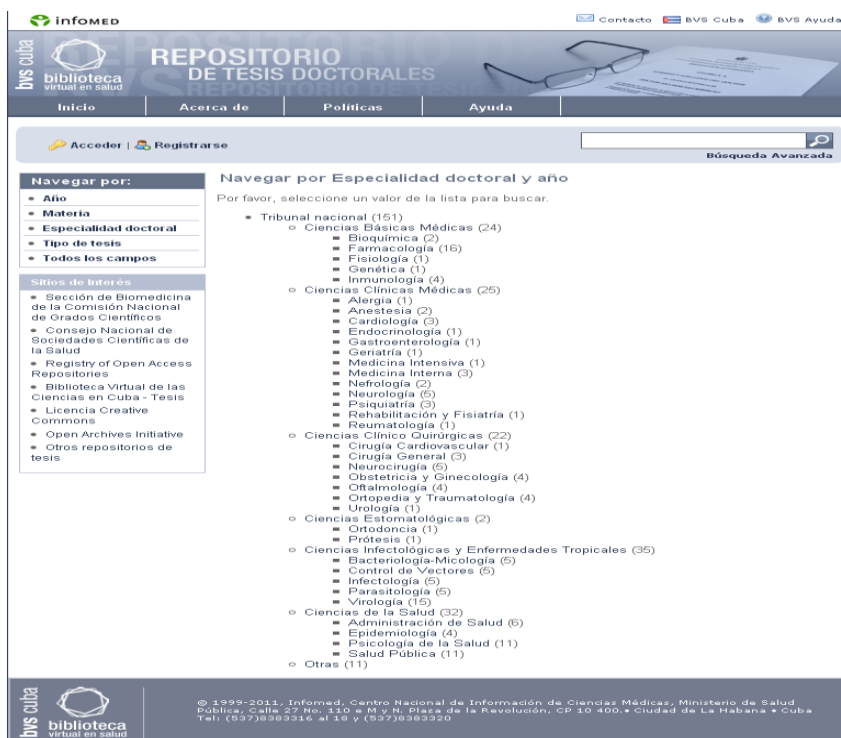


Figura 6. Repositorio de Tesis Doctorales en Ciencias Biomédicas y de la Salud de Cuba (RTD)

## 1.7 Conclusiones del capítulo

En este capítulo se analizó el estado del arte de los entornos tecnológicos para la creación de bibliotecas y repositorios digitales. También se tomó en cuenta las aplicaciones existentes para el desarrollo de los mismos, así como la presentación de varios ejemplos de sistemas que utilizan estas herramientas para un mejor entendimiento de las mismas. Quedando definido como principales herramientas a investigar DSpace, Eprints, Greenstone y Drupal.

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

### **Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales**

#### **2.1 Introducción**

El desarrollo de una biblioteca digital implica responder una serie de cuestionamientos que contextualizan el proyecto, interrogantes como: ¿cuál será la visión de la biblioteca dentro de la institución?, ¿quién interactuará con la biblioteca?, ¿qué soporte informático se necesita?, ¿qué estándares y especificaciones técnicas se usarán? y ¿cuál será el costo del proyecto?

Uno de los principales obstáculos para la creación y proyección de bibliotecas digitales es el alto costo de la tecnología asociada o más específicamente el gran costo que tienen los softwares que administran y gestionan los procesos claves asociados a la creación de las mismas.

Gracias a las tecnologías de código abierto existen ciertos softwares de clase mundial que hacen posible la construcción de bibliotecas digitales a un costo notoriamente bajo. En este contexto nacen softwares de código abierto tales como E-Prints, Dspace, Greenstone y Drupal.

#### **2.2 Características de las herramientas**

##### **2.2.1 DSpace**

DSpace es un repositorio de documentos digitales con preservación a largo plazo. Es un archivo estandarizado de documentos con un sistema de indexación y búsqueda sobre metadatos y en el texto completo. Crea un identificador permanente para cada artículo, colección y comunidad en Dspace, para lo cual utiliza el Sistema de Handles de la Corporación para Iniciativas de Investigación Nacionales (CNRI, por sus siglas en inglés). Permite la realización de copias de seguridad automáticamente de los archivos de una institución a otra.

También permite a los investigadores encontrar lo que están buscando, y suscribirse a colecciones de documentos de su interés. Está programado en Java y corre sobre sistemas tipo Unix y Windows. Precisa como servidor web y motor de servlets Apache Tomcat. Usa una base de datos relacional de código abierto: PostgreSQL. Tiene dos interfaces, una clásica (JSPUI) que está implementada usando Java Servlets para manejo de lógica de negocio y JavaServer Pages para generación del código HTML enviado a los navegadores; y la otra es XMLUI, la cual está implementada basada en el framework Apache Cocoon y usa XML y XSLT (TERÁN Y GONZÁLEZ, 2009). La interfaz del usuario final apoya la búsqueda

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

y visualización de los documentos, esta es totalmente configurable según las necesidades de cada institución o comunidad de usuarios, mediante HTML y Java.

Es totalmente compatible con el protocolo OAI-PMH, y es capaz de exportar paquetes de software METS (Metadata Encoding and Transmission Standard). Los archivos de datos se organizan juntos en sistemas relacionados para su descripción. Los metadatos, información técnica sobre los datos, se almacenan junto a los documentos para apoyar la preservación. Trabaja con objetos digitales que encapsulan atómicamente los documentos del usuario, los metadatos agrupados, y los identificadores digitales. Estos son indexados para permitir al sistema visualizar colecciones similares y buscar documentos independientes. (evalDSpace\_v01)

### **2.2.1.1 Forma de organización de los documentos**

DSpace impone una estructura de organización de la información que pretende ser un reflejo de la estructura de la institución que implementa el sistema. Cada instancia está compuesta por comunidades, las cuales representan el nivel más alto de la jerarquía de contenido. Una comunidad es una agrupación lógica de una o más colecciones y puede a su vez ser dividida en subcomunidades.

Las comunidades contienen colecciones, que son agrupaciones de contenido relacionado lógicamente. Cada colección está compuesta por ítems, los cuales representan el elemento básico de almacenamiento del repositorio. Cada ítem pertenece a una única colección, aunque puede aparecer en colecciones adicionales.

Se pueden agregar varias colecciones a una subcomunidad y varias subcomunidades a una comunidad pero no inversa, también se pueden hacer modificaciones a las ya existentes con una cuenta autorizada.

### **2.2.1.2 Gestión de los usuarios**

DSpace distingue tres tipos de usuarios, por orden de privilegios: anónimo, normal y administrador. Por defecto el sistema asigna un conjunto de privilegios a cada uno.

- Anónimo: Usuario no autenticado en el sistema.
- Normal: Una vez identificado en el sistema, los usuarios normales pueden recibir privilegios según lo disponga el administrador.
- Administrador: Acceso absoluto a todas las funciones de los ítems de Dspace.

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

Organiza a los usuarios en cuentas personales y en grupos, obligando a todos los usuarios a pertenecer a un grupo definido para poder asignarle permisos de escritura, lectura, modificación y eliminación de registros. Estas autorizaciones pueden realizarse a nivel de comunidad o colección.

Un mismo usuario puede pertenecer a más de un grupo, obteniendo los privilegios de ambos e incluso un grupo puede ser miembro de otro grupo.

### **2.2.1.3 Autenticación**

El sistema requiere una dirección de correo electrónico y no permite a múltiples usuarios usar la misma dirección de correo electrónico. Además provee una característica de recordatorio de contraseña.

### **2.2.1.4 Proceso de carga**

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

Luego se procede a definir la descripción y los metadatos del ítem. Para lo cual se debe ingresar tanta información como sea posible y deberán darse valores al menos para todos los campos definidos como requeridos para avanzar a la siguiente etapa.

Posteriormente se procede a subir el(los) documento(s). En el caso de que el ítem contenga más de un archivo, podrá ingresarse una descripción para cada archivo incluido. Una vez seleccionado y subido el(los) archivo(s) el sistema permitirá definir el formato del mismo (en caso de que la detección de formato automática no haya sido posible o el usuario la considere incorrecta) y ofrecerá además la posibilidad de verificar que el archivo fue subido sin error u omisión mediante checksum.

Finalmente se presenta la licencia de distribución y se ofrece a quien realiza la carga la posibilidad de aceptarla o rechazarla. La licencia es personalizable. En caso de aceptación se completa la carga y en caso de rechazo se cancela el proceso salvando el ítem en el espacio de trabajo del usuario.

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

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

### 2.2.1.5 Formatos de archivos soportados

DSpace no impone restricciones sobre formatos de archivos a subir y permite al administrador mantener una lista de formatos soportados por el repositorio. Tampoco impone límite alguno sobre el tamaño de los archivos que se pueden subir, aunque da al administrador la posibilidad de hacerlo mediante edición del archivo de configuración principal.

Aunque soporta todo tipo de formatos de documentos, se recomienda a los administradores de cada sistema decidir qué formatos aceptar con fines de preservación, pudiendo diferenciar así tres categorías de formatos: soportados, conocidos y no soportados. Además los documentos se mantienen accesibles con formatos actuales, mientras se desarrollan y actualizan formatos nuevos.

### 2.2.1.6 Búsqueda y navegación

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

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

Métodos de búsqueda:

- Búsqueda avanzada: Permite ingresar una expresión de búsqueda formada por pares (metadato, valor) conectados mediante operadores booleanos (and/ or / not). Así como buscar en todo el repositorio o limitar la búsqueda a una comunidad. Además se puede crear una expresión combinando a lo sumo tres términos de búsqueda.
- Búsqueda simple: Busca coincidencias con el(los) término(s) ingresado(s) en los valores de varios metadatos de cada ítem. Si la búsqueda de texto completo está habilitada, el texto completo de cada ítem también será considerado en la búsqueda.

### 2.2.1.7 Estándares de metadatos

DSpace provee por defecto un esquema de metadatos Dublin Core, aunque permite configurar múltiples esquemas y seleccionar campos de metadatos entre una combinación de esquemas configurados para describir un ítem.

La información sobre los esquemas es mantenida en lo que la aplicación llama registro de metadatos. El registro de metadatos puede ser administrado directamente desde la interfaz web permitiendo editar

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

esquemas, agregando, quitando o modificando elementos. Cualquier colección de DSpace puede ser importada en Greenstone.

### 2.2.1.8 Comunidad de usuarios

Para llevar a cabo la investigación de esta herramienta se pudo contar con la gran comunidad de usuarios que apoyan a la misma, con el fin de colaborar con el desarrollo y la popularización del producto. Dicha comunidad cuenta con tres listas de correo:

DSpace general: Donde se realizan preguntas generales, anuncios y discusiones sobre cuestiones no técnicas (Dspace-general).

DSpace tech: Donde se realizan preguntas y discusiones técnicas (DSpace-tech)

DSpace devel: Consiste en una lista para desarrolladores trabajando sobre el código de la aplicación (Dspace-devel).

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

### 2.2.1.9 Las ventajas técnicas DSpace son que:

- El código es limpio (o puro) y utiliza los estándares.
- La arquitectura se basa en estándares como Dublin Core, METS, OAI, CNI Handles para la persistencia de los URL y otros.
- El código se encuentra bien comentariado. Se puede ver lo que el programador ha hecho y se puede modificar o corregir fácilmente.
- Posee buena arquitectura, estrato de negocios (business layer) separado del estrato de la aplicación.
- La base de datos permite las búsquedas normales de SQL y hace la corrección de problemas más fácil.
- Posee identificadores persistentes.
- Es orientado a la preservación.

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

- Soporta gran cantidad de formatos.
- Permite realizar suscripciones a colecciones.
- Existe suficiente documentación técnica actualizada.

Algunos de los beneficios que se ven en este sistema con su implementación son:

- **Visibilidad:** Aumenta la difusión nacional e internacional de los documentos, ya que permite que sean fácilmente recuperados y citados por todo el mundo. Gracias al manejo de estándares para intercambiar información (Metadatos Dublin Core y el protocolo OAI-PMH), permite compartir la misma.
- **Permanencia:** Asegura la estabilidad de la localización de los trabajos en línea, mediante la asignación de una dirección única en Internet (identificador Handle) para cada documento, lo que permite un modo de cita sencillo, directo y confiable.
- **Flexibilidad:** Soporta el ingreso de múltiples formatos de archivo y tipos documentales, incluidos documentos de texto, hojas de cálculo, presentaciones, imágenes, programas y videos.
- **Almacenamiento seguro:** Garantiza la conservación, respaldo y perdurabilidad de todos los documentos ingresados, mediante mecanismos estándares de preservación digital automatizada.
- **Control de acceso:** Permite a los administradores de colecciones y a los mismos autores a determinar el nivel de acceso a sus documentos, dejándolos como públicos o privados, y con la posibilidad de variar esa condición en el tiempo.

### **2.2.1.10 Desventajas DSpace:**

- Para cada OA los metadatos tienen que ser ingresados en un formulario.
- La administración del repositorio requiere tiempo y dedicación.

### **2.2.2 E-Prints**

EPrints es un software que 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 creado con esta herramienta es un conjunto de servicios ofrecidos a todos los miembros de



## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

una institución para el acceso abierto, la preservación, la enseñanza, la publicidad, la colaboración, compartir datos, e-learning, e-publishing, entre otros. Además 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 para lo cual se utiliza la operación de copia o la de clonación que ofrece el software. Es totalmente compatible con el protocolo OAI-PMH. Para evitar duplicación de objetos el sistema alerta al usuario en caso de que un objeto existente se vea similar al que se está ingresando.

Actualmente en su versión 3.0, el código está escrito en Perl, utiliza el servidor web Apache y MySQL como base de datos, pudiendo funcionar sobre los sistemas operativos GNU/Linux, GNU o Solaris, así como en Windows gracias a un módulo adicional desarrollado en su última versión. (evalEPrints\_v02)

### **2.2.2.1 Forma de organización de los documentos**

EPrints provee posibilidades de organizar los documentos del repositorio en colecciones.

### **2.2.2.2 Gestión de los 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.

- Visitantes: Son los usuarios con menores privilegios. Pueden navegar el repositorio, realizar búsquedas, visualizar y descargar objetos públicos, y suscribirse a las novedades del repositorio.
- Usuarios registrados: Son los usuarios que crean una cuenta de usuario en el repositorio. Este proceso se realiza a partir de completar un formulario con datos personales y la elección del nombre de usuario y la contraseña. Para finalizar el proceso el usuario debe activar la cuenta mediante un e-mail automático enviado por el sistema a su correo.
- 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.

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

### **2.2.2.3 Autenticación**

El sistema requiere una dirección de correo electrónico. Además permite a los usuarios restablecer sus contraseñas en caso de olvidarlas.

### **2.2.2.4 Proceso de carga**

La carga de objetos está disponible para usuarios registrados, editores y administradores. Se realiza mediante una interfaz web. Este software 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 entre los disponibles.

La creación de un nuevo objeto comienza al seleccionar el tipo de objeto y subir el(los) documento(s), así como, elegir el formato del archivo, entrar una descripción, definir visibilidad, seleccionar el tipo de licencia e ingresar una fecha de embargo.

Posteriormente se procede a ingresar los metadatos necesarios para cargarlo. Deberán darse valores al menos para todos los campos definidos como requeridos e ingresar tanta información como sea posible. En esta etapa resulta de especial utilidad la función de autocompletado.

Para finalizar la carga se debe seleccionar que temas trata el objeto y aceptar la carga del objeto sino existen campos requeridos faltantes.

### **2.2.2.5 Formatos de archivos soportados**

En cuanto a formato, este sistema puede almacenar cualquier tipo de ficheros que el administrador decida aceptar, incluyendo HTML, Adobe PDF, y PostScript, y cada documento puede ser almacenado en varios formatos distintos. 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.

### **2.2.2.6 Búsqueda y navegación**

EPrints 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. Pero también permite definir vistas personalizadas por

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

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. Además permite seleccionar el criterio para ordenar los resultados.

### 2.2.2.7 Estándares de metadatos

Aunque por defecto se emplean metadatos Dublin Core, EPrints permite aplicar cualquier esquema de metadatos, y que el administrador decida qué campos se asignan a cada tipología documental.

### 2.2.2.8 Comunidad de usuarios

Durante la investigación se contó con la posibilidad de poder aclarar dudas por diversas vías, entre las que se encuentran: la documentación de ayuda, las consultas a la comunidad de usuarios y a documentos desarrollados por terceros sobre su experiencia en la implementación de repositorios con dicha herramienta. Además, dispone de una lista de correo para discusiones técnicas sobre el uso y desarrollo del producto donde la suscripción es gratuita.

### 2.2.2.9 Ventajas aportadas por Eprints

- Tiene la posibilidad de ingresar nuevos módulos que se desarrollen.
- La interfaz es modificable a las características de la unidad de información.
- El proceso de la carga de la descripción de un libro en el sistema es muy fácil e intuitivo de usar.
- Permite la pre catalogación, de manera que sea solo un editor el que autorice que un elemento suba al sistema.
- Posibilidad de indicar los tipos de documentos que podrá cargar en el sistema.
- Permite modificar la estructura de la clasificación que viene predeterminada (LC) a conveniencia de la unidad de información.
- Tiene la posibilidad de hacer búsquedas a partir de otras ya realizadas.
- Permite limpiar el formulario para hacer nuevas búsquedas.

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

### 2.2.2.10 Desventajas

- No soporta formatos para objetos complejos como METS, MPEG21 o IMS CP (para el empaquetamiento de objetos de aprendizaje).
- Funciona mejor para documentación textual y no multimedia.

### 2.2.3 Greenstone

El sistema Greenstone es una aplicación pensada para la construcción y administración de bibliotecas digitales accesible desde la Web, proporcionando de esta forma el acceso tanto a documentos textuales como a imágenes, vídeos y sonidos. Su organización permite distribuir los documentos en colecciones que se pueden crear atendiendo a sus características. Por otra parte, favorece la importación y/o asignación de metadatos, de manera manual o automática. Permite configurar el tipo de indexación a utilizar, modalidades de búsqueda y obtener listados de resultados en función a la relevancia con los términos consultados.

Las implementaciones mediante este sistema pueden publicarse en Internet como también en entornos más específicos como las intranets. Es totalmente compatible con el protocolo OAI-PMH. Funciona en todas las versiones de Windows, Unix/Linux y también en Mac OS-X. Es muy sencillo de instalar y trabaja con otros programas de software libres, incluyendo el servidor Web Apache y Perl. Para trabajar con esta aplicación existen dos interfaces interactivas separadas: la página del lector y la interfaz del bibliotecario de Greenstone (GLI).

La descripción de los documentos se puede realizar de manera individual, documento a documento, o describir las distintas agrupaciones en conjunto. De esta última manera se asigna a todos los recursos incluidos en cada agrupación la misma descripción de manera automática, desarrollando así un práctico sistema multinivel. (SOLA, 2008)

#### 2.2.3.1 Forma de organización de los documentos

Greenstone permite organizar los documentos de la biblioteca a partir de las colecciones que posee por defecto o a partir de la creación de nuevas colecciones.

Una biblioteca digital creada con esta herramienta contendrá muchas colecciones, organizadas individualmente, pero muy parecidas entre sí. De fácil mantenimiento, las colecciones pueden aumentarse y reorganizarse automáticamente.

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

### 2.2.3.2 Gestión de los usuarios

Cuando se instala Greenstone, hay un solo usuario llamado admin. Su contraseña se establece durante el proceso de instalación. Este usuario desde la página de administración, puede añadir nuevos usuarios, suprimir antiguos y hacer una lista de los usuarios.

A su vez cada usuario puede pertenecer a grupos diferentes. Los grupos se usan para controlar las acciones que los usuarios están habilitados a ejecutar sobre las colecciones. Inicialmente los únicos existentes son administrator (administrador) y colbuilder (creador de colección). Los miembros del primer grupo pueden añadir y suprimir usuarios, así como modificar sus grupos. Los miembros del segundo pueden acceder a las funciones para crear nuevas colecciones y modificar (o suprimir) colecciones existentes.

### 2.2.3.3 Autenticación

El sistema requiere un usuario y contraseña para registrarse, además una dirección de correo electrónico para utilizarla en caso de que su contraseña sea olvidada.

### 2.2.3.4 Formatos de archivos soportados

Greenstone es capaz de procesar un considerable número de documentos de diversa índole, tales como: textos (.txt, .doc, .pdf, .rtf), páginas Web (html, htm, xhtml, mhtml), imágenes (.jpeg, .jpg, .gif, .png), bases de datos (ISIS), vídeos, audio, mensajes de correo y otros. Tal procesamiento depende, en gran medida, de la utilización de conectores (plugins). A cada tipo de documento corresponde un tipo de conector.

La arquitectura del programa permite la utilización de plugins y clasificadores aún para datos no textuales. Una colección puede tener documentos de origen de diferentes formas. Esta aplicación es capaz de procesar archivos zip, descomprimiéndolos al momento de importarlos.

### 2.2.3.5 Búsqueda y navegación

La opción búsqueda de Greenstone ofrece facilidades para la localización y recuperación de la información contenida en las bibliotecas generadas por este sistema. Para facilitar esta tarea posee un potente motor de búsqueda así como formularios de búsquedas sencilla y avanzada. Además utiliza 2 funciones de búsqueda de información diferente, disponible a partir de:

- Búsquedas a texto completo de los documentos basados en índices según metadatos incluidos en

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

el diseño (título, autor, palabras clave, fecha y otros).

- Mecanismos de navegación, basados en clasificaciones que ofrecen la información ordenada y clasificada conforme a los criterios establecidos en el diseño y vinculados igualmente a los metadatos.

Los formularios de búsqueda permiten seleccionar en su menú desplegable los campos incluidos de acuerdo con el diseño de cada colección y los intereses de los usuarios. Estos campos (metadatos) pueden ser título, autor, fecha de publicación, texto completo, editorial y otros. Los resultados pueden ordenarse por pertinencia o clasificarse mediante un metadato.

Como cada colección es independiente, las búsquedas se pueden ejecutar de manera aislada o entre colecciones por medio de búsquedas cruzadas a condición de que el diseño de las colecciones sea similar. Además, Greenstone admite la ejecución de búsquedas jerárquicas por frases, en ocasiones, considerablemente largas.

### **2.2.3.6 Estándares de metadatos**

Greenstone soporta la utilización de esquemas de metadatos estándar y no estándar. El usuario es quien define el esquema a utilizar a través de la interfaz del bibliotecario. Entre los metadatos que ya vienen predefinidos y pueden seleccionarse para ser utilizados en la colección se encuentran: Dublin Core (esquema básico y cualificado), RFC 1807, NZGLS (New Zealand Government Locator Service), AGLS (Australian Government Locator Service).

También utiliza conectores para importar metadatos preparados externamente en diferentes formatos, por ejemplo: XML, MARC, CDS/ISIS, ProCite, BibTex, Refer, OAI y METS. Incluso, sus registros pueden ser exportados en el formato DSpace para ser directamente importados en un repositorio Dspace.

### **2.2.3.7 Comunidad de usuarios**

Para llevar a cabo la investigación de esta herramienta se pudo contar con la gran comunidad de usuarios que apoyan a la misma, principalmente con la comunidad de usuarios en español de Greenstone (CUEGreenstone) y el grupo de usuarios Greenstone de Latinoamérica (GUGreenstone). En estos sitios se pudo encontrar abundante documentación que sirve de guía para el trabajo con esta aplicación. Además se debe decir que esta ha sido desarrollada y distribuida con la cooperación de la UNESCO. (UNESCO)

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

Actualmente la Red Latinoamericana de soporte para Greenstone se propone crear centros nacionales en todos los países del área, iniciándose con la denominación de tres centros regionales situados en Argentina, Chile y Cuba.

### 2.2.3.8 Ventajas aportadas por Greenstone

- Integración de todos los procesos necesarios para gestionar miles o millones de recursos digitales.
- Potente búsqueda a texto completo y múltiples índices de búsqueda basados en metadatos.
- Disponibilidad de vasta documentación, soporte y mejoramiento.
- Seguridad y control de acceso.
- Integración con otros sistemas.
- Acceso ilimitado a las colecciones vía Internet, intranet o por un CD-ROM
- Posibilidad de automatizar un servicio de tipo biblioteca digital con un mínimo de inversión y sin costos ocultos.
- Posibilidad de compartir experiencias sin restricciones dentro de la comunidad internacional dado su carácter de software fuente abierta.
- Comprende una función de auto instalación que permite instalar fácilmente el programa en Windows y Unix.
- Su aplicación abarca un amplio abanico de posibilidades que cubre las necesidades tanto de pequeñas bibliotecas como las de importantes entidades que requieran programas capaces de satisfacer exigencias más elevadas.

### 2.2.4 Drupal

Drupal es un Sistema de Gestión de Contenido (Content Management System, CMS) basado en módulos implementables, y configurables que permiten mostrar y administrar todo tipo de información. Desde su web se define como una Estructura de Gestión de Contenidos, Content Management Framework (CMF). Puesto que a diferencia de un CMS corriente, se orienta a la capacidad de personalización por parte del

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

administrador, más que a las posibilidades predeterminadas de un CMS. Es de código abierto, con licencia GPL, utiliza php como lenguaje de programación y PostgreSQL o MySQL como gestores de base de datos.

Publica todo tipo de información y utilidades (artículos, imágenes, encuestas, blogs, foros, comercio electrónico) mediante un sencillo sistema de administración de contenido, usuarios y permisos. Es un CMS dinámico, el contenido que muestra y ofrece se almacena en una base de datos que responde a los usuarios mediante un entorno Web. Se destaca por la calidad de sus servicios, su robusta comunidad y las atractivas páginas que genera de una manera sencilla.

Su diseño es especialmente idóneo para construir y gestionar todo tipo de portales en Internet, como webs corporativas, o personales. Además es una herramienta en constante proceso de mejora, para adaptarse a los cambios y tendencias de Internet.

Presenta una curva de aprendizaje más elevada que otros CMS con lenguaje php, pero es un sistema de propósito general diseñado para emplearse en casi cualquier tipo de proyecto, además brinda una serie de componentes para la implementación de requerimientos que están presentes en proyectos de corte bibliotecario. (TRAMULLA, 2010)

Drupal estructura el contenido en una serie de elementos básicos, estos son: nodos, módulos, bloques, menús, permisos de usuario y plantillas.

### **2.2.4.1 Gestión de usuarios**

Drupal dispone de un registro de usuarios y roles que permiten especificar qué tareas pueden realizar de las que ofrecen los módulos instalados en el sitio así como el contenido que se puede crear, al que se puede acceder o el que puede visualizar cada usuario o tipo de usuario (rol).

Con la instalación de Drupal, se crea el usuario administrador y por defecto aparecen dos roles, usuario autenticado y anónimo. Pero es posible crear tantos usuarios adicionales como se quiera y asignarle a nuevos roles personalizados que se crea.

Para cada rol se puede establecer la visibilidad de urls específicas, de tipos de contenido, de menús, de bloques, así como permisos sobre los módulos instalados.

### **2.2.4.2 Módulos**



## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

Los módulos definen las funcionalidades del sitio, los distintos módulos (Información sobre módulos) dan la funcionalidad de Drupal y permiten ampliar la funcionalidad de nuestro sitio o aplicación web. Son archivos que contienen código PHP.

Existen dos tipos de módulos:

Los módulos de core, los cuales se incluyen en Drupal y los contribuidos, proporcionados por la comunidad de Drupal. A excepción de los módulos de core requeridos por Drupal, los módulos pueden ser habilitados o deshabilitados en función de la necesidad del sitio web.

Entre los módulos que provee Drupal para el trabajo en un entorno bibliotecario se encuentran:

**Content Construction Kit (CCK):** Permite crear las distintas estructuras de datos para nuevos tipos de contenido.

**View:** Sirve para la recuperación de información, basado en un asistente para crear las expresiones de búsqueda o selección (tanto en CCK como en VIEW se habla de la familia de módulos, no solo de un módulo).

**Faceted Search:** Permite hacer búsquedas por facetas sin necesidad de instalar Java ni Solr.

**Bibliography:** Manipula bibliografías con la capacidad de importar formatos XML, MARC, Bibtex y otros; maneja diversidad de estilos bibliográficos y se integra con los manejadores de referencias bibliográficas más comunes (end note, zotero, procite).

**Biblio Facets:** Integra los dos anteriores, para búsquedas facetadas en las bibliografías.

**Book Post:** Obtiene información de libros mediante su ISBN, tomada de Open Library Project.

**Marc:** Importa registros MARC, a partir de ficheros mrc o mrk.

**Millenium Integration:** Enlaza con información de Google Books e importa índices de la Biblioteca del Congreso.

**OAI2 for CCK:** Crea un proveedor de datos OAI para tipos de contenidos creados con CCK.

### **Bloques**

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

Los bloques (Bloques) se utilizan para mostrar contenido. Son widget que se ajustan en áreas como las columnas laterales, encabezado, pie de página o cualquiera otra que determine cada tema. Debido a que cada tema puede tener diferentes regiones para ubicar los bloques, hay una página de administración de bloques para cada tema.

Algunos módulos proveen de bloques propios automáticamente, que aparecerán como disponibles en la página de administración de bloques tras la activación del módulo correspondiente.

En la configuración del bloque se pueden determinar diferentes opciones de visibilidad:

- Si los usuarios pueden o no pueden ocultarlo.
- A qué roles de usuarios se les mostrará o no el bloque.
- Las páginas en las cuales se mostrará o no el bloque.

### **Menús**

Los menús (Menús) son colecciones de enlaces que facilitan la navegación por el sitio. Cuando se crea un menú, automáticamente aparece un bloque contenedor que se puede ubicar donde se desee. Como un bloque más, el cual puede ser para uno o varios roles y mostrarse o dejarse de mostrar en una, varias o en todas las páginas de nuestro sitio.

#### **2.2.4.3 Nodos**

El nodo es el tipo de contenido básico. El bloque de construcción de Drupal. Todo tipo de contenido es un nodo. Durante la planificación de una aplicación o sitio web es necesario saber con qué tipos de contenido se va a trabajar, cada uno de ellos será un tipo diferente de nodo.

A cada tipo nodo se le puede asignar permisos de edición y creación basándose en los roles. Existen diferentes tipos de nodos básicos, disponibles sólo si el módulo correspondiente está activado.

Por defecto, están activados tan sólo los nodos de tipo page y story. Los nodos de tipo page no muestran información sobre el autor o la fecha de publicación, los nodos story si lo hacen, estos además por defecto cuando se crea uno, se muestra en la página principal del sitio web.

#### **2.2.4.4 Plantillas**

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

Las plantillas o templates en Drupal son las que establecen la apariencia o estilo de la información a mostrar al usuario. Esta separación entre información y aspecto permite cambiar la apariencia de un sitio web sin modificar el contenido.

### **Tema de un foro**

Los temas de un foro son los bloques de construcción de los foros. Los foros sólo pueden consistir en los temas del foro y sus comentarios, a diferencia de los libros, que puede consistir en casi cualquier tipo de contenido. La información en los foros se clasifica y organiza en contenedores temáticos con una estructura jerárquica y son extremadamente útiles para acoger debates, así como una base comunitaria de apoyo y aprendizaje. Exige que previamente se haya definido una taxonomía que establecerá los distintos foros que contendrán las discusiones.

### **Encuesta**

El tipo encuesta proporciona la posibilidad de hacer preguntas, y proporciona un conjunto de respuestas que se presentan en formato gráfico.

### **Temas**

Los temas (SHREVES, 2008) son la capa de presentación de Drupal. CSS (Cascading Style Sheets), HTML y Javascript pueden ser redefinidos y organizados por un tema en Drupal. Cada tema tiene una serie de características, además de existir unas características definidas que son aplicables a todos los temas. Como el logotipo, el nombre y el eslogan del sitio, los enlaces primarios y los secundarios.

## **2.5 Comparación de las herramientas**

Según las características analizadas de estos programas, en base a las pruebas realizadas, cualquiera de ellos puede utilizarse para satisfacer las necesidades de creación de colecciones digitales.

Se pudo apreciar que todas las herramientas presentan variedad en su interfaz, al ofrecer formatos diferentes en un ambiente web para sus distintos momentos, así los administradores poseen para su trabajo una interfaz y los usuarios finales pueden disfrutar de otra, que resulta amigable, para un buen uso de las colecciones.

Greenstone, Dspace y EPrints, al ser programas de código abierto, gratis y con licencia pública, permiten que los especialistas que los utilicen dispongan de su código fuente, a través del cual pueden configurar

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

sus colecciones a sus necesidades específicas y hacer mejoras a los programas. Además la recuperación de la información en ellos es a través de varios índices, generalmente autor, título, materia y en el texto completo, y mediante la navegación por listados.

Otra concordancia unánime entre estas herramientas es la forma de procesar documentos, pues todas permiten la asignación de metadatos en base a estándares ampliamente distribuidos, como Dublin Core, aunque también brindan la posibilidad de que los especialistas declaren otros, si así lo necesitan. Cabe expresar, también que todas permiten crear colecciones de variedad de formatos y tipos de documentos.

Todas soportan diversos sistemas operativos, así Windows y GNU/Linux son los sistemas operativos que todas pueden usar, además de instalarse en servidores Apache o Tomcat, como generalidad, con otras opciones disponibles. Los usuarios siempre pueden visualizar las colecciones en sus navegadores web, sin necesidad de instalar nuevos programas.

Una vez analizadas y comparadas todas las herramientas, puede afirmarse que EPrints es un sistema que funciona mejor para almacenar documentación textual y no multimedia, así como que no soporta formatos para objetos complejos como METS, MPEG21 o IMS CP, lo que conlleva a que no sea recomendable utilizarlo para la creación de biblioteca digitales, repositorios de objetos y por repositorios que deseen almacenar documentación gráfica, sonora o audiovisual

Por su parte, Dspace y Greenstone son herramientas que soportan cualquier formato de documentos. E incluso los documentos registrados en Greenstone pueden exportarse al formato de Dspace y viceversa.

De ellas Dspace es la más utilizada para la creación de repositorios digitales y la características que más la define es la organización jerárquica de contenidos que propone. La jerarquía comunidad / colección que permite particional el repositorio, pudiendo dar a cada partición diferentes características, permisos de acceso y aspecto. Además, su código se encuentra bien comentariado lo que permite ver lo que el programador ha hecho y se puede modificar o corregir fácilmente.

Greenstone es un sistema para la creación de bibliotecas digitales que permite crear colecciones que pueden publicarse en Internet como también en entornos más específicos como las intranets, así como para acceder a ellas puede ser desde estos sitios o por un CD-ROM. Su desarrollo y distribución son garantizados por la UNESCO. Además permite importar los metadatos de los documentos de forma manual o automática.

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

Por lo anteriormente expuesto el software propuesto para la creación de repositorios digitales es DSpace y para el desarrollo de bibliotecas digitales es Greenstone.

### 2.6 Definición de posibles escenarios

Un escenario es un espacio informático en donde operan determinados comandos, funciones o características comunes. En informática, un escenario puede ser uno de los múltiples lugares posibles en donde se siguen una serie de reglas o se suceden acciones similares de acuerdo con parámetros predeterminados. Además, un escenario permite actuar de manera predecible, ya que el usuario suele contar con información de antemano sobre las características o normas que regulan el funcionamiento de este espacio.

Una vez que ya se conoce el concepto de escenario se puede proceder a definir los posibles escenarios donde pueden interactuar las herramientas analizadas, ya sea de manera independiente o a partir de la integración de algunas de ellas. Los escenarios definidos fueron: escenario en Java, escenario de un repositorio, escenario de una biblioteca digital con interacción de usuarios y escenario en PHP.

#### 2.6.1 Escenario en Java

Un escenario java es basado en la plataforma de Java, es decir, una plataforma basada en el software. Una plataforma es el ambiente de hardware o software en el cual un programa funciona. La plataforma de java está compuesta por tres componentes, de los cuales el más importante es:

- El lenguaje de programación Java.

##### 2.6.1.1 El lenguaje de programación Java

Java es un lenguaje de programación orientado a objetos y posee una curva de aprendizaje muy rápida. Es compilado e interpretado. Proporciona una colección de clases para su uso en aplicaciones de red, que permiten abrir sockets y establecer y aceptar conexiones con servidores o clientes remotos, facilitando así la creación de aplicaciones distribuidas. También proporciona numerosas comprobaciones en compilación y en tiempo de ejecución que le permiten crear software altamente fiable. Posee implementadas barreras de seguridad en el lenguaje y en el sistema de ejecución en tiempo real.

Java está diseñado para soportar aplicaciones que serán ejecutadas en los más variados entornos de red, desde Unix a Windows Nt, pasando por Mac y estaciones de trabajo, sobre arquitecturas distintas y con sistemas operativos diversos.

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

Al integrarse en una página web, las aplicaciones de Java tienen acceso a:

- Recreación de gráficos expertos.
- Interacción en tiempo real con los usuarios.
- Actualización en vivo de la información.
- Interacción instantánea con los servidores a través de la red.

Las aplicaciones de Java pueden obtenerse en cualquier servidor con esta tecnología y funcionan de forma segura bajo cualquier plataforma, permitiendo introducirlas en páginas HTML.

### 2.6.1.2 Herramientas necesarias para trabajar con Java

- **Java JDK**

Java Development Kit (JDK), es un software que permite desarrollar (compilar, ejecutar, generar documentación) programas en lenguaje Java. Posee versiones para prácticamente todos los sistemas operativos e incluye entre sus programas:

- Appletviewer: Es un visor de applet para generar sus vistas previas, ya que un applet carece de método *main* y no se puede ejecutar con el programa java.
- Javac: Es el compilador de JAVA.
- Java: Es el intérprete de JAVA.
- Javadoc: Genera la documentación de las clases java de un programa.

- **NetBeans**

NetBeans es un entorno de desarrollo, hecho principalmente para el lenguaje de programación Java. Es un producto libre y gratuito sin restricciones de uso.

La plataforma NetBeans permite que las aplicaciones sean desarrolladas a partir de un conjunto de componentes de software llamados módulos. Un módulo es un archivo Java que contiene clases de java escritas para interactuar con las APIs de NetBeans y un archivo especial que lo identifica como módulo. Las aplicaciones construidas a partir de módulos pueden ser extendidas agregándole nuevos módulos. Debido a que los módulos pueden ser desarrollados independientemente, las aplicaciones basadas en la

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

plataforma NetBeans pueden ser extendidas fácilmente por otros desarrolladores de software. NetBeans se conoce como la solución más completa para programar en Java.

- **Tomcat**

Tomcat es el servidor web más utilizado a la hora de trabajar con Java en entornos web; es una implementación completamente funcional de los estándares de JSP (JavaServer Pages) y Servlets. También puede especificarse como el manejador de las peticiones de JSP y servlets recibidas por servidores Web populares, como el servidor Apache HTTP de la Fundación de software de Apache o el servidor Microsoft Internet Information Server (IIS).

### **2.6.1.3 Herramienta propuesta y requisitos necesarios para su correcto funcionamiento**

Para el desarrollo de este escenario se propone como herramienta a utilizar a DSpace ya que esta es una herramienta desarrollada en Java y que para la modificación y entendimiento de su código es necesario poseer conocimientos de este lenguaje.

Para la instalación de DSpace se puede usar sistema operativo Windows o GNU/Linux; se requiere un servidor web: Apache Tomcat 4 o superior; entorno de Java: Java JDK 5 o superior, herramienta para la compilación de Java: Apache Ant 1.6.2 o superior; Perl y un servidor de base de datos: PostgreSQL.

### **2.6.2 Escenario de un repositorio**

Para el desarrollo de este escenario se debe tener en cuenta una serie de parámetros que permitirán crear un repositorio digital de acuerdo a las necesidades de la institución o grupo de usuarios y encaminados a satisfacer las requerimientos de información de los mismos.

#### **2.6.2.1 Creación de un equipo de trabajo**

La creación de un equipo de trabajo constituye un principio cuyo cumplimiento garantiza alcanzar el objetivo propuesto, teniendo en cuenta que permite la administración de encuestas o entrevistas académicas y de personal, el procesamiento de los resultados y el desarrollo del modelo de servicio, además de diseñar y ofrecer directrices o incluso ayuda para esclarecer asuntos de derechos de propiedad intelectual. (BARTON Y WATERS, 2005)

#### **2.6.2.2 Estudio de homólogos**

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

El examen de otros repositorios en funcionamiento pondría al equipo al corriente de las tendencias actuales y del estado del arte sobre la implementación del servicio, que junto a la evaluación de las necesidades de la institución conllevarán a un mejor cumplimiento de los objetivos.

### **2.6.2.3 Caracterización de los usuarios y la identificación de las necesidades informativas de los mismos**

Para la caracterización de los usuarios deberán investigarse variables como la edad, el sexo, la ocupación y el nivel de escolaridad. El lugar de conexión a Internet, la frecuencia de uso y la utilización que se hace del recurso, la disponibilidad de buzón de correo electrónico y de los recursos informáticos a su alcance.

### **2.6.2.4 Servicios que se ofrecerán**

Para definir los servicios que se ofrecerán deberá proponerse, a priori, la arquitectura del repositorio, la modalidad de publicación, el formato de los contenidos a archivar, la organización del contenido y la definición de quienes aportan los metadatos.

### **2.6.2.5 Arquitectura del repositorio**

En cuanto a la arquitectura del repositorio, son dos las modalidades que comúnmente se describen:

- Sistema centralizado:
  - Todos los contenidos en una base de datos.
  - Un único repositorio para todas las unidades.
- Sistema distribuido:
  - Una base de datos para cada tipo de documento.
  - Una base de datos para cada unidad organizativa.

Se recomienda la utilización de una arquitectura basada en un sistema centralizado, donde toda la documentación digital se encuentre en un mismo lugar y sea accesible a los usuarios.

### **2.6.2.6 Forma de organizar el contenido**



## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

En cuanto a la forma de organizar el contenido es común el criterio de que el servicio de repositorio lo organice de la forma que mejor responda a los intereses de la institución y de su estructura académica (MONGE Y OVELAR, 2007)

López Medina (BARITÉ, 2005), en relación a la organización del contenido propone dos variantes:

- Por índice: Permite la organización según materias, módulos, entidades, tipos de documentos y otros.
- Jerárquica: Permite la organización según comunidad, subcomunidad, colección, ítem (registro bibliográfico) y fichero.

De las cuales se recomienda usar la segunda porque permite organizar los contenidos conforme a los departamentos u oficinas de la institución, lo cual conllevaría a que cada usuario pueda encontrar con mayor facilidad lo que busca. Además, que de esta forma la búsqueda o recuperación de la información sería más rápida y abarcadora.

### **2.6.2.7 Definición de roles de usuarios**

El funcionamiento de un repositorio se basa en un flujo de trabajo que garantice el control y depósito de los contenidos, se recomienda el establecimiento de diferentes roles de usuarios: creador, editor, administrador, editor de metadatos, anotador y revisor. Para cada uno de los cuales deberán definirse sus respectivas funciones. (BARTON Y WATERS, 2005)

### **2.6.2.8 Selección del software a utilizar**

De las herramientas anteriormente analizadas se propone utilizar para este escenario DSpace porque está construida en base al concepto de comunidades y sub comunidades de una institución y sus colecciones. Además, organiza los usuarios en cuentas personales y grupos, obligando a todos los usuarios a pertenecer a un grupo definido para poder asignarle permisos de escritura, lectura, modificación y eliminación de registros. Los formatos soportados favorece la preservación permitiendo guiar a los usuarios sobre cómo ayudar para que su trabajo depositado perdure.

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

### **2.6.3 Escenario de una biblioteca digital con interacción de usuarios**

Para el desarrollo de este escenario se debe tener en cuenta una serie de requisitos que permitirán crear una biblioteca digital encaminada a satisfacer las necesidades y perspectivas de los usuarios, así como a almacenar y preservar los documentos a largo plazo.

#### **2.6.3.1 Creación de un equipo de trabajo**

Primeramente se debe crear un equipo de trabajo preparado y comprometido con la labor propuesta. Así como elegir los roles que desempeñarán cada uno según los conocimientos y la preparación que presentan, lo cual conllevará un mejor trabajo, con mayor eficiencia y rapidez, pudiendo así cumplir con cada uno de los objetivos propuestos y en el tiempo requerido.

#### **2.6.3.2 Caracterización de los usuarios**

Las necesidades de los usuarios y sus expectativas varían ampliamente. Las bibliotecas digitales deben reconocer, diferenciar y satisfacer sus requerimientos. Idealmente, para el desarrollo de sistemas de búsqueda, debe existir un conocimiento previo de las necesidades más comunes que se presentan, según los tipos de usuarios de la biblioteca, así como una seguridad sobre la satisfacción de sus necesidades.

#### **2.6.3.3 Estudio de homólogos**

El análisis de otras bibliotecas en funcionamiento pondría al equipo al corriente de las tendencias actuales y del estado del arte sobre la implementación del servicio, que junto a la evaluación de las necesidades de los usuarios conllevarán a un mejor cumplimiento de los objetivos.

#### **2.6.3.4 Recolección de documentos**

Es necesario recolectar, con la mayor exhaustividad posible, la información necesaria. Debe seleccionarse y actualizarse regularmente la información útil, sólida y significativa para el trabajo científico o práctico de una o varias disciplinas, en una tendencia de investigación o rama del conocimiento, entre la colosal cantidad de información existente.

#### **2.6.3.6 Seleccionar los documentos a almacenar**

Seleccionar que documentos se deben almacenar es de gran importancia ya que se debe saber discernir entre documentos serios y charlantes de manera que la información sea revisada por expertos en la

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

materia y no escritos pensados al vapor sin tener una base de conocimientos. Además, se debe especificar al tipo de publicación que pertenece. Para ello se debe realizar dos tipos de selecciones.

### ➤ **Selección de los recursos disponibles localmente**

Es esencial que las bibliotecas digitales identifiquen los recursos de información existentes en línea, localmente en sus organizaciones o en organizaciones homólogas, y que los incorporen a sus colecciones. Para esto, se deben realizar esfuerzos sistemáticos y acciones de coordinación entre las organizaciones y los especialistas.

### ➤ **Selección de los recursos disponibles en Internet**

En este paso, debe buscarse la implementación de una política coherente de desarrollo de las colecciones, que persiga el mayor nivel de accesibilidad posible. Es el caso concreto del arte, las redes de galerías y museos, las universidades de arte, entre otras, porque, aun cuando existen muchas editoriales relevantes internacionalmente que colocan en la red sus producciones con acceso a los textos completos, ellas generalmente no incluyen las artes y humanidades.

#### **2.6.3.6 Recuperación de la información**

Es un aspecto muy importante del trabajo de información, que consiste en una serie de operaciones lógicas que garantizan la localización de la información que se necesita. Cualquier procedimiento empleado para desarrollar esta tarea debe permitir, en cualquier momento, recuperar de una colección, aquella información que proporcione una respuesta directa a la pregunta formulada por el usuario.

#### **2.6.3.7 Creación de colecciones**

La construcción de una colección de documentos no es una tarea sencilla, involucra un análisis exhaustivo de la información que se desea almacenar así como las necesidades de los usuarios de dicha colección.

Se debe tener en cuenta que una colección puede estar formada por información de distinta índole (libros, revistas, artículos, foros, videos) y en una variedad inmensa de usuarios. Por lo cual es necesario definir bien que colecciones se van a crear y los metadatos que se utilizarán en ellas para la recuperación de información.

#### **2.6.3.8 Servicios a prestar por la biblioteca**

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

Para que una biblioteca digital tenga éxito debe proveer mecanismos que permitan explotar al máximo las colecciones que la componen. Estos mecanismos no son exclusivos para el acceso a la información sino también para poder compartirla y comentar sobre ella con otros miembros de la comunidad.

Entre los servicios, se puede encontrar los de búsqueda, autenticación, visualización, y uno de sumo interés para la comunidad bibliotecaria: los servicios de interoperabilidad, de manera que las biblioteca pueda compartir sus acervos y poder hacer búsquedas no solo en los archivos locales, sino también en otros.

Otro servicio de gran importancia para el éxito de la biblioteca es la posibilidad de que el usuario pueda tener su biblioteca personal para almacenar los documentos obtenidos en una búsqueda para su posterior consulta sin necesidad de realizar nuevamente la búsqueda.

### ➤ **Biblioteca personal**

La biblioteca personal es el componente que más identidad le da a la biblioteca, tiene el objetivo de que el usuario interactúe con los recursos de la biblioteca, con elementos de infotecnología y con otros usuarios.

Los documentos se pueden organizar en colecciones personales que les permiten a los usuarios acceder de forma inmediata a estos, siempre que estén autenticados en la aplicación. También se pueden añadir amigos y acceder a sus colecciones personales.

#### **2.6.3.9 Selección del software a utilizar**

De las herramientas anteriormente analizadas se propone utilizar para este escenario la integración de Greenstone con Drupal. Greenstone será la herramienta para la creación de la biblioteca digital ya que permite distribuir los documentos en colecciones que se pueden crear atendiendo a sus características, así como la importación y asignación de metadatos, de manera manual o automática. Permite configurar el tipo de indexación a utilizar, modalidades de búsqueda y obtener listados de resultados en función a la relevancia con los términos consultados. Es totalmente compatible con el protocolo OAI-PMH, el cual permite la interoperabilidad, permitiendo el intercambio de información entre diferentes fuentes bajo el principio del libre acceso, pero respetando los derechos de autor.

Drupal se utilizará para crear la interfaz gráfica con que interactuará el usuario, así como permitirá la creación de la biblioteca personal del usuario, la asignación y creación de roles y usuarios. Además de permitir la creación de foros y otras funcionalidades que proveen los módulos desarrollados en él.

Para integrar Drupal con Greenstone se puede realizar de dos formas:

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

Mediante la incorporación de un *iframe* a la web desarrollada con *Drupal*. De esta manera se acopla *Greenstone* en el gestor de contenidos, y aparenta ser el mismo portal, pero funcionan de manera independiente.

Otra forma más compleja consiste en crear un formulario en *Drupal* donde se describan los documentos, e importarlos a *Greenstone* de manera automática, usando XML o SOAP (siglas de Simple Object Access Protocol) soportado por Perl (*Greentsone*) y php (*Drupal*). De esta manera tanto la gestión del repositorio, como su consulta se desarrollan íntegramente en *Drupal*. Además se deben de utilizar módulos que mejoren la presentación y uso de la información. Por otro lado *Drupal* y *Greenstone* se pueden unir con las posibilidades que permiten algunos módulos (nodewords, biblio, marc, OAI-PMH).

### 2.6.4 Escenario en PHP

Un escenario PHP es basado en el entorno LAMP, que es conjunto de softwares necesarios para alcanzar una solución global. Para lograr este entorno es necesaria la unión de las siguientes tecnologías:

- GNU/Linux: El sistema operativo.
- Apache: El servidor web.
- MySQL o PostgreSQL: El gestor de bases de datos.
- PHP: Lenguaje de programación.

#### 2.6.4.1 El lenguaje de programación PHP

PHP es el acrónimo de Hypertext Preprocessor (Preprocesador de Hipertexto). Se trata de un lenguaje interpretado de alto nivel embebido en páginas HTML y ejecutado en el servidor. Está muy orientado al desarrollo de aplicaciones web y permite insertar contenidos dinámicos en las páginas. Funciona tanto en sistemas Windows como GNU/Linux utilizando como servidor web Apache. Sin embargo, las mejores prestaciones del lenguaje se obtienen trabajando en un entorno Unix o GNU/Linux.

Es un lenguaje basado en herramientas con licencia de software libre, es decir, no hay que pagar licencias, ni está limitado en su distribución y, es posible ampliarlo con nuevas funcionalidades si así se desea. Tiene soporte para la programación orientada a objetos. Su característica más fuerte es su amplio soporte para una gran cantidad de bases de datos, tales como MySQL y PostgreSQL.

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

Es usado principalmente en interpretación del lado del servidor, pero actualmente puede ser utilizado desde una interfaz de línea de comandos o en la creación de otros tipos de programas incluyendo aplicaciones con interfaz gráfica usando las bibliotecas Qt o GTK+.

### **2.6.4.2 Apache**

Apache es un servidor web HTTP de código abierto para plataformas Unix (BSD, GNU/Linux.), Microsoft Windows, Macintosh y otras, que implementa el protocolo HTTP. Es usado principalmente para enviar páginas web estáticas y dinámicas en Internet. Es personalizable, su arquitectura modular permite construir un servidor hecho a la medida. Además permite la implementación de los últimos y más nuevos protocolos.

Es el servidor web hecho por excelencia, su configurabilidad, robustez y estabilidad hacen que cada vez millones de servidores reiteren su confianza en este programa. La licencia Apache es una descendiente de la licencias BSD. Esta licencia permite hacer lo que se quiera con el código fuente. Trabaja con gran cantidad de lenguajes como Perl, PHP y otros lenguajes de script.

### **2.6.4.3 PostgreSQL**

PostgreSQL es un sistema de gestión de bases de datos objeto-relacional, que incluye características de la orientación a objetos, como son la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional. Pero a pesar de ello, PostgreSQL no es un sistema de gestión de bases de datos puramente orientado a objetos.

Soporta distintos tipos de datos, además del soporte para los tipos base, también soporta datos de tipo fecha, monetarios, elementos gráficos, datos sobre redes (MAC, IP), cadenas de bits. También permite la creación de tipos propios. Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos.

### **2.6.4.4 Herramienta propuesta y requisitos necesarios para su correcto funcionamiento**

Para el desarrollo de este escenario se propone como herramienta a utilizar a Drupal ya que esta es una herramienta desarrollada en PHP y que para la modificación y entendimiento de su código es necesario poseer conocimientos de este lenguaje. Además para su correcto funcionamiento necesita del servidor web Apache y una base de datos que puede ser PostgreSQL.

## **2.7 Conclusiones del capítulo**

## Capítulo II: Propuesta de las herramientas a utilizar en el desarrollo de catálogos y bibliotecas digitales

---

Con el fin de este capítulo se logró obtener un análisis de las herramientas a utilizar para la creación de repositorios y bibliotecas digitales. A partir de este se llevó a cabo una comparación entre ellas y una propuesta por posibles escenarios donde se demostró que para el desarrollo de repositorios DSpace es la herramienta más indicada y para la creación de bibliotecas digitales la mejor opción es la integración de Greenstone y Drupal; Greenstone para el desarrollo de las colecciones y almacenamientos de los documentos y Drupal para la creación de la interfaz gráfica y el desarrollo de algunas funcionalidades como la biblioteca personal del usuario.

### **Conclusiones Generales**

Con la realización del trabajo expuesto, de forma general se le ha dado cumplimiento al objetivo general y a cada uno de los objetivos específicos de la investigación que habían sido definidos. Se realizó un estudio del estado del arte de los principales conceptos relacionados con los sistemas de información tipo biblioteca, principalmente los correspondientes con la creación de repositorios y bibliotecas digitales, pudiéndose así definir las principales herramientas que se utilizan para el desarrollo de estos últimos, las cuales son: DSpace, Greenstone y EPrints.

También se realizó un estudio de las principales características de las herramientas seleccionadas para la gestión de repositorios y bibliotecas digitales, así como unas series de pruebas que permitieron conocer mejor de las principales funcionalidades que realizan cada una de ellas.

Además se definieron 4 escenarios principales donde se definió que para el desarrollo de repositorios digitales DSpace es la herramienta más indicada y para la creación de bibliotecas digitales la mejor opción es la integración de Greenstone y Drupal.



### **Recomendaciones**

Luego de finalizada la investigación y cumplido el objetivo general de la misma, en vista de lograr alcanzar mayor conocimiento y preparación del tema, se recomienda:

- Mantener un estudio sostenido de la evolución de DSpace, Greenstone y el resto de las herramientas más utilizadas y reconocidas por su eficiencia para el desarrollo de repositorios y bibliotecas digitales.
- Preparar un equipo de trabajo en los diferentes componentes y características de las herramientas DSpace y Greenstone, así como lenguajes y herramientas que utilizan, con el objetivo de agregar nuevas funcionalidades a los repositorios y bibliotecas digitales desarrolladas con las mismas.

### Bibliografías

ACUÑA, MARÍA JOSÉ DE Y AGENJO, XAVIER. 2005. Archivos en la era digital: problema (y solución) de los recursos electrónicos. En: El profesional de la información. [Consultado el: 13 de enero 2011]. Disponible en: <[http://scielo.sld.cu/scielo.php?pid=S1024-94352007001200006&script=sci\\_arttext](http://scielo.sld.cu/scielo.php?pid=S1024-94352007001200006&script=sci_arttext)>.

BARITÉ, M. 2005. Diccionario de Organización y Representación del Conocimiento. KODictionary. [Consultado el: 2 marzo 2011]. Disponible en:<[http://www.eubca.edu.uy/diccionario/letra\\_r.htm#Le tra%20A](http://www.eubca.edu.uy/diccionario/letra_r.htm#Le tra%20A)>.

BARTON, MARY R. Y WATERS, MARGARET M. 2005. Cómo crear un Repositorio Institucional.pdf. [Consultado el: 28 febrero 2011]. Disponible en: <<http://www.recolecta.net/buscador/documentos/mit.pdf>>.

BEITRA OLIVA, ELENA MARGARITA. 2008. Selección de recursos disponibles en Internet para el desarrollo de colecciones de la biblioteca virtual. [Consultado el: 15 de noviembre de 2010]. Disponible en: <<http://www.scielo.org.pe/pdf/biblios/n31/a07n31.pdf>>.

Bloques. [Consultado el: 9 febrero 2011]. Disponible en: <<http://drupal.org/handbook/modules/block>>.

BNJM. Biblioteca Nacional José Martí. [Consultado el: 16 enero 2011]. Disponible en: <<http://www.bnjm.cu/salas.htm>>.

CABRAL VARGAS, BRENDA, RAMÍREZ MEDINA, REYNA Y FIGUEROA ALCÁNTARA, HUGO ALBERTO. 2007. La biblioteca digital. [Consultado el: 8 de enero 2011]. Disponible en: <<http://ru.ffyl.unam.mx:8080/jspui/bitstream/10391/949/1/Brenda%20Cabral%20-%20Reyna%20Ram%C3%ADrez%20-%20%20Hugo%20Figueroa%20-%20La%20biblioteca%20digital.pdf>>.

CLACSO. Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe de la Red de Centros Miembros de CLACSO. [Consultado el: 16 enero 2011]. Disponible en: <<http://www.biblioteca.clacso.edu.ar/>>.

CLIFFORD A. LYNCH. 2003. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. [Consultado el: 11 de enero 2011]. Disponible en: <<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>>.

## Bibliografías

---

CUEGreenstone. Comunidad de usuarios Greenstone en Español. [Consultado el: 3 febrero 2011]. Disponible en: <<http://www.prodigioconsultores.com/?p=72>>.

Digital Libraries Federation. DLF draft strategy and business plan. 2001. [Consultado el: 15 de noviembre de 2010]. Disponible en: <<http://old.diglib.org/about/strategic.htm>>.

Dspace-general Info Page. [Consultado el: 28 enero 2011]. Disponible en: <<http://mailman.mit.edu/mailman/listinfo/dspace-general>>.

DSpace-tech Info Page. [Consultado el: 28 enero 2011]. Disponible en: <<https://lists.sourceforge.net/lists/listinfo/dspace-tech>>.

Dspace-devel Info Page. [Consultado el: 28 enero 2011]. Disponible en: <<https://lists.sourceforge.net/lists/listinfo/dspace-devel>>.

EcuRed. Enciclopedia cubana - EcuRed. [Consultado el: 16 enero 2011]. Disponible en: <[http://www.ecured.cu/index.php/EcuRed:Enciclopedia\\_cubana](http://www.ecured.cu/index.php/EcuRed:Enciclopedia_cubana)>.

E-Prints C. Archivo Institucional E-Prints Complutense. [Consultado el: 16 enero 2011]. Disponible en: <<http://eprints.ucm.es/>>.

ESCOLAR SOBRINO, H. 1990. Historia de las bibliotecas. Madrid: Fundación G. Sánchez Ruipérez. [Consultado el: 15 de noviembre de 2010].

evalDSpace\_v01. 2009. Evaluación de software para organización de recursos digitales [Consultado el: 27 Enero 2011]. Disponible en: <[http://ricabib.cab.cnea.gov.ar/90/1/evalDSpace\\_v01.pdf](http://ricabib.cab.cnea.gov.ar/90/1/evalDSpace_v01.pdf)>.

evalEPrints\_v02.pdf. 2009. Evaluación de software para organización de recursos digitales [Consultado el: 30 Enero 2011]. Disponible en: <[http://ricabib.cab.cnea.gov.ar/91/1/evalEprints\\_v02.pdf](http://ricabib.cab.cnea.gov.ar/91/1/evalEprints_v02.pdf)>.

FLORRES CUESTA, GILDA Y SÁNCHEZ TARRAGÓ, NANCY. 2007. Los repositorios institucionales: análisis de la situación internacional y principios generales para Cuba. ACIMED v.16 n.6. [Consultado el: 11 de enero 2011]. Disponible en: <[http://scielo.sld.cu/scielo.php?pid=S1024-94352007001200006&script=sci\\_arttext](http://scielo.sld.cu/scielo.php?pid=S1024-94352007001200006&script=sci_arttext)>.

GÓMEZ, LAUREANO FELIPE. 2007. Interoperabilidad en los Sistemas de Información Documental (SID): La Información Debe Fluir. [Consultado el: 13 de enero 2011]. Disponible en: <<http://redalyc.uaemex.mx/pdf/953/95330103.pdf>>.

## Bibliografías

---

---

GUGreenstone. Grupo de Usuarios Greenstone de Latinoamérica. [Consultado el: 3 febrero 2011]. Disponible en: <<http://www.greenstone-la.org/>>.

HERVÉ LE CROSNIER. 2006. Bibliotecas digitales. [Consultado el: 8 de enero 2011]. Disponible en: <<http://vecam.org/article627.html>>.

Información sobre módulos. [Consultado el: 8 febrero 2011]. Disponible en: <<http://drupalmodules.com/>>.

LOOMS, THELMA Y CHRISTENSEN, CLARK. 2002. Advanced Distributed Learning Emerging and Enabling Technologies for the Design of Learning Object Repositories Report. [Consultado el: 11 de enero 2011]. Disponible en: <<http://xml.coverpages.org/ADLRepositoryTIR.pdf>>.

LÓPEZ GUZMÁN, CLARA Y GARCÍA PEÑALVO, FRANCISCO J. 2006. Repositorios de objetos de aprendizaje: bibliotecas para compartir y reutilizar recursos en los entornos e-learning. [Consultado el: 11 de enero 2011]. Disponible en: <<http://redalyc.uaemex.mx/pdf/285/28590202.pdf>>.

LÓPEZ MEDINA, A. 2007. Guía para la puesta en marcha de un repositorio institucional. Madrid: SEDIC. [Consultado el: 10 de enero 2011].

MELERO, REMEDIOS. 2005. Significado del acceso abierto a las publicaciones científicas: definición, recursos, copyright e impacto. [Consultado el: 11 de enero 2011]. Disponible en: <<http://eprints.rclis.org/bitstream/10760/6571/1/EPI-rmelero.pdf>>.

MELERO, R. 2005. Acceso abierto a las publicaciones científicas: definición, recursos, copyright e impacto. En: El Profesional de la Información. [Consultado el: 15 de enero 2011].

Menús. [Consultado el: 9 febrero 2011]. Disponible en: <<http://drupal.org/handbook/modules/menu> y <http://drupal.org/node/120632>>.

MONGE BENITO, S. Y OVELAR BELTRÁN, R. 2007. Repositorio 2.0: Dinámicas sociales para favorecer el desarrollo de comunidad en torno a un repositorio de contenidos educativos digitales. [Consultado el: 28 Febrero 2011]. Disponible en: <<http://spdece07.ehu.es/actas/Monge.pdf>>.

MUÑOZ CASALS, VELMOUR. 2008. Repositorio de documentos. [Consultado el: 9 de enero 2011]. Disponible en: <<http://www.monografias.com/trabajos-pdf/repositorio-documentos/repositorio-documentos.pdf>>.

NÜRNBERG, P. J. 1995. Digital libraries: issues and architecture. En The Second Annual Conference on Theory and Practice of Digital Libraries, Texas A&M University, College Station.

## Bibliografías

---

---

ORTIZ REPISO, VIRGINIA Y MOSCOSO, PURIFICACIÓN. 2006. La biblioteca digital: inventando el futuro. [Consultado el: 8 de enero 2011]. Disponible en: <<http://www.bibliociencias.cu/gsd/collect/eventos/import/La%20Biblioteca%20Digital.pdf>>.

PÉREZ LÓPEZ, ANA. 1995. HACIA UNA ECOLOGÍA DE LA DOCUMENTACIÓN: La Biblioteca como Sistema de Información. [Consultado el: 15 de noviembre de 2010]. Disponible en: <<http://eprints.rclis.org/bitstream/10760/7745/1/ECOBIBLIOTECA.pdf>>.

PÉREZ, M. Y SILVA, LF. 2007. Cómo funciona el protocolo OAI – PMH en la recuperación de información. [Consultado el: 15 de enero 2011].

RTD. Repositorio de Tesis Doctorales. [Consultado el: 16 enero 2011]. Disponible en: <<http://tesis.repo.sld.cu/information.html>>.

SHREVES, RIC. 2008. Drupal 6 Themes. Packt Publishing. [Consultado el: 10 febrero 2011].

SOLA MARTÍNEZ, MARIA JOSÉ. 2008. Greenstone: distribución de colecciones digitales. [Consultado el: 2 febrero 2011]. Disponible en: <<http://www.ub.edu/bid/21/sola2.htm>>.

TERÁN, JORGE Y GONZÁLEZ, VICTOR. 2009. Dspace. [Consultado el: 25 Enero 2011]. Disponible en: <<http://www.slideshare.net/jvelizster/dspace-2004449>>.

TRAMULLA, J. 2010. Drupal para Bibliotecas y Archivos. [Consultado el: 8 febrero 2011].

UASB. Repositorio UASB-Digital. [Consultado el: 16 enero 2011]. Disponible en: <<http://repositorio.uasb.edu.ec/>>.

UNESCO. 2010. Greenstone: Un software libre de código abierto para la construcción de bibliotecas digitales. Experiencias en América Latina y el Caribe. [Consultado el: 4 febrero 2011]. Disponible en: <<http://unesdoc.unesco.org/images/0018/001887/188719s.pdf>>.

VERA, F. 2006. Construcción de Bibliotecas Digitales, modelo de capa piramidal. [Consultado el: 20 enero 2011].

VOUSSÁS MÁRQUEZ, JUAN. 2006. Bibliotecas y publicaciones digitales. México. UNAM, Centro Universitario de Investigaciones Bibliotecológicas. [Consultado el: 15 de noviembre de 2010].

### Glosario de Términos

**Acceso abierto / Open Access:** Disponibilidad gratuita en el Internet público, que permita que cualquier usuario lea, descargue, copie, distribuya, imprima, busque, o ligue los textos completos de los artículos, pasarlos como datos a un software, o utilizarlos con cualquier otro propósito legal, sin ningún tipo de barrera económica, legal o técnica.

**Código abierto:** Aplicaciones computacionales que tienen su código fuente liberado; en general, los programas de código abierto suelen ser libres.

**Dublin Core:** Esquema 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.

**GPL:** La GNU General Public License (inglés: Licencia Pública General) es una licencia creada por la Free Software Foundation y orientada principalmente a los términos de distribución, modificación y uso de software. Su propósito es declarar que el software cubierto por esta licencia es Software Libre.

**MARC21:** Es un registro catalográfico legible por máquina (MACHINE-Readable Cataloging) "Legible por máquina" significa que un tipo particular de máquina, una computadora, puede leer e interpretar los datos contenidos en un registro catalográfico.

**Movimiento de Acceso Abierto:** Se basa fundamentalmente en colocar a disposición de la comunidad científica y del público general, en forma gratuita y libre, tanto los artículos científicos como materiales docentes y para la investigación por medio de su publicación en revistas de acceso abierto y el depósito de estos en repositorios institucionales o temáticos con igual acceso.

**Multimedia:** Combinación de texto, sonido, gráficos, animación y video en un programa. Debe incluir una interfaz de usuario interactiva.

**Metadatos:** El término «metadatos» no tiene una definición única. Según la definición más difundida de metadatos es que son «datos sobre datos». También hay muchas declaraciones como «informaciones sobre datos», «datos sobre informaciones» e «informaciones sobre informaciones».

**Módulo:** Es un software que agrupa un conjunto de subprogramas y estructuras de datos.

## Glosario de Términos

---

---

**OPAC:** (Online Public Access Catalog) Es un catálogo automatizado de acceso público en línea de los materiales de una biblioteca. Generalmente, tanto el personal de la biblioteca como el público tienen acceso a él en varias terminales dentro de la biblioteca o desde el hogar vía Internet.

**Servicio:** Acción de una organización destinada a satisfacer necesidades.

**Software Libre:** Se refiere a la libertad de los usuarios para ejecutar, copiar, distribuir, estudiar, cambiar y mejorar el software.

**XML:** siglas en inglés de Extensible Markup Language (lenguaje de marcas extensible), es un metalenguaje extensible de etiquetas desarrollado por el World Wide Web Consortium (W3C). XML no ha nacido sólo para su aplicación en Internet, sino que se propone como un estándar para el intercambio de información estructurada entre diferentes plataformas.

**Z3950:** Es el nombre de un estándar definido por ANSI/NISO que permite comunicar sistemas que funcionan en distinto hardware y usan distinto software. Fue diseñado para solucionar los problemas asociados a la búsqueda en múltiples bases de datos con diferentes lenguajes y procedimientos.