

Bibliotecas digitales: situación actual y problemas

Por Pedro Hípola, Benjamín Vargas-Quesada y José A. Senso

Resumen: La producción bibliográfica sobre las bibliotecas digitales se está incrementando de forma espectacular. El objetivo del presente trabajo es analizar las principales definiciones existentes sobre este tipo de bibliotecas y presentar los problemas a los que se enfrentan, todo ello con el fin de proporcionar una visión general de su situación actual.

Palabras clave: Biblioteca digital, Sistemas bibliotecarios, Diseño e implementación, Sistemas de información, Metadatos.

Title: Digital libraries: current situation and problems

Abstract: Literature on digital libraries is increasing dramatically. The objective of the present work is to provide a global vision of the current status of digital libraries by analysing the main definitions currently used to describe them as well as by exploring the principal problems encountered.

Keywords: Digital libraries, Library systems, Design and implementation, Information systems, Metadata.

Hípola, Pedro; Vargas-Quesada, Benjamín; Senso, José A. «Bibliotecas digitales: situación actual y problemas». En: *El Profesional de la información*, 2000, abril, v. 9, n. 4, pp. 4-13.

Tabla 1

Algunas fuentes de información electrónicas sobre bibliotecas digitales

- American memory
<http://lcweb2.loc.gov/ammem/ammemhome.html>
- Annotated bibliography of digital library related sources
<http://robotics.stanford.edu/users/ketchpel/annbib.html>
- ARL digital initiatives database
<http://www.arl.org/did>
- California digital library
<http://www.cdlib.org>
- Canadian initiative on digital libraries
<http://www.nlc-bnc.ca/cidl/cidle.htm>
- Digital libraries initiative
<http://www.dli2.nsf.gov>
- Digital library federation
<http://www.clir.org/diglib/dlfhomepage.htm>
- Ariadne
<http://www.ariadne.ac.uk>
- D-Lib magazine
<http://www.dlib.org>
- Journal of digital information
<http://journals.ecs.soton.ac.uk/jodi>

Es uno de los principales temas de estudio, investigación y desarrollo en todo el mundo bibliotecario. Está provocando la aparición de un gran número de proyectos apoyados y financiados por agencias gubernamentales, tanto nacionales como internacionales, así como por otras entidades privadas.

La Comisión Europea apoya y financia proyectos de este tipo a través de *Delos* (*Digital library working group*), un grupo de trabajo sobre bibliotecas digitales que forma parte de *Ercim* (*European research consortium for informatics and mathematics*).

<http://www.ercim.org/activity/index.html>

En el Reino Unido, *Elib* (*Electronic libraries programme*) ayuda y respalda la implantación de dieciséis proyectos de bibliotecas digitales. En Estados Unidos, los seis grandes proyectos recogidos en la *DLI-1* (la primera *Digital library initiative*) apoyados por *NSF*, *Nasa* y *Darpa*, se encuentran ya en su fase final de desarrollo. La *DLI-2* (la segunda *Digital libraries initiative*) comienza a funcionar este año. Se diferencia de su predecesora en que el número de agencias encargadas de sostener el proyecto se ha visto aumentado con la *National Library of Medicine*, la *Library of Congress* y la *National Endowment for the Humanities*. Además, el presupuesto con respecto a la fase anterior se ha duplicado.

<http://www.ukoln.ac.uk/services/elib/>

<http://www.dli2.nsf.gov/>

http://www.cise.nsf.gov/iis/dli_home.html

El hecho de que entidades, tanto públicas como privadas, se encuentren involucradas en la materialización de las bibliotecas digitales ha dado lugar a la apa-

Tabla 2. Normas y protocolos relacionados

Función	Norma
Consulta	z39.50
Petición	Http (hypertext transfer protocol), ISO 10160/10161 (préstamo interbibliotecario), z39.50
Entrega	EDI (electronic data interchange), Http
Metadatos	Dublin Core EAD (encoding archival description), Marc MCF (meta content framework), RDF (resource description format), TEI (text encoding initiative)
Identificadores	DOI (digital object identifier), Isbn (international standard book number), Issn (international standard serial number), Sici (serial item and contribution identifier), URI (uniform resource identifier), URN (uniform resource name)
Seguridad	SET (secure electronic transaction), SSL (secure socket layer)
Derechos de autor	DOI (digital object identifier), Dublin Core RDF (resource description format)
Información	GIF (graphics interchange format), Html (hypertext markup language), Jpeg (joint photographic experts group), PDF (portable document format), Sgml (standard generalized markup language), Unicode XML (extensible markup language)

rición de sitios web creados y mantenidos por ellas mismas y que recogen información sobre estudios, proyectos y avances realizados en esta materia. Éste es el caso de *Ukoln* (*UK office for library and information networking*), el de *Ifla*, etc.

<http://www.ukoln.ac.uk>

<http://www.ifla.org/III/diglib.htm>

Actualmente existe una gran cantidad de bibliografía sobre los diferentes aspectos que afectan a este tipo de bibliotecas, y que puede ayudarnos a seguir la pista de la creciente masa de información, tanto electrónica como impresa, que se genera acerca del tema. Un ejemplo de ello se recoge en la tabla 1.

Como consecuencia del constante crecimiento del número de publicaciones relacionadas con las distintas

áreas de la biblioteca digital, cada día es más difícil para los investigadores, los profesionales de la información y, en definitiva, para el usuario en general, estar al día sobre los avances y nuevas expectativas creadas por este tipo de bibliotecas y todos los proyectos que se generan a su alrededor.

El objetivo de este trabajo no es sólo presentar el estado del arte sobre la definición de las bibliotecas digitales, sino también informar de los principales problemas a los que se enfrentan y de qué modo lo hacen. Para ello seguiremos el esquema utilizado por autores como **Lynch** y **García-Molina** (1995), **Arms, Bianchi**, y **Overly** (1997), **Faba** y **Moya** (1999) o **Chowdhury** y **Chowdhury** (1999) entre otros.

¿Qué es la biblioteca digital?

Las expresiones “biblioteca digital” y “biblioteca electrónica” son usadas de forma indistinta en numerosas publicaciones. Incluso la segunda forma ha llegado a ser en los últimos años más popular que la primera. Pero no deben ser confundidas, ya que la biblioteca digital implica el uso de las telecomunicaciones, es decir, el acceso a la información puede ser realizado de forma remota e independiente del lugar y número de conexiones. Otra expresión es la de “biblioteca virtual”, que se utiliza muy a menudo para hacer referencia a la biblioteca digital. Aunque muy parecida, tampoco es equivalente, puesto que una colección de información digital normalmente se encuentra disponible en una sola localización, mientras que una bi-

Tabla 3. Elementos de Dublin Core

Contenido	Propiedad intelectual	Otros
Título	Autor	Fecha
Materia	Editor	Tipo
Resumen	Contribución	Formato
Fuente	Derechos	Identificador
Idioma		
Relación		
Cobertura		

URLs de interés

American Memory.
<http://lcweb2.loc.gov/ammem/ammemhome.html>

Annotated Bibliography of Digital Library Related Sources.
<http://robotics.stanford.edu/users/ketchpel/annbib.html>

Ariadne.
<http://www.ariadne.ac.uk>

ARL Digital Initiatives Database.
<http://www.arl.org/did>

California Digital Library.
<http://www.cdlib.org>

Canadian Initiative on Digital Libraries.
<http://www.nlc-bnc.ca/cidl/cidle.htm>

Digital Libraries Initiative.
<http://www.dli2.nsf.gov>

Digital Libraries: Resources and Projects.
<http://www.ifla.org/II/diglib.htm>

Digital Library Federation.
<http://www.dlib.org/diglib/dlffhomepage.htm>

D-Lib Magazine.
<http://www.dlib.org>

Dublin Core.
http://purl.org/metadata/dublin_core

ET-Map.
<http://ai.bpa.arizona.edu/et>

European Research Consortium for Informatics and Mathematics.
<http://www.ercim.org/activity/index.html>

First Digital Library Initiative.
http://www.cise.nsf.gov/iis/dli_home.html

Ifla.
<http://www.ifla.org/II/diglib.htm>

Internet Public Library. Scorpion.
<http://orc.rsch.oclc.org:6109>

Journal of Digital Information.
<http://journals.ecs.soton.ac.uk/jodi>

Ncstrl.
<http://www.ncstrl.org>

Resource Description Framework (RDF).
<http://www.w3.org/Press/RDF>

The Big Picture.
<http://www.public.iastate.edu/~CYBERSTACKS/BigPic.htm>

The Electronic Engineering Library.
<http://eels.lub.lu.se>

The Internet Scout Project.
<http://scout.cs.wisc.edu/index.html>

UK Office for Library and Information Networking.
<http://www.ukoln.ac.uk>

Visual Site Map.
<http://lislin.gws.uky.edu/Sitemap/Sitemap.html>

World Wide Web Consortium.
<http://www3.org>

bliblioteca virtual suele estar formada por un conjunto de conexiones a varias fuentes de información recogidas en internet.

También se ha usado la expresión “biblioteca híbrida”, abanderada por el programa *eLib*, que recoge sobre todo la visión británica de lo que es la biblioteca digital. Con este nombre se designa la evolución/transición de las actuales unidades informativas hacia las bibliotecas digitales, pues sus fondos no son totalmente impresos, ni electrónicos ni digitales, y por tanto se debe hacer uso de las últimas tecnologías para conseguir la mejor conjunción de ambos mundos (**Rusbridge**, 1998; **Oppenheim** y **Smithson**, 1999).

A la hora de definir la biblioteca digital se pueden observar varias tendencias. Existe una preferencia entre los investigadores a considerarla como una *colección* cuyo contenido representa a una comunidad de usuarios, mientras que los profesionales de la información la ven más bien como una *entidad de servicios* (**Borgman**, 1999).

Además de estas opiniones, que son las más aceptadas, hay también quien considera que ni la biblioteca digital, ni la electrónica, ni la virtual, tienen por qué llamarse biblioteca, sino que todas ellas, de forma individual o conjunta, deberían denominarse sistema de información digital, o sistema digital de acceso a la información (**Miksa** y **Doty**, 1995).

Por otra parte, nos encontramos con defensores de la idea de que la biblioteca digital debe ser un punto de encuentro entre profesionales de la información, informáticos, ingenieros electrónicos, sociólogos, antropólogos..., cuyo objetivo fundamental es mejorar la organización y distribución de los sistemas de información (**Cronin**, 1998).

La *Digital Library Federation (DLF)*, recogiendo la definición de **Waters** (1998), señala a la biblioteca digital como una organización que suministra recursos informativos para seleccionar, estructurar, ofrecer acceso intelectual, interpretar, distribuir, preservar la integridad y asegurar la existencia a través del tiempo de las colecciones, todo ello con el fin de que puedan ser utilizadas por una comunidad determinada o un grupo de comunidades, de la manera más fácil y económica posible.

Para el propósito de este trabajo nosotros nos adherimos a la que surgió del *Digital library workshop* de marzo de 1994 (**Gladney et al.**, 1994): “una biblioteca digital es una colección de información almacenada de tal forma que, mediante el uso del software adecuado, pueda reproducir, emular y ampliar los servicios suministrados por las bibliotecas actuales basadas en el papel y en otros sistemas de recopilación, ca-

atalogación, búsqueda y difusión de la información. La biblioteca digital debe suministrar todos los servicios tradicionales de las actuales bibliotecas y explotar al máximo las ventajas del almacenamiento, recuperación y difusión digital de la información”.

De acuerdo con esta definición podemos asumir que la información asociada a las bibliotecas digitales será: texto, datos numéricos, ilustraciones, fotografías, sonido, vídeo, diapositivas..., en definitiva, cualquier ejemplo de “objeto digital”. Estos objetos representan la información almacenada en los depósitos documentales y están compuestos por su propio contenido (bits), metadatos o propiedades asociadas a ellos, un registro de operaciones realizadas y etiquetas para su identificación.

«Butler defiende que recopilar todo tipo de información digital no es sostenible, ni incluso deseable, a largo plazo»

Una biblioteca digital debería contemplar cualquiera de los siguientes puntos a la hora de ofrecer sus servicios:

— Los usuarios podrán solicitar información de todo tipo, independientemente de su área geográfica y, a menudo, desearán relacionarla con otra que sea de su interés.

— Los usuarios serán muy variados: autores, editores, centros de información tradicionales, usuarios finales...

— Gran cantidad de información no está todavía en formato digital, por lo que en lugar de digitalizarla sin más, lo mejor será recoger los metadatos de índices y catálogos de monografías y publicaciones periódicas no digitalizadas.

— Estará construida sobre sistemas y plataformas convencionales que permitan la compatibilidad e integración con otros estándares y procedimientos integrados.

— La biblioteca digital deberá incluir servicios de seguridad y autenticación en las transmisiones, control de la propiedad intelectual, facturación de servicios, pagos, etc.

Como se puede observar, el funcionamiento que debe regir una biblioteca digital está basado en suposiciones o hechos deseables. Es decir, se trata de metas a las que se aspira pero que están en fase de consecución. Gran parte de los esfuerzos e investigaciones se centran en resolver los escollos o problemas necesarios para conseguir que estos hechos deseables se conviertan en realidades.

Componentes

La mayoría de los aspectos relacionados con la gestión de una biblioteca digital están íntimamente relacionados con aquellos que se desarrollan en las bibliotecas “reales” (gestión de la colección, adquisición...).

Ampliando el modelo propuesto por Murray (1999), y sin ánimo de establecer un orden o jerarquía entre cada componente, podemos distinguir entre módulo *básico* y *extendido*. El primero de ellos, que incluye los componentes esenciales para que la biblioteca pueda ofrecer sus servicios de forma eficaz, estaría formado por:

— Difusión selectiva de la información. Para ello sería necesario el uso de tecnologías push que permitieran gestionar de la mejor forma posible todos los perfiles de usuario y que agilizaran la comunicación entre éstos y el centro.

— Seguridad. Por un lado referida al control de entrada de usuarios al sistema y, por otro, al pago de servicios. Dentro de este apartado se contempla también un sistema que lleve a cabo la encriptación de datos.

— Préstamo interbibliotecario.

— Digitalización. Con el fin de suministrar de forma electrónica artículos publicados en papel y solicitados por los usuarios.

En el segundo caso (módulos extendidos) podemos encontrar:

— Servidores de metadatos. Si la biblioteca ha optado por separar los metadatos de los documentos, es necesaria la utilización de servidores que sean capaces de trabajar con grandes cantidades de metadatos.

— Gestión de datos jerárquicos. En numerosas ocasiones las bibliotecas digitales optan por editar la información siguiendo algún lenguaje de etiquetado (sgml, XML...). Estos sistemas permiten la generación de árboles temáticos que tienen como finalidad principal recuperar la información de forma más precisa.

— Portales. Si actualmente se utilizan como punto de partida para navegar por internet, en una biblioteca digital se pueden usar como puentes que permitan el acceso a una o varias colecciones.

Problemas actuales

Aspectos económicos

Las bibliotecas digitales ¿nacen o se hacen? Esta pregunta es importante, pues mucha gente considera este tipo de centros como los sustitutos del formato impreso y de las bibliotecas tal y como las conocemos hoy en día. Esta opinión es expresada no sólo por los

entusiastas de la distribución electrónica de la información, sino por las propias Administraciones que ven una oportunidad ideal para dejar de construir nuevos edificios y ahorrarse la adquisición de colecciones para cada nuevo centro. Aunque también es verdad que son el vehículo ideal para la transmisión del conocimiento en aquellos países que no pueden construir una adecuada red de bibliotecas por falta de recursos.

En el caso de que la biblioteca digital se haga, es decir, surja a partir de fondos impresos, los aspectos a tener en cuenta son tan amplios como costosos: selección de fuentes para la digitalización, necesidades hardware y software, costes de digitalización, infraestructura, requerimientos legales (copyright y derechos de autor), selección y suscripción de nuevos soportes electrónicos, investigación y desarrollo... Para los interesados en conocer de una forma más pormenorizada los aspectos económicos a tener en cuenta a la hora de convertir/transformar una biblioteca de fondos impresos en una digital, **Andreoni et al.** (1999), **Kapidakis** (1999) y **Lloret** (1999), hacen un estudio bastante detallado de esta situación.

«Gran parte de las bibliotecas digitales se han decantado por la opción de crear sus propios conjuntos de metadatos»

En el caso de que la biblioteca digital nazca, las cuestiones a considerar son muy parecidas a las expuestas en el párrafo anterior. Aunque es cierto que hay que dedicar un mayor presupuesto para la adquisición de fondos electrónicos, también es verdad que los costes derivados de la digitalización (compra o alquiler de equipos, gastos de personal, etc.) son casi nulos, por lo que la inversión será siempre menor. Para los interesados en obtener información más detallada acerca de esta materia, **Spink** (1995), **Arms** (1995) y **Hayes** (1999) hacen un estudio serio y profundo sobre los aspectos económicos de las bibliotecas digitales (costes directos, indirectos, operativos...), y su repercusión en otros asuntos como los sistemas y medios de distribución actuales.

Diseño metodológico

Las bibliotecas digitales, por su naturaleza y características, exigen dispositivos dinámicos de adaptación al siempre cambiante entorno operacional. Para la puesta en marcha de los sistemas de bibliotecas digitales se han propuesto varios modelos. Por ejemplo, tal y como se recoge en el sexto proyecto *DLI-I*, la *Universidad de Stanford* las considera como un bus de datos compartido que conecta varios depósitos digitales. Por otro lado, la *Universidad de Michigan* las ve como un conjunto de agentes cooperativos. El aspecto co-

mún de los distintos puntos de vista es que todas ellas se basan en una agrupación de servidores que utilizan internet, y que interactúan unas con otras para satisfacer las necesidades de sus usuarios.

http://www.cise.nsf.gov/iis/dli/_home.html

Personal digital library (PDL) es un conjunto de bibliotecas digitales implementadas dentro del proyecto *Elinor* y gestionadas por un mismo sistema de recuperación de información (**Ramsden et al.**, 1998). La particularidad de *PDL* se encuentra en que la ubicación del motor de búsqueda y de la interfaz es distinta e independiente del servidor en el que se hallan situadas cada una de las bases de datos, por lo que las consultas pueden ser atendidas por una o varias bibliotecas del sistema en función de la disponibilidad de objetos que se adecúen a ella.

Fedora es un software con arquitectura de depósito llevado a cabo por la *Universidad de Cornell* para aquellas bibliotecas digitales basadas en un sistema abierto que necesitan gestionar, e incluso integrar, objetos digitales de todo tipo (**Payette, Lagoze**, 1998). La característica principal de *Fedora* es que el funcionamiento de todos sus servicios se basa en interfaces:

Depósito documental que suministra los mecanismos adecuados para el almacenamiento y acceso a los objetos digitales.

Indización que facilita la recuperación de los objetos.

Conjunto de servicios que permiten el tratamiento e integración de los objetos digitales en una misma colección.

Interfaz de usuarios que actúa como pasarela hacia otros servicios.

Desarrollo de la colección

Es uno de los aspectos más importantes, y a menudo el que más problemas plantea en la biblioteca digital. Como consecuencia de ello son bastantes las publicaciones, la mayoría de ellas en revistas especializadas, que recogen investigaciones centradas en conseguir una política efectiva del desarrollo de la colección. Algunas sólo describen los procedimientos adoptados en proyectos concretos, mientras que otras se centran en los problemas relacionados con la evolución de la colección digital en general.

Weiner (1997) habla de la importancia de las revistas electrónicas haciendo hincapié en el precio, implementación, distribución, derechos de autor y formato desde la perspectiva de los bibliotecarios, editores y distribuidores. **Galbreath** (1997) debate sobre el papel que juegan los bibliotecarios en el desarrollo de la colección en las bibliotecas tradicionales, y de cómo és-

te está cambiando a causa de la aparición de las bibliotecas digitales, donde la mayoría de los materiales está disponible en formato electrónico.

Pandora, el proyecto de la *Biblioteca Nacional de Australia*, utiliza unas guías de selección particulares elaboradas por el *Selection committee on online Australian publications*. Este proyecto tiene en cuenta que la aplicación y gestión de las políticas de desarrollo de la colección puede consumir mucho tiempo y recursos a corto plazo. Por este motivo defiende que recopilar todo tipo de información digital no es sostenible, ni incluso deseable, a largo plazo.

Butler (1996) sugiere dialogar con autores y editores con el fin de llegar a un acuerdo que permita a las bibliotecas digitales convertir a formato electrónico parte de su colección bibliográfica para que pueda llegar a un mayor número de usuarios. Esta idea puede ser fundamental para la formación de las bibliotecas digitales, pues la conexión de muchas de ellas permitiría atender a un público muy amplio y variado. **Enderle** (1996) describe el desarrollo de la colección desde el punto de vista del proyecto de la *Biblioteca digital alemana*, que se basa en la compra y acceso compartido de los fondos digitales, la realización de servicios de información, estandarización del formato de los documentos y el límite entre el acceso gratuito para los socios y el de pago para determinados servicios o productos.

«La clasificación del conocimiento por medio de sistemas tradicionales como listas de términos controlados, sistemas de clasificación y catalogación, etc., está volviendo a tener un fuerte auge en los entornos digitales»

La necesidad de compartir recursos y fondos (documentales) queda patente en el proyecto de la *Ncstrl* (*Networked computer science technical reference library*). Se trata de una federación de más de cien bibliotecas públicas y privadas cuyo único objetivo es crear una biblioteca federal de informática. Para ello han implementado una arquitectura abierta de biblioteca digital. Esto significa que la funcionalidad de la biblioteca digital depende de un conjunto de servicios asociados a un protocolo de comunicaciones adoptado por todos, y de una interfaz que permite la puesta en marcha de los servicios. La evolución organizativa de cada biblioteca, así como el diseño y desarrollo de las futuras y necesarias implementaciones del sistema, dependerá fundamentalmente de la coherencia existente entre las interfaces y el protocolo adoptado.

<http://www.ncstrl.org>

El coste económico de la colección es un aspecto a tener muy en cuenta. Es necesario tomar decisiones, por ejemplo, sobre si se debe mantener la versión impresa de los recursos electrónicos cuando éstos están disponibles en la Red, o almacenados en un soporte óptico o magnético dentro del centro. Por otra parte, la adquisición compartida de recursos entre bibliotecas puede reducir los costes, pero cabe la posibilidad de que el consumo en telecomunicaciones (instalación de nuevas redes, uso y mantenimiento) dispare los gastos.

Interfaz de usuario

Se trata de uno de los temas que más ha atraído la atención de los investigadores en recuperación de información en la última década. Prueba de ello son los avances realizados tanto en el diseño de interfaces para el acceso en línea a la información, como para la recuperación de información de bases de datos en cd-rom. Las bibliotecas digitales se enfrentan a la necesidad de crear una interfaz única que sea capaz de gestionar los objetos independientemente de su origen, forma, formato e idioma utilizado por los usuarios. Hoy día son bastantes las bibliotecas que cuentan con una interfaz de usuario capaz de gestionar los objetos de sus depósitos documentales (en algunos casos distribuidos), pero ninguna puede ser aplicable a la totalidad de los proyectos existentes.

Clasificación e indización

Se han cumplido ya 123 años desde la aparición de la primera edición de la *Clasificación Decimal de Dewey* (*CDD*) y aún hoy a las puertas del siglo XXI tanto las bibliotecas tradicionales como las digitales siguen necesitando sistemas de catalogación, clasificación e indización. Desde la aparición de la *CDD* hasta hoy, se han realizado grandes esfuerzos con el objetivo de mejorar estos sistemas. La *University of Michigan School of Information* ofrece a sus usuarios una gran cantidad de objetos digitales organizados según la *CDD* por medio de un sistema parecido al *world wide web*, llamado *Internet public library*. *Scorpion* es un sistema de clasificación basado en la *CDD* y mantenido por la *Oclc*, capaz de generar y asociar de forma automática un código de clasificación a los distintos objetos. Cada nuevo objeto se enfrenta a la base de datos como si se tratase de una consulta, de tal forma que los resultados obtenidos pueden ser utilizados como posibles materias que representan a ese documento en concreto. *Scorpion* utiliza técnicas de clustering centradas en la medición de la similaridad entre los documentos, que se calcula en función de las palabras contenidas en cada uno.

<http://orc.rsch.oclc.org:6109>

Existen también otras herramientas desarrolladas única y exclusivamente con el fin de seguir utilizando los sistemas tradicionales de clasificación e indización, como por ejemplo: *Eels* (*The electronic engineering library*) o *The internet scout project*, con el que se realiza un sistema de clasificación tradicional basado en la lista de materias de la *Biblioteca del Congreso*.

<http://eels.lub.lu.se>

<http://scout.cs.wisc.edu/index.html>

Metadatos

Podemos definirlos como “datos sobre los datos que contienen los documentos”. El ejemplo más claro lo tenemos en el formato Marc. En el caso de las bibliotecas digitales nos estaríamos refiriendo a datos sobre los objetos almacenados. La distinción entre datos y metadatos es muy útil, sobre todo en el entorno digital, ya que mediante ellos es posible diferenciar no sólo los objetos almacenados, sino también sus distintas estructuras. Pero ¿cuáles son esos metadatos que permiten identificar a un objeto o documento? Con el objetivo claro de dar respuesta a esta pregunta surgieron varios sistemas de metadatos, entre los que destacó rápidamente *Dublin Core*.

<http://purl.org/dc>

Dublin Core es producto del consenso internacional entre un gran número de expertos sobre la elección de quince elementos que, con menos nivel de complejidad que otros sistemas más orientados hacia el mundo bibliotecario clásico (Marc), facilitan la descripción e identificación de cualquier fuente electrónica (ver tabla 3). Está claro que, para que un objeto pueda ser consultado y recuperado a partir de sus metadatos, es necesario que alguien los haya incluido. Esto ocurre tanto con los documentos digitalizados como con los que no se van a digitalizar nunca, pero de los que se quiere incluir referencias en la biblioteca digital.

Para determinadas ocasiones, los metadatos *Dublin Core* pueden ser personalizados con calificadores (que especifican el tesoro utilizado para la indización, lenguaje del texto, sistema de codificación del fichero...), aunque estas modificaciones afectarían a la interoperatividad de las distintas versiones.

Otros sistemas de metadatos que también han tenido una amplia aceptación han sido:

— *EAD* (*encoded archival description*). Se trata de una *DTD* (*document type definition*) para sgml con vistas a almacenar datos sobre documentos archivísticos.

<http://lcweb.loc.gov/ead>

— *TEI* (*text encoding initiative*). Otra *DTD*, pero en este caso orientada (en sus inicios) especialmente a

documentos sobre lingüística y literatura. En la actualidad permite la inclusión de datos siguiendo normas de carácter nacional como las *Aacr2* o las *Isbd*.

<http://www-tei.uic.edu/orgs/tei>

— *MCF* (*meta content framework*). Sistema desarrollado por la empresa *Netscape Corporation*.

<http://www.w3.org/TR/NOTE-MCF-XML/MCF-tutorial.html>

La mayoría de estos sistemas posee una estructura fija y, en ocasiones, poco flexible. Esto ha provocado que gran parte de las bibliotecas digitales se decantaran por la opción de crear sus propios conjuntos de metadatos. De esta forma se ajustaban mejor a los requerimientos de cada centro pero, sin embargo, era imposible la unión de metadatos de dos bibliotecas diferentes.

«En muchos centros con un espíritu práctico plausible han decidido digitalizar grandes cantidades de información a la espera de poder contar en el futuro con mejores herramientas de acceso»

Con el fin de acabar con este caos, el *W3C* (*World Wide Web Consortium*) estudió una norma que permitiera la creación de un marco genérico de descripción: *RDF* (*resource description framework*) sirve para intercambiar información entre varios sistemas de metadatos, facilitando la integración entre diversas colecciones.

Por otra parte hay que recordar que los metadatos pueden ubicarse en numerosos puntos del sistema de información:

- en la cabecera de cada documento,
- en un documento separado,
- en un depósito de metadatos.

El esquema seleccionado determinará las condiciones en las que se debe llevar a cabo el intercambio de datos entre cualquier sistema.

Recuperación de información

La recuperación de información (RI) ha sido uno de los campos de investigación más amplios en las últimas décadas. La aparición de internet, del *www* y ahora de las bibliotecas digitales, ha provocado un relanzamiento de la misma.

El acceso y recuperación de la información en el entorno de la Red y de las bibliotecas digitales puede

ser resuelto mediante pequeñas adaptaciones a los actuales sistemas de indización de los documentos a texto completo. No obstante hay una serie de problemas que deben ser tenidos muy en cuenta, pues pueden convertirse en un escollo importante para la RI:

— Los sistemas de recuperación de información (SRI) suelen conllevar un lenguaje de consulta y una serie de procesos cognitivos, por lo que es muy difícil su automatización de forma completa.

— Las necesidades de los usuarios varían con el tiempo, entorno..., por lo que los SRI deben poder adaptarse a estos cambios y evolucionar con los usuarios.

Para solucionar estos problemas se están realizando muchos y variados proyectos de investigación, que según **Chowdhury** (1999) son los siguientes:

— Se intenta conseguir un conjunto de metadatos relacionados con una interfaz común que permita a los usuarios finales conocer qué información está disponible y, si les interesa, poder recuperarla, independientemente del hardware y software utilizado y de las posibles variaciones que éste pueda experimentar.

— Los estudios de usuarios para el desarrollo e implementación de los SRI son cada día más importantes, sobre todo en el entorno de las bibliotecas digitales.

— La clasificación del conocimiento por medio de sistemas tradicionales como listas de términos controlados, sistemas de clasificación y catalogación, etc., está volviendo a tener un fuerte auge en los medios digitales.

— Se están diseñando agentes inteligentes para la búsqueda y recuperación de información que permitan mejorar el nivel de satisfacción de los usuarios en el entorno digital.

— Se está teniendo muy en cuenta la experiencia de los profesionales de la información en la búsqueda y recuperación, así como su capacidad de identificar las necesidades de los usuarios para organizar la estructura caótica de la Red.

«Después de siglos organizando la información, las herramientas para estructurarla todavía dejan mucho que desear»

Como ejemplo de estas investigaciones podemos citar las técnicas de visualización, que han sido desarrolladas con el fin de ofrecer nuevos métodos de consulta en multimedia y en las bibliotecas digitales: *The Big Picture* es un centro de intercambio de información de distintas actividades, en donde se aplican téc-

nicas de visualización de información para mejorar el acceso y uso de la misma. Otros ejemplos del mismo tipo son el de **Xia Lin**, de la *Universidad de Drexel*, el de **Hsinchun Chen**, de la *Universidad de Arizona*, o el *Websom* de **Teuvo Kohonen**, de la *Universidad Tecnológica de Helsinki*.

<http://ai.bpa.arizona.edu/et>

<http://lislin.gws.uky.edu/Sitemap/Sitemap.html>

<http://www.public.iastate.edu/~CYBERSTACKS/BigPic.htm>

<http://websom.hut.fi/websom>

Aspectos legales

De todos los problemas existentes, posiblemente sean éstos el mayor escollo para el desarrollo de las bibliotecas digitales. Pueden distinguirse los siguientes:

— Originalidad de la información: como consecuencia del desarrollo tecnológico, cualquier persona podría con gran facilidad utilizar las herramientas adecuadas para apropiarse de una información creada por otros y decir que es suya (imágenes, textos, sonidos...). Una de las técnicas propuestas para verificar y proteger la autoría de las obras consiste en la adopción de filigranas o marcas digitales (digital watermarks), que consisten en unos mapas de bits, para el caso de los gráficos, textos y vídeos, y de unos sonidos de fondo en el caso de las grabaciones de audio, imperceptibles por los sentidos humanos pero identificables por un software concreto. Desgraciadamente la localización y eliminación de este tipo de marcas en los documentos es tan fácil como su integración (**Vargas-Quesada**, 1998), por lo que la originalidad de la información aún no está asegurada.

— Derechos de autor: las bibliotecas digitales se encuentran con el enorme problema de tener que respetar la legislación vigente sobre la propiedad intelectual a la vez de que deben satisfacer las necesidades informativas de sus usuarios. Son conocidos los problemas que surgen respecto a los derechos de derechos de autor en el momento en que un usuario recibe una copia electrónica de un documento. A partir de ese instante la información podrá ser copiada, transformada y difundida cuantas veces se quiera, sin que el autor o el propietario reciban ningún tipo de beneficio por ello.

Existen algunos proyectos como *Acorn*¹, *DOI* (*digital object identifier*)². En España, *Cedro* (*Centro Español de Derechos Reprográficos*)³ garantiza los derechos de propiedad intelectual de los documentos, aunque de momento sólo se ocupa de los impresos.

Preservación

Hace ya tiempo que los formatos electrónicos y digitales se están viendo no sólo como unos medios ideales para la preservación de los formatos impresos tradicionales, sino también como sistemas que pueden facilitar el acceso y la difusión de la información de forma ilimitada. No obstante, existen algunos problemas de criterio y tecnológicos relacionados con la preservación de la información:

¿Deben las bibliotecas tradicionales digitalizar la totalidad de sus fondos con la inversión que esto supone? En el caso de realizar una selección de los fondos a digitalizar ¿qué criterio seguir? ¿Ha de llegarse a un acuerdo con el resto de bibliotecas de la comunidad o del país para que cada una digitalice unos documentos determinados y así garantizar la preservación de la colección? ¿Deberían realizar esta labor las Bibliotecas Nacionales de cada país?

En el caso de las revistas electrónicas hay que destacar la iniciativa *ECO* (*Electronic collection online*). Desde 1997 este servicio de *Oclc* sirve de puente entre las editoriales y las bibliotecas. Las primeras entregan a *Oclc* el texto de sus revistas electrónicas, y es este consorcio internacional el encargado de mantener el software de recuperación, gestionar los archivos y almacenarlos de forma pertinente. Los usuarios de las bibliotecas pueden consultar los textos almacenados en el servidor de *Oclc*, garantizando de esta forma el acceso continuado en el caso de una cancelación de suscripción o desaparición de la revista o de la editorial. Si bien es cierto que el sistema no está orientado a cubrir totalmente las necesidades de una biblioteca digital, deja constancia del hecho de que a través de la coordinación y la cooperación internacional es posible solucionar buena parte del problema de la preservación de la colección.

<http://www.oclc.org/oclc/menu/eco.htm>

Otra cuestión: ¿qué soporte y formato utilizar para almacenar la información digitalizada? A mediados de los 90 el formato que reinaba era PDF. Más adelante, con el avance de internet, se redescubrieron los lenguajes de etiquetado de la mano de HTML. El tiempo puso en evidencia las limitaciones que ofrecía este lenguaje y apareció XML...

Situación actual

En la *Tercera conferencia europea sobre bibliotecas digitales (Ecdl'99)* celebrada en París se pudieron observar los últimos avances tecnológicos relacionados con este tipo de bibliotecas. Es llamativo que cada institución o grupo de instituciones, cada país o comunidad de países, estén afrontando el problema desde un mismo punto de vista pero con diferentes soluciones.

Quizá cuando en un futuro todos los problemas relacionados con las bibliotecas digitales queden solucionados, las medidas utilizadas para resolver una cuestión concreta difícilmente serán extrapolables a otras.

Después de siglos organizando la información, las herramientas para estructurarla todavía dejan mucho que desear. Éste es realmente el reto de nuestra profesión: organizar el caos de datos, o por lo menos hacerlos fácilmente recuperables. Mientras tanto, en muchos centros con un espíritu práctico plausible han decidido digitalizar grandes cantidades de información a la espera de poder contar en el futuro con mejores herramientas de acceso.

Digitalizar una colección no es sólo un trabajo informático. Es necesario contar con profesionales de la organización del conocimiento, ya sean éstos bibliotecarios, documentalistas o como prefiera denominarseles. Insistimos: la parte fundamental no es sólo aumentar el tamaño del banco de datos, sino que es básico el trabajo de descripción (metainformación) de las fuentes almacenadas.

Conclusiones

Resulta muy difícil establecer conclusiones definitivas sobre un área emergente como la que aquí nos ocupa, pues se encuentra sometida a un constante cambio debido a las múltiples líneas de investigación y desarrollo abiertas. No obstante se vislumbra una serie de aspectos que merece la pena comentar.

La definición de biblioteca digital como punto de encuentro entre profesionales de la información, informáticos, ingenieros electrónicos, sociólogos, antropólogos..., para una mejor organización y distribución de los sistemas de información, parece todavía un objetivo a medio plazo. Fundamentalmente porque se basa en la implementación de nuevas tecnologías que a su vez implican la apertura de nuevas líneas de investigación sobre otras en estado de gestación.

La mayor parte de las investigaciones relacionadas con las bibliotecas digitales se basan en la producción y mejora del hardware, redes de comunicaciones, interfaces de usuario, herramientas y técnicas relacionadas con el almacenamiento y la recuperación de la información..., pero a veces se llevan a cabo de una forma anárquica y poco cooperativa, como si los muchos proyectos de bibliotecas digitales existentes no tuviesen nada que ver unos con otros.

Es necesario que los responsables de cada uno de estos proyectos se reúnan y sigan adoptando acuerdos sobre estándares: protocolos (por ejemplo z39.50), elaboración común de una interfaz única de usuario, etc. Si esto no se hace así, se corre el peligro de que, una vez que estas bibliotecas estén en funcionamien-

to, sean zonas aisladas de conocimiento digital, y que el proyecto de construcción de la biblioteca digital mundial, cuyo objetivo es la preservación y difusión del conocimiento, independientemente de la localización de sus usuarios, se retrase por falta de cooperación y metodología en el trabajo.

Notas

1. *Acorn* es un proyecto de *Elib* (*Electronic libraries programme*) en el que está implicada *Swets & Zeitlinger BV* con una serie de bibliotecas universitarias del Reino Unido. El objetivo principal de *Acorn* es obtener información de primera mano sobre cómo se gestionan los permisos de reproducción, el suministro de copias electrónicas y el correspondiente pago de los derechos de autor en los distintos tipos de bibliotecas con el fin de desarrollar un modelo de transferencia y gestión de derechos para los artículos de revistas electrónicas.

2. *DOI* fue desarrollado por la *Association of American Publishers (AAP)* en colaboración con la *Corporation for National Research Initiative (Cnri)*. En la actualidad es gestionado por la *DOI Foundation*. El sistema propone utilizar un directorio que asigna un número a cada publicación. Una vez que se ha seleccionado un ítem, el usuario recibe información sobre los derechos de autor a los que están sujetos esos datos.

El uso de este tipo de sistemas implica, por un lado, aplicar el concepto de *DLO* (*document like object*) para todos los materiales de la biblioteca digital. Por otro, es necesaria la creación de delegaciones —a imitación de las agencias que gestionan el *Isbn*— que se encarguen de establecer las relaciones entre el ítem y su *DOI* correspondiente, tal y como se explica en el trabajo de *Rosa* (1999). Según nuestra opinión, este último requisito pone en evidencia la duplicidad de esfuerzos que se están produciendo dentro de este campo.

3. Organismo encargado de controlar la recaudación de los derechos de autor generados por la reprografía autorizada en España. Fundado en 1988, no comenzó a realizar su actividad plenamente hasta dos años más tarde. *Cedro* sólo tiene en cuenta trabajos impresos y reproducción por fotocopia. En ningún caso contempla la digitalización de revistas, la copia de publicaciones electrónicas y un sinnúmero de casos que se producen constantemente en cualquier biblioteca digital.

Referencias

Andreoni, Antonella; et al. "Developing an European technical reference digital library". En: *Research and advanced technology for digital libraries*, 1999, pp. 343-362.

Arms, William Y. *Key concepts in the architecture of the digital library*. Consultado en: 06/03/2000.
<http://www.dlib.org/dlib/july95/07arms.html>

Arms, William Y.; Bianchi, Christophe; Overly, Edward A. An architecture for information in digital libraries. Consultado: 08/02/2000.
<http://www.dlib.org/dlib/february97/cnri02arms1.html>

Borgman, C. L.; Bates, M. J.; Cloonan, M. V.; Efthimiadis, E. N.; Guillianland-Swetland, A; Kafai, Y; Leazer, G. L.; Madox, A. *Social aspects of digital libraries*. Consultado en: 15-10-99.
<http://dli.grainger.uiuc.edu/national.htm>

Borgman, C. L. "What are digital libraries? Competing visions". En: *Information processing and management*, 1999, v. 35, pp. 227-243.

Butler, B. "Electronic course reserves and digital libraries: progenitor and prognosis". En: *Journal of academic librarianship*, 1996, v. 22, n. 2, pp. 124-127.

Chowdhury, G. G. "Internet and information retrieval". En: *Journal of Documentation*, 1999, v. 55, n. 2, pp. 209-225

Chowdhury, G. G.; Chowdhury, Sudatta. "Digital library research: major issues and trends". En: *Journal of documentation*, 1999, v. 55, n. 4, pp. 409-448.

Cronin, Blaise. "Information professionals in the digital age". En: *International information and library review*, 1998, v. 30, n. 1, pp. 37-50.

Enderle, W. "Shared digital collection development: the first steps to the virtual library". En: *European research libraries co-operation*, 1996, v.6, n. 4, pp. 367-382.

Faba Pérez, Cristina; Moya Anegón, Félix de. "Bibliotecas digitales: concepto y principales proyectos". En: *Investigación bibliotecológica*, 1999, enero-junio, v. 13, n. 26, pp. 64-78.

Galbreath, R. "Nailing Jell-O to the wall? Collection management in the electronic era". En: *North Carolina Libraries*, 1997, spring, v. 55, n. 1, pp. 18-21.

Gladney, H. M., Fox, E. A., Ahmed, Z., Ashany, R., Belkin, N. J. y Zemankova, M. Digital library: gross structure and requirements: report from a March 1994 Workshop. Consultado en: 07-12-99.
<http://www.csdl.tamu.edu/DLP4/paper/fox.html>

Hayes, Robert M. *The economics of digital libraries*. Consultado en: 28-06-2000.
<http://www.usp.br/sibileconomics.html>

Kapidakis, Sarantos. "Issues in the development and operation of a digital library". En: *Research and advanced technology for digital libraries*, 1999, pp. 363-382.

Keefer, Alice. "La preservació digital". En: *Item*, 1999, jul.-dic., nº 25, pp. 27-41.

Lynch, Clifford; García-Molina, Héctor. "Interoperability, scaling, and the digital libraries research agenda". En: *Iita digital libraries workshop*, 1995, pp. 18-19. Consultado en: 08/02/2000.
<http://www.diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html>

Lloret Romero, Nuria. "Cómo plantear un proyecto para el desarrollo de una biblioteca digital". En: *II Jornadas andaluzas de documentación*, 1999, pp. 307-317.

Miksa, F. L.; Doty, P. *Intellectual realities and the digital library*. Consultado en: 12-10-99.
<http://csdl.tamu.edu/DL94/paper/miksa.html>

Murray, Robin. "The digital library jigsaw: fitting the pieces together". En: *Online information proceedings*, 1999, diciembre, pp. 175-180.

Oppenheim, Charles; Smithson, Daniel. "What is the hybrid library?". En: *Journal of information science*, 1999, febrero, v. 25, pp. 97-112.

Payette, Sandra; Lagoze, Carl. *Flexible and extensible digital object and repository architecture (Fedora)*. Consultado en: 04/03/2000.
<http://www2.cs.cornell.edu/payette/papers/ECDL98/FEDORA.html>

Ramsden, A.; et al. *Elinor: electronic library project*. London: Bowker-Saur for the British Library, 1998.

Rosa, Antonio de la. "XML orientado a objetos". En: *El profesional de la información*, 1999, septiembre, v. 8, n. 9, pp. 4-23.

Rusbridge, Chris. *Toward the hybrid library*. Consultado en: 03-03-2000.
<http://www.dlib.org/dlib/july98/rusbridge/07rusbridge.html>

Spink, Amanda. *Digital libraries and sustainable development?* Consultado en: 06/03/2000.
<http://csdl.tamu.edu/DL95/papers/spink/spink.html>

Vargas-Quesada, Benjamín. "Marcas digitales: ¿sistema de copyright en internet?". En: *El profesional de la información*, 1998, v. 7, n. 12, pp. 31-32

Waters, D. J. What are digital libraries? Consultado en: 12-10-99.
<http://www.clir.org/pubs/issues/issues04.html>

Weiner, S. T. "Electronic journal, four parts series: an introduction". En: *Science and technology libraries*, 1997, v. 16, pp. 65-68.

Pedro Hípola. *Universidad de Granada, Facultad de Biblioteconomía y Documentación.*
phipola@ugr.es

Benjamín Vargas-Quesada. *Universidad de Granada. Centro de Documentación Europea.*
bvargas@ugr.es

José A. Senso. *Universidad de Granada, Facultad de Biblioteconomía y Documentación.*
jsenso@ugr.es