

**Informe Técnico – Technical Report**

**DPTOIA-IT-2007-001**

**Junio, 2007**

# **Las bibliotecas digitales y su aplicación en la gestión de contenidos**

**Sergio Bravo Martín**

**Francisco José García Peñalvo**



Departamento de Informática y Automática

Universidad de Salamanca

Revisado por:

Dr. D. Carlos García de Figuerola Paniagua

Área de Lenguajes y Sistemas Informáticos

Departamento de Informática y Automática

Facultad de Traducción y Documentación – Universidad de Salamanca

[figue@usal.es](mailto:figue@usal.es)

Dr. D. José Luis Alonso Berrocal

Área de Lenguajes y Sistemas Informáticos

Departamento de Informática y Automática

Facultad de Traducción y Documentación – Universidad de Salamanca

[berrocal@usal.es](mailto:berrocal@usal.es)

Aprobado en el Consejo de Departamento de 7 de junio de 2007

Información de los autores:

D. Sergio Bravo Martín

Área de Lenguajes y Sistemas Informáticos

Departamento de Informática y Automática

Facultad de Ciencias – Universidad de Salamanca

Plaza de los Caídos S/N – 37008 – Salamanca

[ser@usal.es](mailto:ser@usal.es)

Dr. D. Francisco José García Peñalvo

Área de Ciencia de la Computación e Inteligencia Artificial

Departamento de Informática y Automática

Facultad de Ciencias – Universidad de Salamanca

Plaza de los Caídos S/N – 37008 – Salamanca

[fgarcia@usal.es](mailto:fgarcia@usal.es)

Este trabajo ha sido parcialmente financiado por el Programa Nacional de Tecnologías de Servicios de la Sociedad de la Información (proyecto ref. TSI2005-00960).

Este documento puede ser libremente distribuido.

© 2007 Departamento de Informática y Automática - Universidad de Salamanca.

## Resumen

Este documento muestra el estado actual de las bibliotecas digitales. Además hace un recorrido a través de las diferentes definiciones enunciadas en estos últimos años. Dado que los campos de actuación así como la naturaleza de las bibliotecas digitales son muy variados, se proponen varias clasificaciones. También se tratan aspectos clave sobre la construcción y el diseño de bibliotecas digitales. En este sentido, se exponen algunos de los conceptos más interesantes de su aplicación en los sistemas distribuidos. Sin embargo, la construcción de bibliotecas digitales no es algo trivial, por esto mismo, se exponen algunas de las problemáticas más importantes que emergen al respecto.

Uno de los campos de actuación más relevantes de las bibliotecas digitales es el universo Internet y, más concretamente, la gestión de contenidos web. Por este motivo se incluye todo un capítulo dedicado a los denominados Sistemas de Gestión de Contenidos (CMS). No cabe ninguna duda, que los servicios de información y documentación disponibles en Internet aumentan y evolucionan continuamente. El usuario poco a poco, se vuelve más exigente y demanda información correcta, actualizada y accesible en todos los sentidos. Estos y otros factores implican un incremento del esfuerzo empleado en la gestión y mantenimiento de la información, y en particular, de la información contenida en la Web. En consecuencia, los procesos de publicación y creación de contenido, han ido cambiando e incorporando nuevas técnicas y procedimientos relacionados con la gestión de información. En esta evolución, los sistemas de gestión documental y de recuperación de información se funden para adoptar soluciones globales que soporten el proceso de gestión de información de cualquier organización. En este marco de trabajo emergen los sistemas de gestión de contenidos. Uno de los puntos fuertes de este capítulo ha sido la construcción de un producto software que muestra las funcionalidades básicas de un sistema gestión de contenidos web. Bajo un enfoque arquitectónico común al de las bibliotecas digitales, el software se presenta con una interfaz amigable, sencilla y accesible. El producto lleva por nombre *MyCMS (My Content Management System)* y pretende ser una herramienta útil para la difusión de información en entornos de colaboración y especialmente en el área de la educación. Finalmente el artículo expone las experiencias obtenidas de la implantación de MyCMS en un centro de enseñanza secundaria, como parte de un intento de acercar las TIC y la tele-formación a este sector de la educación.

## **Abstract**

This document shows the current state of the digital libraries. Besides, it does a tour across the different definitions enunciated in the latter years. Provided that the fields of performance as well as the nature of the digital libraries are very varied, it proposes several classifications. Also it treats other fix aspects on the construction and design of digital libraries. In this respect, there are exposed some of the most interesting concepts of the application in the distributed systems. Nevertheless, the construction of digital libraries is not anything trivial, for this reason, some of the more important troubles are exposed.

One of the more important performance fields of the digital libraries is the universe of Internet and, more specifically, the web content management. For this reason, there is included a whole chapter dedicated to systems called Content Management Systems (CMS). There is not any doubt, which the available services of information and documentation on Internet increase and evolve constantly. The user gradually becomes more perfectionist and demands correct, updated and accessible information in every sense. These and other factors imply an increase of the effort used in the management and maintenance of the information, and especially, of the information contained in the Web. In consequence, the processes of publication and creation of content have been changing and incorporating new technologies and procedures related to the information management. In this evolution, the systems of documental management and information retrieval are mixed to adopt global solutions that support the content management of any organization. In this framework the content management systems emerges. One of the strong points of this chapter has been the construction of a software product that shows the basic functionalities of a web content management system. Under a common architectural approach to that of the digital libraries, the software appears with one useful, simple and accessible interface. The product is named MyCMS (My Content Management System) and tries to be an useful tool for the diffusion of information in collaboration environments and specially in the area of education. Finally the article exposes the experiences obtained of MyCMS's implantation in a secondary education school, as part of an attempt of bringing the TIC and virtual formation over this sector of the education.

## Tabla de Contenidos

<i>Las bibliotecas digitales y su aplicación en la gestión de contenidos</i> .....	1
<b>1 Las bibliotecas digitales</b> .....	1
<b>1.1 Introducción</b> .....	1
<b>1.2 Evolución</b> .....	2
<b>1.3 Concepto</b> .....	3
<b>1.4 Clasificación</b> .....	6
<b>1.5 Problemática</b> .....	7
<b>1.6 Construcción de una biblioteca digital</b> .....	9
1.6.1 Digitalización y conversión a formato digital .....	9
1.6.2 Almacenamiento.....	10
1.6.3 Recuperación .....	10
1.6.4 Metadatos .....	11
1.6.5 Identificación.....	13
1.6.6 Software de base.....	15
<b>1.7 Componentes</b> .....	16
<b>1.8 Arquitectura OAI-PMH</b> .....	17
1.8.1 Proveedor de datos .....	19
1.8.2 Proveedor de servicios .....	20
1.8.3 Proveedores de datos y servicios.....	22
<b>2 Los sistemas de gestión de contenidos</b> .....	23
<b>2.1 Conceptos previos</b> .....	23
2.1.1 Consideraciones generales .....	23
2.1.2 Definición.....	23
2.1.3 Beneficios.....	24
2.1.4 Clasificación.....	24
2.1.5 Objetivos del CMS .....	25
2.1.6 Aplicación en el e-Learning .....	26
<b>2.2 Descripción funcional de un CMS</b> .....	26
2.2.1 Creación de contenido.....	27

2.2.2	Gestión de contenido.....	27
2.2.3	Publicación.....	27
2.2.4	Presentación .....	28
<b>3</b>	<b><i>MyCMS v1.0: una experiencia en la ESO.....</i></b>	<b>28</b>
<b>3.1</b>	<b>Conceptos previos.....</b>	<b>28</b>
3.1.1	Contenido .....	29
3.1.2	Comunicación.....	30
3.1.3	Actividad .....	30
<b>3.2</b>	<b>Características de MyCMS v1.0.....</b>	<b>30</b>
3.2.1	Características generales del sistema .....	30
3.2.2	Características de la publicación de contenidos .....	32
3.2.3	Arquitectura de la información.....	33
3.2.4	Estructura de navegación: el objeto <i>carpeta</i> .....	33
<b>3.3</b>	<b>Arquitectura de MyCMS.....</b>	<b>35</b>
3.3.1	Arquitectura multi-capa distribuida.....	35
3.3.2	Patrones de diseño .....	37
<b>3.4</b>	<b>Presente y futuro .....</b>	<b>38</b>
<b>4</b>	<b><i>Referencias</i> .....</b>	<b>40</b>

# 1 Las bibliotecas digitales

## 1.1 Introducción

Son diversos factores los que han incitado a investigadores, especialistas, bibliotecarios, técnicos de la computación y de las telecomunicaciones, a elaborar un mecanismo válido para la representación en formato telemático de los sistemas y elementos de información instaurados en las bibliotecas tradicionales, con el propósito de satisfacer las necesidades de información de los usuarios. Estos factores están relacionados fundamentalmente con necesidades prácticas, como son por ejemplo, el almacenamiento de grandes volúmenes de información, la diversidad de formatos, así como las dificultades añadidas para su recuperación. De acuerdo con esta perspectiva, se han llevado a cabo múltiples iniciativas, se han desarrollado numerosos productos y servicios digitales, se ha impulsado la investigación en tecnologías telemáticas aplicadas y se han lanzado numerosos proyectos de bibliotecas digitales, entre los que se pueden destacar los proyectos *Greenstone* y *Fedora*.

La creación de un repositorio documental para la biblioteca digital es una difícil tarea y siempre expuesta a errores, sin embargo, simplemente es parte de una de las cinco funcionalidades básicas de toda biblioteca digital: gestión de la información, presentación de la información, intercomunicación, operatividad y protección de la información. Es en cada una de estas materias en las que se centran prácticamente la gran mayoría de las líneas de investigación actuales del campo de las bibliotecas digitales.

Sobrepasamos ya quince años de duro trabajo en el que se han planteado numerosos problemas, algunos de los cuales no han sido resueltos aún. Una de las cuestiones más importantes en la creación de bibliotecas digitales es su concepción basada en repositorios dispersos que trabajan de forma conjunta e incluso con la posibilidad de ofrecer su información a otras organizaciones externas, es decir, interoperabilidad. Éste ha sido un problema crítico desde los años noventa y sigue estando a la orden del día, en la medida que el número de sistemas computacionales, repositorios de información, aplicaciones y usuarios se multiplican a ratios elevadísimas. Así pues, las bibliotecas digitales ya no sólo se concebirán como componentes individuales diseñados para operar de forma autónoma, sino también, como elementos que pueden operar de forma conjunta. En definitiva, se trata de un mecanismo de trabajo mucho más cercano al de los sistemas distribuidos. La interoperabilidad ha sido tratada ya anteriormente, pero principalmente en el contexto de un dominio específico como el de los sistemas de base de datos o incluso el de lenguajes de programación. Sin embargo, dada la naturaleza heterogénea de los recursos y servicios proporcionados por las bibliotecas digitales, es necesario afrontar la interoperabilidad desde un punto de vista más amplio (Paepcke A. et al., 1998). En esta línea surgen estándares de comunicación y nuevas iniciativas para dar finalmente a este problema una respuesta efectiva.

Por otro lado, y siempre sin olvidar el valor intrínseco de las bibliotecas digitales como auténticos repositorios de información digital, surgen otros ámbitos de aplicación bajo un marco de trabajo más acotado. En este sentido se halla la creación y administración de información de un sitio *web*, un marco de trabajo que dado su gran interés, adquiere un nombre propio: Sistemas de Gestión de Contenidos (*Content Management System*). En estos sistemas, una sola interfaz controla uno o varios repositorios de información en los que se aloja el contenido de un sitio *web*. El sistema permite manejar de manera independiente el contenido de una parte y el diseño por otra. Así, es posible manejar el contenido y darle en cualquier momento una apariencia distinta al sitio sin tener que darle formato al contenido de nuevo, además de permitir la fácil y controlada publicación a varios editores. En este sentido, los gestores de contenidos ofrecen un valor añadido a los repositorios de información digital, ofreciendo al usuario, diversas formas de presentarle la información almacenada. Como se puede observar, se potencia

sobremañera una de las funcionalidades principales en el diseño de bibliotecas digitales, la presentación de la información.

Aunque no forme parte específica de esta investigación cabe destacar también la importancia de la gestión documental en lo que a repositorios de información digital se refiere. Este sistema permite la gestión y el acceso unificado a información generada en la organización sobre muy diversos aspectos. La implantación de un sistema de gestión documental es un elemento clave en la actividad de una empresa, tanto en grandes organizaciones como en la pequeña y mediana empresa. En este sentido el alcance de la solución es muy amplio pasando desde el simple archivo masivo de documentos hasta el análisis de la información y gestión del conocimiento. En este caso la idea de compartir la información con otras organizaciones prácticamente desaparece, aunque cabe destacar un nuevo enfoque en el tratamiento de la información y que define un ciclo de vida completo de los documentos<sup>1</sup>. En este caso, la funcionalidad básica en el diseño de bibliotecas digitales que prima, es la de gestión de la información.

No cabe duda, que tanto la gestión documental, como los gestores de contenidos y las bibliotecas digitales, son tres sistemas de información estrechamente relacionados con una arquitectura común, pero que potencian unas funcionalidades más que otras.

## 1.2 Evolución

En el mundo de las bibliotecas y de sus proyectos digitales, la década de los años 80 es clave, debido a la introducción masiva de las computadoras personales en el trabajo diario de bibliotecarios y especialistas en información. Además el acceso en línea a los catálogos de las bibliotecas progresa con el empleo de diversas tecnologías.

En 1988, la Norma Z39.50, desarrollada por la *National Standard Organization*, (NISO) ([www.niso.org](http://www.niso.org)) y adoptada por el *American Standard Institute* (ANSI) ([www.ansi.org](http://www.ansi.org)), ordena el acceso en línea a los catálogos. Se trata de la primera especificación que permite la conexión a bases de datos con independencia del sistema que las almacene y de las plataformas de trabajo existentes.

Posteriormente toman protagonismo los protocolos TCP/IP de la mano de Internet. Se pueden destacar: el HTTP (*Hypertext Transfer Protocol*), SMTP (*Simple Mail Transfer Protocol*), FTP (*File Transfer Protocol*) y MIME (*Multipurpose Internet Mail Extension*). Todos ellos han contribuido notablemente a la expansión de las bibliotecas digitales en Internet, entre otras razones por las garantías para el intercambio de cualquier tipo de documento digital o electrónico a través de un entorno distribuido y colaborativo como es la red de redes.

Una de las partes fundamentales en el diseño de bibliotecas digitales dentro de su concepción como sistemas distribuidos implica también la computación remota. En este sentido, CORBA y DCOM son protocolos que proporcionan la portabilidad de componentes de diferentes lenguajes entre plataformas totalmente distintas. Estos protocolos garantizan la interoperabilidad y a su vez facilitan mecanismos para la invocación de operaciones remotas en cualquiera de los nodos que forman parte del sistema distribuido, y por tanto de la biblioteca digital.

Paralelamente, se investiga en la identificación y definición de un conjunto de elementos, lo más simple posible, para describir cualquier recurso de información. Este conjunto de elementos constituyen una meta-información crucial para la catalogación y recuperación posterior de los recursos recopilados. Aquí encontramos estándares como USMARC (MARC 21, MARC-XML) o más recientemente DUBLIN CORE.

---

<sup>1</sup> Del inglés *workflow*, flujos de trabajo para el tratamiento en este caso de la información.



El tema de las bibliotecas digitales es actualmente uno de los puntos de giro en cuanto al cambio de paradigmas en la Ciencia de la Información y de la Bibliotecología, en tanto que supone el acceso universal a la información. Esto engloba varios aspectos:

- La biblioteca digital de nuevo no existe como una institución aislada, sino en cooperación con bibliotecas o instituciones afines, para que el intercambio de información se convierta en algo fluido.
- La tecnología, como elemento indispensable en el que se sustenta la evolución de la biblioteca tradicional.
- La mentalidad de los profesionales y usuarios de las bibliotecas debe progresar aún más y orientarse hacia las nuevas metodologías y procedimientos de la sociedad de la información, e incorporarse a la nueva faceta digital de las bibliotecas.
- A raíz del nacimiento de las bibliotecas digitales, conceptos como: "usuarios", "colecciones" y "demandas de información", adoptan connotaciones nuevas para su análisis y rompen los límites, las barreras empíricas y pragmáticas establecidas por décadas, para cambiar radicalmente a la manera en que socialmente se asumen.
- La biblioteca digital es interdisciplinaria, no pertenece sólo a la labor de los bibliotecarios aunque estos deben tener un lugar determinante en su diseño, construcción y funcionamiento, y no se debe desechar la labor de siglos de la biblioteca tradicional, así como todo el legado conceptual de la Ciencia de la Información y la Bibliotecología.

Ocupa así, la biblioteca digital, un lugar importante dentro del universo de recursos electrónicos presentes en Internet; que ofrece al usuario la perspectiva del acceso electrónico a los recursos de información, al conocimiento acumulado y a la inteligencia colectiva, desde su conveniencia temporal y espacial.

### 1.3 Concepto

Antes de profundizar en las bibliotecas digitales, definimos el concepto de biblioteca en general y de biblioteca digital en particular.

La definición de biblioteca, según Taly Sharon & Ariel J. Frank (Sharon T, Frank A. J., 2000), incluye seis características principales:

1. Colección de objetos de datos: Una biblioteca aloja una colección de objetos de datos, también llamados fondos, ítems, recursos o materiales. Los ítems pueden ser: libros y periódicos, documentos (p. ej., páginas HTML) y objetos multimedia (pinturas o imágenes, cintas o archivos de vídeo, etc.). Los objetos bibliotecarios pueden ser accesibles localmente en la biblioteca o, indirectamente, utilizando una red para acceder a ellos.
2. Colección de estructuras de metadatos: Una biblioteca contiene una colección de estructuras de metadatos, como son catálogos, guías, diccionarios, *thesauros*, índices, sumarios, anotaciones, glosarios, etc.
3. Colección de servicios: Una biblioteca proporciona un conjunto de servicios, como: diversos métodos de acceso para distintos usuarios (búsqueda, etc.), gestión de la biblioteca, estadísticas y evaluación de las actividades de la biblioteca y Difusión Selectiva de la Información (DSI).
4. Dominio especializado: Una biblioteca tiene un dominio especializado y su colección tiene un objetivo. Por ejemplo: arte, ciencia o literatura. Además, normalmente se crea

para servir a una comunidad de usuarios y está sutilmente estructurada. Por ejemplo: universitaria, pública, especializada, escolar, nacional o estatal.

5. Control de calidad: Una biblioteca utiliza el control de calidad en el sentido de que todo su material se verifica y es consistente con el perfil o estereotipo de la biblioteca. El material se filtra antes de incluirlo en la biblioteca y sus metadatos generalmente se enriquecen (p. ej., se anota), etc.
6. Preservación: Bibliotecas y archivos han servido como focos institucionales centrales para la preservación y para ambas instituciones la preservación es una de sus funciones fundamentales. El objetivo de la preservación (Hedstrom M., 1998) es asegurar la protección de la información de valor permanente de manera que pueda ser accesible para las generaciones presentes y futuras. La preservación incluye la distribución de los recursos para su permanencia, medidas preventivas para detener el deterioro de los materiales y medidas restauradoras que permitan la utilización de los materiales seleccionados.

Durante los primeros años de la década de los 90 son muchas las definiciones que se han enunciado. En 1993, Edward A. Fox (1993, p.65), presenta a la biblioteca digital como “un entorno distribuido de tecnología que reduce las barreras para la creación, diseminación, manipulación, almacenamiento, integración y reutilización de la información por individuos y grupos”.

Un informe publicado por Saffady (Saffady W., 1995), cita del orden de treinta definiciones de biblioteca digital de entre las que se citan términos como: “grandes depósitos de información”, “bases de datos heterogéneas”, “productos informativos en línea”, “depósitos de información activa”, “mecanismos de almacenamiento computerizados en los que residen depósitos de información”, “bases de datos accesibles por medio de Internet”, entre otros muchos.

Realizando una puesta en común de, podemos observar:

- Un gran énfasis en la tecnología.
- Un enfoque totalmente dirigido hacia los recursos de información (en lugar de orientar el enfoque hacia los servicios al usuario).
- Un tratamiento pobre hacia la labor de los bibliotecarios y demás personal adscrito a la biblioteca.

Por otro lado, en el sitio web del Proyecto de Biblioteca Digital (*Digital Library Project*, <http://www.dli2.nsf.gov>), aparece una definición de biblioteca digital, que hace referencia también a un entorno distribuido, pero que completa algunas de las carencias mencionadas anteriormente: “El concepto de biblioteca digital no es únicamente el equivalente de repertorios digitalizados con métodos de gestión de documental. Es más bien un entorno donde se reúnen colecciones, servicios y personal que favorece el ciclo completo de la creación, difusión, uso y preservación de los datos para la información el conocimiento”.

Más adelante, otro informe de Edward A. Fox (Edward A. Fox, 1990-2001), recopila otra serie de definiciones enunciadas entre 1994 y 2001. En 1994 Bill Birmingham, por ejemplo, define a la biblioteca digital como “el nombre genérico para unas estructuras federadas que suministran a usuarios el acceso físico e intelectual a redes mundiales inmensas con la información codificada en formatos digitales multimedia”. El concepto de estructuras federadas aparece bastante frecuente a lo largo de estas definiciones e implica una estrecha relación e interconexión entre bibliotecas dispersas por la red y que trabajan de una forma colaborativa.

Más adelante, hacia 1996, se habla de las bibliotecas digitales como un conjunto electrónico de recursos y capacidades técnicas asociadas para la creación, búsqueda y

utilización de información. En este sentido son una extensión y realce de sistemas de almacenamiento y recuperación de información que manipulan los datos digitales en cualquier medio (texto, imágenes, sonidos, vídeo) y existen en redes distribuidas. El contenido de bibliotecas digitales incluye los datos, metadatos que describen los varios aspectos de los datos (descripción, creador, propietario, derechos de reproducción), y metadatos que constan de enlaces a otros datos o metadatos (internos o externos a la biblioteca digital).

Otro investigadores, como Gary Marchionini en 1998, argumentan que el concepto de biblioteca digital varía dependiendo del ojo con que se mire. Así pues, comunidades como la de los ingenieros de ciencias de la computación definen a la biblioteca digital como una metamorfosis de diferentes clases de servicios de bases de datos distribuidas para la gestión de datos multimedia no estructurados; otras comunidades del ámbito de la política y de los negocios, afirman que representa un nuevo mercado para los servicios y recursos de información del mundo; finalmente comunidades futuristas ven las bibliotecas digitales como la manifestación del conocimiento mundial<sup>1</sup>.

También nos encontramos asociaciones como la ASIS (*American Society for Information Science*) que se atreve a dar una definición muy directa, pero hasta cierto punto escasa: “bibliotecas cuyos contenidos se encuentran en formato electrónico y que son accesible local o remotamente mediante redes”.

Por otro lado, nos encontramos otras definiciones orientadas al proceso, como la enunciada por el consorcio DLF (*Digital Library Foundation*) en la que se define a las bibliotecas digitales como “organizaciones que proporcionan recursos digitales y para ello cuentan con el personal especializado para seleccionar, estructurar, ofrecer el acceso, interpretar, distribuir, preservar la integridad y asegurar la preservación de las colecciones digitales de manera que estén disponibles de forma fácil y económica para el uso de una comunidad definida o de un conjunto de comunidades”.

A su vez, Tim Bernes-Lee<sup>2</sup>, define a las bibliotecas digitales como una extensión de la WWW (*World Wide Web*), un entorno en el que se ofrece la información con un significado bien definido y que facilita la cooperación entre las computadoras y las personas.

Desde un punto de vista semántico, a finales de los 90, recogemos definiciones orientadas a modelos o patrones de diseño. Así, por ejemplo, Fox en compañía de Marchionini, argumentan en mayo de 1999, que el trabajo de la biblioteca digital se desarrolla en un complejo diseño espacial formado por cuatro dimensiones: comunidad, tecnología, servicios y contenido (Gary Marchionini and Edward A. Fox, 1999).

Por último, una de las definiciones más aceptadas y seguida por numerosos proyectos, como por ejemplo el de Greenstone, presenta a las bibliotecas digitales como “complejos sistemas de conocimiento, información y datos que ayudan en: satisfacer las necesidades de información de los usuarios (Sociedades), proporcionan servicios de información (eScenarios), organizan la información de forma accesible (eStructuras), gestionan el almacenamiento de la información (eSpacios), y hacen llegar la información a los usuarios (Secuencias). Se trata del modelo 5S<sup>3</sup> propuesto por Fox y otros en Virginia Tech. en julio de 1999 (Edward A. Fox, 1999).

Finalmente para concluir este capítulo, a continuación se enumeran algunas de las ideas fundamentales recopiladas del análisis de las definiciones revisadas:

<sup>2</sup> Tim Bernes-Lee, actualmente director del W3C (*World Wide Web Consortium*), concibió la idea de un proyecto de hipertexto global, que años más tarde se convertiría en las WWW.

<sup>3</sup> Modelo 5S ó modelo de las cinco eses argumenta la construcción de las bibliotecas digitales en torno a cinco conceptos fundamentales: *Sociedades*, *eScenarios*, *eStructuras*, *eSpacios* y *Secuencias*.

- Las colecciones de bibliotecas digitales no incluyen solamente sustitutos de libros y otros objetos físicos; también incluyen objetos digitales que no se pueden representar o distribuir en formato impreso.
- Las bibliotecas digitales no son entidades individuales.
- Requieren de una tecnología para interconectar y enlazar los servicios y recursos de información distribuidos entre las muchas bibliotecas digitales repartidas y que trabajan en un entorno de colaboración.
- A su vez, estos enlaces deben resultar transparentes a los usuarios, con el fin de percibir un acceso universal a los servicios y recursos de información disponibles en todo el universo digital.
- Proporcionan acceso a colecciones de información en formato texto, imagen, gráfico, audio y vídeo.
- Mejoran en materias de colaboración como la investigación y el aprendizaje, al reducir las barreras de tiempo y distancia.
- La biblioteca tradicional, con sus procesos habituales de selección, organización, catalogación y preservación, no desaparece. Simplemente se adapta a la era digital, para lo cual, entre otras tareas, será necesario añadir una gestión digital y codificación binaria de los recursos de información.

## 1.4 Clasificación

Tal y como hemos podido observar en el capítulo anterior, el concepto de biblioteca digital no es hoy, un concepto nuevo. Desde el punto literario, la biblioteca digital es una evolución de conceptos, de uso de términos que han ido surgiendo en la literatura, de neologismos que aparecen sucesivamente tales como: biblioteca electrónica, híbrida, digital o virtual. En esta línea aparecen diversas clasificaciones que abordan diferentes tipos de bibliotecas según su grado de automatización:

- Biblioteca clásica
- Biblioteca automatizada
- Biblioteca electrónica
- Biblioteca digital
- Biblioteca híbrida

En concreto, las bibliotecas híbridas reúnen colecciones impresas y colecciones digitales. Más que un tipo de biblioteca es un modelo de armonización entre los medios en papel y digital y sus formas de gestión (Díaz Hermoso L. M., 2001). Este término, se utilizó mucho en los últimos años de la década de los 90, sin embargo ha caído en desuso al profundizar en el concepto de la biblioteca digital. En ocasiones, se han llegado a utilizar indistintamente los términos de biblioteca electrónica, digital y virtual.

Este análisis sobre el concepto de biblioteca digital, hace que en el año 2000, Taly Sharon y Ariel J. Frank (Sharon T, Frank A. J., 2000), hablen de las bibliotecas digitales como un tipo más de biblioteca, pero que en ningún momento emerge como sustituta de la biblioteca clásica y tradicional. Como consecuencia de este enfoque surge una nueva clasificación, en la que se distinguen tres grandes tipos de biblioteca:

- Biblioteca analógica: biblioteca de papel clásica con sus ficheros y archivos manuales.

- Biblioteca automatizada o híbrida: biblioteca analógica pero con un catálogo automatizado.
- Biblioteca digital: biblioteca automatizada en la que la mayor parte de la información es digital.

A su vez, dentro de las bibliotecas digitales se puede ahondar en una nueva clasificación atendiendo a sus interrelaciones y al grado de complicidad con Internet, con otras tres clases:

- Biblioteca digital autónoma (BDA): biblioteca clásica normal que se implementa de manera completamente automatizada. Sus fondos son digitales (escaneados o digitalizados) y el material está centralizado.
- Biblioteca digital federada (BDF): federación de varias bibliotecas digitales autónomas, interconectadas y centradas en un tema común.
- Biblioteca digital recolectada<sup>4</sup> (BDR): biblioteca virtual que proporciona acceso resumido al material relacionado disperso en la red. Este tipo de biblioteca únicamente maneja metadatos con hiperenlaces a través de la red a los fondos.

A partir de estas clasificaciones, se puede deducir que las BDA y BDF pueden existir tanto fuera como dentro de la red. Además, los ítems se adquieren electrónicamente o están completamente digitalizados y/o escaneados. En cambio, las BDR sólo existirán en la red. Además los ítems se encuentran dispersos en numerosos servidores y se accede a ellos vía recuperación directa utilizando protocolos estándar como HTTP, FTP., etc. La BDR sólo aloja metadatos de los ítems y, además, su depósito es pequeño y compacto. Obviamente, persiguiendo objetivos como el de la disponibilidad y el acceso universal a cualquier recurso de información, este último tipo de biblioteca digital juega un papel crucial de cara a las demandas actuales de la sociedad actual.

Es también notable la similitud que existe entre la biblioteca digital recolectada y los actuales motores de búsqueda en la web. Los motores de búsqueda exigen un enorme esfuerzo organizativo, proporcionan al usuario demasiada información ruidosa, pero son útiles para información que se necesita muy rápidamente. Por otra parte, las bibliotecas digitales necesitan un modesto esfuerzo de apoyo, proporcionan al usuario información especializada, pero deben estar disponibles de antemano si bien son superiores en cuanto a calidad y facilidad de uso. Es importante observar que estos dos paradigmas no entran en conflicto ni son exclusivos sino que son, por naturaleza, complementarios (Sharon T, Frank A. J., 2000).

## 1.5 Problemática

Las bibliotecas digitales no traen consigo sólo ventajas. Normalmente cualquier avance en tecnología implica también un sometimiento a ciertas amenazas como pueden ser las barreras económicas y sociales del mundo real, que sin ninguna duda, marcan una nueva brecha en el mundo virtual. Esta brecha se define como la separación que existe entre las personas (comunidades, estados, países y regiones) que utilizan las nuevas tecnologías de la información (NTI) como una parte rutinaria de su vida diaria, y aquellas que no tienen acceso a ellas y que aun cuando las tengan, no saben como utilizarlas. Puede definirse en términos de la desigualdad de posibilidades que existen para acceder a la información, al conocimiento y a la educación mediante las NTI. La brecha digital no se relaciona sólo con aspectos exclusivamente de

---

<sup>4</sup> Del inglés *harvested*. Uno de los protocolos más importantes empleado en este subtipo de bibliotecas digitales es el OAI-PMH (*Open Archives Initiative – Protocol for Metadata Harvesting*) cuya tarea fundamentalmente se reduce al intercambio de metadatos asociados a los recursos de información de las diferentes bibliotecas digitales.

carácter tecnológico, es un reflejo de una combinación de factores socioeconómicos y, en particular, de limitaciones y falta de infraestructura de telecomunicaciones e informática.

Ahora bien, para que ocurra la evolución de la biblioteca tradicional a la biblioteca digital, son necesarios innumerables estudios e investigaciones y muchas acciones de carácter formativo y evaluativo, para que se produzca también una evolución de los aspectos sociales que rodean a las bibliotecas.

En resumen, algunos de los problemas que frenan la expansión de las bibliotecas digitales son:

- Disponibilidad: Todo lo que existe registrado (impresos, fotografías, películas, pinturas, grabados, música escrita y grabada, etc.) tendría que convertirse a digital de forma que estuviera disponible a todos los usuarios dentro y fuera de la red.
- Recuperación: Cada usuario tiene que poder acceder rápida y fácilmente a todos los documentos de este universo digital.
- Coste: Los precios de los documentos, deben ser asequibles y en ningún caso superar sus equivalentes en el mercado tradicional.
- Autenticidad: Hay que tener la certeza de que el documento al que se quiere acceder es el auténtico y original.
- Integridad: Los documentos tienen que conservar su integridad y fiabilidad una vez recuperados.
- Factores legales: La protección de los derechos de autor de los recursos de información accesibles universalmente deben estar garantizados. Las nuevas tecnologías de la información, se han convertido hoy en día en uno de los bienes mejor cotizados y más comercializables.

Continuando con los factores legales, en la Era Digital la guerra por la comercialización del conocimiento de las industrias culturales se hace cada vez más áspera pues alega que la “producción cultural” está en peligro a causa de los circuitos alternos para la distribución del mismo, algunos con fines lucrativos, otros absolutamente no. En este contexto, las licencias *Creative Commons*<sup>5</sup> son una alternativa legal a la actual legislación de Derechos de autor de cara al TLC (Tratado del Libre Comercio) -dentro del ATPDA<sup>6</sup> con los EE.UU.- donde el capítulo sobre propiedad intelectual es extremadamente sensible pues lo que se pretende es restringir aun mas los derechos del creador respecto al derecho de copia.

Esto no ayuda en absoluto al autor cuando este quiere conseguir la exposición o distribución amplia de su obra, pues los consumidores, no pueden acceder a ella sin delinquir. Estas licencias tienen como objetivo difundir la experiencia y vía que tomaron muchos empresarios y artistas, quienes, al no poder crear nuevas formas de negocio en la era digital debido a las limitaciones de las actuales leyes sobre derecho de autor, eligieron el modelo que

---

<sup>5</sup> En español Creatividades Comunes, está inspirada en la licencia GPL (Licencia General Pública) de la Fundación por el Software Libre. El usuario puede copiar modificar y difundir una obra sin fines de lucro, siempre y cuando respete la autoría de la creación bajo la misma licencia. *Creative Commons*, por lo tanto, va más allá de ser una alternativa legal para el creador y consumidor, ya que es una innovadora manera de salvaguardar las creaciones e innovaciones para que alimenten los conocimientos de Dominio Público, los que están en peligro si no se realizan los cambios pertinentes en las actuales legislaciones sobre propiedad intelectual.

<sup>6</sup> Sigla para la Ley de Preferencias Arancelarias.

ofrece las licencias *Creative Commons* para asegurarse un beneficio en su inversión creativa sin caer en la piratería.

El proyecto mas conocido para resguardar la distribución y copias de una creación a través de una red digital es el DOI (*Digital Object Identifier*)<sup>7</sup>. Este estándar resguarda la propiedad intelectual del documento, pero por el momento sólo se preocupa del impreso. Se le ha venido a considerar como “el código de barras de los recursos digitales”. El DOI difiere del mecanismo utilizado habitualmente en Internet para apuntar a cualquier recurso<sup>8</sup> ya que identifica al objeto desde el punto de vista de entidad, en lugar de hacerlo desde el punto de vista de donde puede localizarse. Además cada DOI está asociado a una serie de meta-información que consiste en un conjunto de información bibliográfica y comercial concerniente al contenido. Las ventajas que se obtienen con la adopción del DOI para con los recursos digitales son básicamente tres:

- Persistencia de originalidad e identificación de la información digital.
- Portabilidad a entornos de redes de información digital.
- Interoperabilidad semántica a través de la meta-información asociada al recurso.

En este sentido también surgen normas para garantizar la originalidad de la información por medio de filigranas, o marcas digitales (*Digital Watermarks*) mapas de bits para texto, imagen, vídeo y sonido de fondo en el caso de grabaciones de audio. Estas marcas son imperceptibles por el ser humano pero detectables por máquinas, el problema es que con la tecnología adecuada se elimina, así que aún no hay total seguridad.

En el campo de los recursos de información musicales se han apoyado diversas medidas entre las que destaca la Iniciativa Segura de Música Digital (SDMI), que tiene por objeto desarrollar especificaciones para la distribución y reproducción de música digital que protejan los derechos de autor. Por otro lado la especificación OPIMA (*Open Platform Initiative for Multimedia Access*), desarrollada bajo el amparo de la Comisión Internacional de Electrotecnia (IEC), pretende elaborar un marco donde los proveedores y usuarios puedan intercambiar información y productos de forma segura.

Todas estas iniciativas persiguen trasladar el modelo de negocio tradicional al mundo de la Red.

## **1.6 Construcción de una biblioteca digital**

### **1.6.1 Digitalización y conversión a formato digital**

Uno de los mayores retos para la digitalización de documentos de diversos tipos, es la generación de la propia información digital, debido a que en la mayoría de los casos ésta no se encuentra en tal formato, es decir, se hace necesario la realización de una transformación de la información que se encuentra en forma analógica, ocasionando la necesidad de que se establezcan reglas y procedimientos para su digitalización.

Sin embargo ésta no es una tarea sencilla, pues es necesario considerar varios aspectos como el estado físico y legal que guarda la información a digitalizar, la comunidad y el uso a la que está

---

<sup>7</sup> La IDF (*International DOI Foundation*) es una organización sin ánimo de lucro creada en 1998, que da carácter gubernamental al sistema DOI.

<sup>8</sup> *Uniform Resource Locator* (URL), es decir, localizador uniforme de recurso. Es una secuencia de caracteres, de acuerdo a un formato estándar, que se usa para nombrar recursos, como documentos e imágenes en Internet, por su localización.

dirigida, el tipo de formato analógico en que se encuentra (audio, video, microfilm, papel, etc.), además de otros factores propios de la digitalización, como los problemas asociados tales como el tamaño de archivo (peso), transporte, despliegue y almacenamiento de la información generada.

Es importante mencionar que al momento de pensar en digitalizar un documento, cualquiera que éste sea, también deberá estar implícita la parte de la vigencia y portabilidad. Esto significa que cuando un formato no es un estándar tiende a desaparecer con el paso del tiempo y en consecuencia resulta difícil su conversión hacia otro formato que pudiera ser más estable.

Los medios técnicos que básicamente se utilizan en esta fase de digitalización son: escáneres, cámaras digitales y dispositivos de reconocimiento óptico de caracteres. Por norma general, el documento una vez digitalizado, no suele estar aún en un estado apropiado para su almacenamiento y catalogación, por ello, posteriormente siempre son necesarias otras fases de preparación. Estas tareas son principalmente:

- Edición: Se elimina “ruido” procedente de la fase de digitalización.
- Conversión: A fin de preservar el documento digital el mayor tiempo posible a lo largo de los tiempos, conviene escoger el formato de almacenamiento más adecuado de acuerdo a su naturaleza original del documento.
- Compresión: Normalmente a fin de preservar el documento original con suficiente calidad su tamaño es considerablemente alto, por tanto, es necesaria una compresión de la información a fin de ocupar el menor espacio posible y así agilizar tanto su almacenamiento como posterior recuperación.

Sin embargo, y a pesar de que parecen estar las ideas muy claras, aún no existe un mecanismo estándar o una solución acabada que resuelva el problema de forma completa.

### 1.6.2 Almacenamiento

El objetivo es garantizar que la información seleccionada ocupe el menor espacio posible, que los medios de registro ofrezcan un largo período de almacenamiento y que puedan disponerse en un sistema ordenado que permita la rápida y fácil recuperación de los datos relevantes.

Las tendencias generales de todos los dispositivos de almacenamiento masivo de información son: el incremento continuo de su capacidad, una mayor velocidad, más economía, menor tamaño y mayor fiabilidad. El tipo de formato y el dispositivo de almacenamiento son elementos de fundamental consideración en toda biblioteca digital. Debe considerarse el volumen de información inicial y sus proyecciones de crecimiento a corto, medio y largo plazo. Además, será necesario estimar, el número de accesos que se realizarán a los documentos y la rapidez con que éstos deben recuperarse.

### 1.6.3 Recuperación

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.

En este campo existe una gran experiencia acumulada por los bibliotecarios quienes por años han estudiado los hábitos y formas diversas de expresión de las necesidades de información de los usuarios. Así, planteamientos como los de *Mijailov* y *Guiliariekskii*: "recuperar la información en cualquier momento", adquieren una gran relevancia porque las bibliotecas digitales permiten al acceso a la información en cualquier momento y además desde cualquier sitio.



A pesar de la gran experiencia acumulada, la nueva realidad impone nuevas formas de abordarla. En la línea de estos nuevos desarrollos, un amplísimo campo de investigación, gira en torno al campo de los metadatos<sup>9</sup>. El uso de esta meta-información es fundamentalmente para refinar de alguna forma las consultas y así conseguir que los resultados sean mucho más precisos. A su vez los metadatos proporcionan flujos de trabajo que van a permitir la conversión automática de formatos, para lo cual, es necesario que describan el contenido y estructura de los datos.

#### 1.6.4 Metadatos

Un registro de metadatos (Rust G., 1998) consiste en un conjunto de atributos o elementos, necesarios para poder describir un objeto digital. Continuando con este planteamiento, existen diferentes tipos de metadatos:

- Administrativos: Usados en la gestión y administración de recursos de información (documento fuente original, fecha de creación, autor, condiciones de uso, formato del documento digital, relaciones con otros objetos y operaciones permitidas).
- Descriptivos y estructurales: Utilizados para representar recursos de información (bibliográficos, temáticos y de evaluación).
- Preservación: Para salvaguardar los recursos de información.
- Técnicos: Relativos a cómo funcionan los sistemas o el comportamiento de los metadatos.
- De uso: Relativos al nivel y tipo de uso que se hace con los recursos informativos.

La idea del uso de estos metadatos fundamental viene marcada no sólo por la recuperación de datos, sino también, para marcar asociaciones con otros objetos y garantizar la interoperabilidad con otros sistemas de información.

Algunos de los estándares para la gestión de metadatos son: TEI (*Text Encoding Initiative*), MET (*Metadata Encoding and Transmission Standard*), MARC-XML y una de las iniciativas más difundidas es la DMCI, Iniciativa de Metadatos *Dublin Core*<sup>10</sup>, una organización cuyos fines son la promoción y difusión de normas interoperables sobre metadatos, así como el desarrollo de vocabularios especializados controlados para representar recursos que permitan el desarrollo de sistemas de recuperación más inteligentes.

La DCMI es la iniciativa internacional de metadatos más sólida e importante para la organización y recuperación de información en Internet en forma normalizada, eficaz y con un propósito general; además, el conjunto de elementos *Dublin Core* es un estándar internacional ISO 15836-2003, con independencia de que, en cada ámbito de información, se desarrollen esquemas propios de meta-información para mejorar la recuperación en la red.

Las principales características del *Dublin Core* son:

- Simplicidad: Reduce considerablemente los costos y promueve la interoperatividad. La simplicidad no significa acomodar o desechar la semántica y las funciones enriquecidas, soportadas por otros sistemas complejos de metadatos: De hecho, DC estimula el uso de formatos de metadatos enriquecidos en

<sup>9</sup> Literalmente los metadatos son "*datos sobre los datos*", por ejemplo, un catálogo de biblioteca es un registro de metadatos, en el sentido de que describen publicaciones. Algunas definiciones más atrevidas hablan de descripciones estructuradas sobre instancias de algún contenido que ayudan a identificarlo de entre otros muchos.

<sup>10</sup> <http://www.dublincore.org>

combinación con DC y, a menudo, es también, el punto de partida para la creación de descripciones más complejas.

- Interoperatividad semántica: Establece vínculos y relaciones con otras normas, sin sacrificar su autonomía.
- Consenso internacional: DC se ha convertido en una parte importante de la infraestructura de Internet y su uso se ha extendido a escala internacional.
- Extensibilidad: Propiedad que deriva, en parte, de la flexibilidad y, en parte, de la definición de elementos estructurados con distintos grados de complejidad y requerimientos. También, se relaciona con su capacidad de convertirse fácilmente a otros formatos, entre ellos y a pesar de su complejidad, al propio formato MARC.
- Flexibilidad: Nada en el DC es obligatorio, todos los elementos son opcionales y repetibles, así el usuario elige la profundidad de su descripción. El formato además, permite incorporar desde estructuras simples hasta aquellas más elaboradas semánticamente.

La asignación de metadatos a los objetos de información que conforman las colecciones de las bibliotecas digitales es uno de los aspectos clave para la gestión y recuperación de información en los nuevos servicios y sistemas de información digital

Estos esquemas de metadatos, junto a la codificación sintáctica de lenguajes: HTML, XML<sup>11</sup>, RDF<sup>12</sup>, así como la Open Archives Initiative (OAI), ofrecen normas de descripción de contenido -ontologías, *topic maps*, tesauros, etc.- y toda una serie de protocolos para el intercambio de información. Se verá un poco más a fondo algunas de estas propuestas:

- *Extensible Markup Language* (XML) (Lenguaje de marcado extensible) es un metalenguaje permite definir la estructura de tipos específicos de documentos electrónicos (o datos) por medio de Definiciones de Tipo de Documento (DTD o XML Schemas), definiendo conjuntos de etiquetas y atributos (vocabularios).
- *Resource Description Framework* (RDF) (Infraestructura para la Descripción de Recursos) es una base para procesar metadatos; proporciona interoperabilidad entre aplicaciones que intercambian información legible por máquina en la Web. RDF destaca por la facilidad para habilitar el procesamiento automatizado de los recursos Web. RDF puede utilizarse en distintas áreas de aplicación; por ejemplo: en *recuperación de recursos* para proporcionar mejores prestaciones a los motores de búsqueda, en *catalogación* para describir el contenido y las relaciones de contenido disponibles en un sitio Web, una página Web, o una biblioteca digital particular, por los *agentes de software inteligentes* para facilitar el intercambio y para compartir conocimiento; en la *calificación de contenido*, en la descripción de *colecciones de páginas* que representan un "documento" lógico individual, para describir los *derechos de propiedad intelectual* de las páginas web, y para expresar las *preferencias de privacidad* de un usuario, así como las *políticas de privacidad* de un sitio Web. RDF junto con las *firmas digitales* será la clave para construir el "Web de confianza" para el comercio electrónico, la cooperación y otras aplicaciones.
- *Open Archives Initiative* (OAI) es una iniciativa para el depósito de documentos científicos a texto completo desde un punto de vista arquitectónico de un sistema abierto. La idea fundamental de esta organización es desarrollar y promover

---

<sup>11</sup> Lenguaje de marcado extensible.

<sup>12</sup> Infraestructura para la Descripción de Recursos.

estándares de interoperabilidad que faciliten la disponibilidad y difusión de contenidos en Internet procedentes de una gran variedad de proveedores. A su vez pretende mejorar el acceso a publicaciones digitales (*eprints*) con la idea de abrir el acceso a un rango de materiales digitales mucho más amplio. OAI trata la comunicación de metadatos sobre cualquier material almacenado en soporte electrónico. La propuesta de OAI para el intercambio de metadatos se centra en el protocolo OAI-PMH<sup>13</sup> (*OAI Protocol for Metadata Harvesting*). Se trata de un protocolo para la recolección (“*harvesting*”) de metadatos. Tiene una arquitectura basada en el modelo cliente – servidor donde los primeros, llamados archivos, ponen a disposición del público metadatos en formato *Dublin Core* para que puedan ser recuperados por los segundos. La comunicación se realiza mediante el protocolo HTTP y las respuestas están codificadas en XML (Barrueco J. M., 2003).

En esta línea, las Bibliotecas Digitales Recolectadas (BDR), encuentran una base muy importante para solucionar los problemas de interoperabilidad presentes en todo momento. A su vez, muchos de los motores de búsqueda de la web han comenzado a incorporar ya a sus procedimientos de catalogación automáticos estos mecanismos de intercambio de metadatos. Este es el caso por ejemplo de Google, quien ha comenzado a incorporar como mecanismos de indexación el protocolo OAI-PMH<sup>14</sup>.

### 1.6.5 Identificación

Una de las metas fundamentales para una buena gestión del catálogo de la biblioteca digital, es la elección adecuada de los identificadores unívocos o persistentes. Esta idea de utilizar identificadores persistentes, surge como respuesta para solucionar el problema originado por los cambios de ubicación y/o nombre de los archivos. El objetivo de estos identificadores es redireccionar a los documentos, aunque éstos hayan cambiado de ubicación en la red (cambio de URL<sup>15</sup>).

Tres ejemplos de esquemas propuestos para solucionar el problema de la persistencia de los identificadores son los PURL’s, URN’s e Identificadores de Objetos Digitales (DOI).

Los PURL’s, son URL’s persistentes. Se trata de un esquema desarrollado por la OCLC (*On Line Computer Library Center*) con la intención de separar el nombre de un documento de su ubicación y, por lo tanto, incrementar la probabilidad de que sea localizado siempre. Los PURL’s trabajan por medio del mapeo de un único y nunca cambiante PURL a un verdadero URL. Si se mueve un documento, se actualiza el URL, pero el PURL permanece igual. Un usuario podría solicitar un documento a través de un PURL, un servidor de este tipo busca el correspondiente URL en una base de datos, y entonces éste se utiliza para transferir el documento al usuario.

El URN, nombre uniforme de recurso (*Uniform Resource Name*), es un sistema de identificación para la información electrónica no limitada a las publicaciones. La sintaxis del

<sup>13</sup> <http://www.openarchives.com>

<sup>14</sup> *Using OAI-PMH with Google Sitemaps* (<http://sitemaps.blogspot.com/2005/09/using-oai-pmh-with-google-sitemaps.html>).

<sup>15</sup> URL, *Uniform Resource Locator*, es decir, localizador uniforme de recurso. Es una secuencia de caracteres, de acuerdo a un formato estándar, que se usa para nombrar recursos, como documentos e imágenes en Internet, por su localización. El concepto de URL ha sido incorporado dentro del más general de URI (*Uniform Resource Identifier* – Identificador Uniforme de Recurso), pero el término URL aún se utiliza ampliamente. Puede obtenerse más información al respecto en la RFC 2396 (<http://www.ietf.org/rfc/rfc2396.txt>).

URN permite, si se desea, la integración de un ISBN y otros números normalizados internacionales.

De todos los identificadores persistentes, el más extendido es el DOI (*Digital Object Identifier*). Es un dígito de control creado para identificar unívocamente objetos digitales, entre los que se encuentran las revistas electrónicas o partes de ellas. Se trata de un proyecto gestionado por la *Internacional DOI Foundation*.

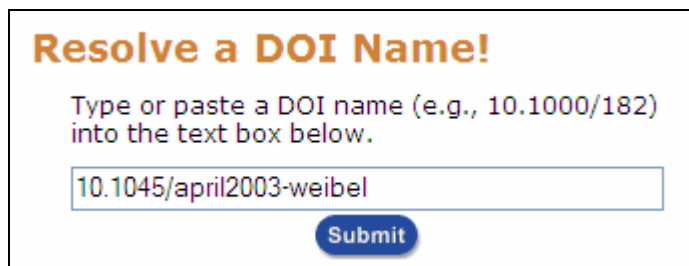
Los objetivos planteados con la creación de este dígito son proporcionar un marco para la gestión de la propiedad intelectual, favorecer la comunicación entre clientes (lectores) y editores, facilitar el comercio electrónico y posibilitar la gestión del copyright de forma automatizada. Como el ISBN y el ISSN, la razón fundamental para la creación de este sistema es permitir la compra y venta de publicaciones electrónicas. Se trata de proteger los derechos de autor, fundamentalmente de cara a la distribución de copias digitales.

El DOI es básicamente un código alfanumérico que incluye dos partes: prefijo y sufijo. El prefijo a su vez consta de dos componentes: la primera componente identifica la secuencia como DOI y la segunda identifica la editorial. El sufijo identifica el objeto digital. El sufijo también puede integrar el ISSN.

La utilidad del DOI no solamente se basa en la posibilidad de identificar y describir objetos digitales sino que también se deriva de una serie de servicios adicionales asociados al propio registro. El uso del DOI está en continua expansión por todo el mundo, particularmente en el sector editorial científico-técnico y académico así como en el ámbito referente a la publicación de información pública.

Recientemente se ha adoptado el DOI como base para el intercambio de información en el marco del proyecto Crossref<sup>16</sup>.

Un ejemplo de referencia DOI es la siguiente: 10.1045/april2003-weibel. Para resolver el DOI y acceder al documento relacionado, se puede utilizar el formulario de resolución de DOI's de la página de inicio del sitio web de la *Internacional DOI Foundation*.



**Ilustración 1. Formulario de resolución de DOI - <http://www.doi.org>**

El proyecto Crossref también dispone de un formulario para resolver DOI's, tal y como se puede observar en la siguiente figura:

---

<sup>16</sup> El objetivo del proyecto Crossref (<http://www.crossref.org>), es relacionar los distintos artículos pertenecientes a las revistas científicas que ofrece cada uno, de forma que, un usuario pueda consultar las publicaciones de los distintos editores seleccionando el enlace de las referencias citadas en los artículos. En definitiva, se accede al texto completo de un artículo desde la cita o referencia mencionada en otro artículo diferente, que se encuentre dentro de los fondos digitales de las entidades asociadas. En la actualidad existen más de 11 millones de DOI's registrados en el consorcio Crossref, que cubre unas 10.000 revistas.

**Ilustración 2. Formulario de resolución de DOI - <http://www.crossref.org>**

### 1.6.6 Software de base

Con el fin de poder gestionar completamente la biblioteca digital, necesitamos un software que nos resuelva fundamentalmente las siguientes funcionalidades:

- Interfaz para la administración y gestión.
- Sistema eficiente de búsqueda.
- Posibilidad de manipulación de los objetos.
- Gestión de los derechos digitales.

*Greenstone* es un completo sistema para la creación y presentación de colecciones compuesta de miles o millones de documentos, que incluyen texto, imágenes, sonido y video. Es software libre bajo licencia GNU GPL, desarrollado por el equipo del Computer Science Dep de la University of Waikato, Nueva Zelanda, con el apoyo de la UNESCO.

Algunas de las características fundamentales de *Greenstone* son las siguientes:

- Es multiplataforma: funciona en Linux, Unix, Windows y Mac.
- Es multilingüe.
- Potente sistema de información que trabaja con gigabytes.
- Capacidad para recuperar y gestionar sitios web.
- Capacidad para integrar y gestionar diversos estándares de metadatos.

Ventajas en el uso de *Greenstone*:

- Capacidad de trabajar con colecciones.
- Colecciones autónomas y con estructura propia.
- Capacidad para procesar diferentes tipos de formatos de documentos.
- Herramienta fácil y sencilla para dar acceso a las colecciones digitales.
- La interfaz es posible de personalizar.
- Armado de búsquedas globales.
- Tiene incorporado OAI para poder recolectar metadatos.
- Soporte a Metadatos.
- Gestiona y trata sitios web como colecciones.
- Posibilita exportar las colecciones a dispositivos multimedia como el CD-ROM.
- Continuo desarrollo ya que posee el respaldo de la UNESCO.
- Existencia de comunidades y listas de discusión.

## 1.7 Componentes

En este capítulo vamos a ver el soporte tecnológico de la biblioteca digital que incluye el software y el hardware necesario para la existencia de la misma.

Estrechamente relacionado con el concepto de “arquitectura de la información”, se utiliza para referirse al diseño de sistemas de organización, etiquetado, navegación y búsqueda, con el fin de ayudar a las personas a encontrar y manejar la información con éxito.

Los elementos que integran la biblioteca digital son numerosos y guardan relación con la representación, almacenamiento, tratamiento y recuperación de la información. Entre ellos se encuentran los siguientes:

- Servidores donde residirá la colección digital. Formatos utilizados en la edición electrónica con el fin de hacer compatibles los documentos digitales que circulan en entornos informáticos heterogéneos, tanto desde el punto de vista físico como lógico, tales como HTML, SGML o XML.
- Motores o sistemas de búsqueda que permiten al usuario recuperar la información almacenada. Protocolos tales como el Z39.50, centrado en una búsqueda distribuida dirigida por el usuario, o incluso más recientemente, el ya citado OAI-PMH que rechaza la búsqueda distribuida, son piezas fundamentales para la recuperación de la información. Los motores de búsqueda tienen gran importancia en la recuperación de la información, ya que permite a los programas clientes interrogar a servidores distantes y recibir los resultados.
- Base de datos. La base de datos de la biblioteca digital contiene todos los documentos que forman su colección, relacionados y organizados con una lógica coherente, para que puedan ser consultados por distintos usuarios y a través de distintas aplicaciones. Para llevar a cabo el diseño de una base de datos es necesario un buen análisis para definir los campos necesarios y la relación que habrá entre los diferentes elementos.
- Seguridad. El problema de la seguridad no es nuevo en las bibliotecas. Pero en las bibliotecas digitales este problema presenta otras características. Afecta a dos aspectos: la seguridad de la información y la seguridad física. La primera se logra con la utilización de mecanismos tales como: cortafuegos, codificadores, protocolos seguros, etc., que permiten ejercer un control sobre quien accede a la información. En cuanto a la seguridad física, está fuera del apartado de la arquitectura técnica de la que hablamos aquí y se logra protegiendo instalaciones y equipos a través de vigilancia, tarjetas magnéticas, códigos de acceso, etc.
- La interfaz gráfica. El desarrollo de una interfaz adecuada es muy importante en el desarrollo de la biblioteca, ya que de ella dependerá en gran medida el éxito de la misma. Para su diseño es importante contar, en primer lugar, con las necesidades del usuario, incluso su formación en el campo de la búsqueda de información. En la interfaz de usuario se pueden distinguir dos partes. La primera se refiere a la aplicación local o distribuida (sitio web) desde el que interactuará el usuario, durante su navegación por la biblioteca digital. La segunda la constituye el diseño de la interfaz de búsqueda en el que hay que tener en cuenta ciertos aspectos como: el nivel de experiencia o especialización de los usuarios; el tipo de información que el usuario busca; la cantidad de información para la búsqueda.
- Metadatos, a los que ya nos hemos referido.

- Otros elementos como equipamiento (escáneres, impresoras), redes, etc., son también necesarios para la implementación de las bibliotecas digitales.

## 1.8 Arquitectura OAI-PMH

El protocolo OAI-PMH (*Open Archives Initiative – Protocol for Metadata Harvesting*) permite la transmisión y recolección de información, en un entorno distribuido y de colaboración como es Internet, con el fin de permitir su traspaso de un repositorio<sup>17</sup> a otro. El protocolo propuesto por OAI se sustenta sobre una arquitectura cliente - servidor (típica de la familia de protocolos TCP/IP de Internet) y define un marco de trabajo de interoperabilidad para dos tipos de participantes: proveedores de datos (*data providers*) y los proveedores de servicios (*services providers*).

De un lado, los proveedores de datos mantienen por medio del protocolo OAI-PMH uno o más repositorios de información (localizados en un servidor web). De otro lado, los proveedores de servicios interrogan a los proveedores de datos y utilizan los metadatos como base para generar servicios de valor añadido. Los metadatos que se transmiten vía OAI-PMH deberán codificarse en *Dublin Core* sin calificar con el fin de minimizar los problemas derivados de las conversaciones entre múltiples formatos.

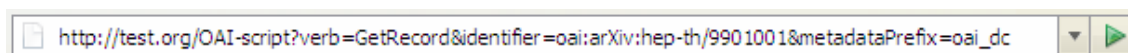
Básicamente OAI-PMH utiliza transacciones HTTP para emitir preguntas y obtener respuestas entre un servidor de datos y un cliente o servicio recolector de metadatos. El segundo puede pedir al primero que le envíe metadatos según diferentes criterios como por ejemplo la fecha de creación de los metadatos.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

Las peticiones se realizan utilizando los métodos GET o POST del protocolo HTTP y constan de una lista de opciones (o verbos<sup>18</sup>) con la forma de pares del tipo: clave=valor. Existen seis posibles peticiones del cliente al servidor:

- *GetRecord*. Utilizado para recuperar un registro concreto.
- *Identify*. Utilizado para recuperar información sobre el servidor: nombre, versión del protocolo que utiliza, dirección del administrador, etc.
- *ListIdentifiers*. Recupera los encabezamientos de los registros, en lugar de los registros completos.
- *ListRecords*. Igual que el anterior pero recupera los registros completos.
- *ListSets*. Recupera un conjunto de registros.
- *ListMetadataFormats*. Devuelve la lista de formatos bibliográficos que utiliza el servidor.

En la siguiente figura puede verse un ejemplo de petición y respuesta:

Petición:



<sup>17</sup> Depósito de documentos de cualquier tipo a texto completo.

<sup>18</sup> La lista de opciones se puede leer también como verbos. De ahí la clave: *verb*, como se puede observar en el ejemplo de petición, *verb=GetRecord*

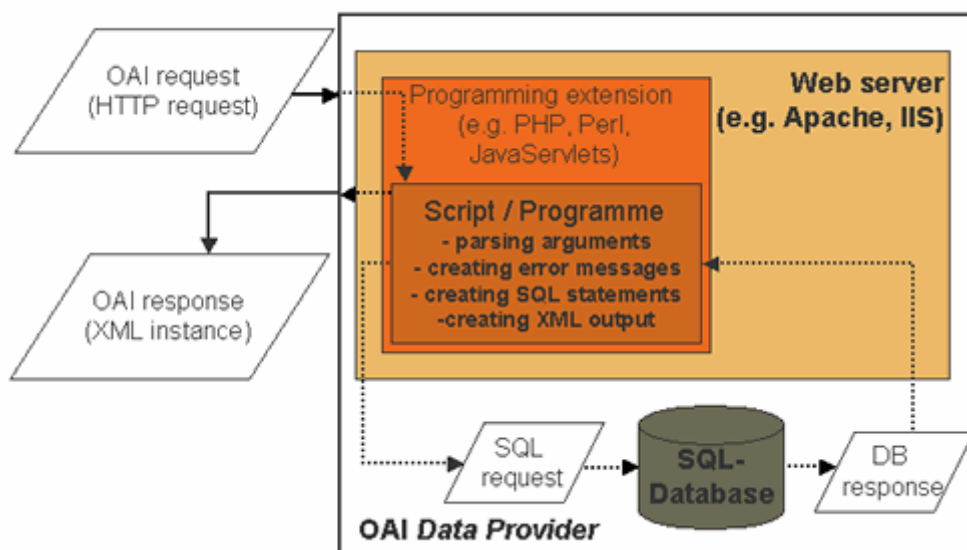
Respuesta:

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
  http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2002-05-01T19:20:30Z</responseDate>
  <request verb="GetRecord" identifier="oai:arXiv:hep-th/9901001"
    metadataPrefix="oai_dc">http://an.oa.org/OAI-script</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:arXiv:cs/0112017</identifier>
        <timestamp>2001-12-14</timestamp>
        <setSpec>cs</setSpec>
        <setSpec>math</setSpec>
      </header>
      <metadata>
        <oai_dc:dc
          xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
          xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
          http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          <dc:title>Using Structural Metadata to Localize Experience of Digital Content
        </dc:title>
        <dc:creator>Dushay, Naomi</dc:creator>
        <dc:subject>Digital Libraries</dc:subject>
        <dc:description>With the increasing technical sophistication of
          both information consumers and providers, there is
          increasing demand for more meaningful experiences of digital
          information. We present a framework that separates digital
          object experience, or rendering, from digital object storage
          and manipulation, so the rendering can be tailored to
          particular communities of users.
        </dc:description>
        <dc:description>Comment: 23 pages including 2 appendices,
          8 figures</dc:description>
        <dc:date>2001-12-14</dc:date>
        </oai_dc:dc>
      </metadata>
    </record>
  </GetRecord>
</OAI-PMH>
```

Por otro lado, hay aspectos que no cubre el protocolo, como por ejemplo la gestión y seguridad de acceso de los clientes, o también, la manera de localizar aquellos servidores que contengan la información requerida por el cliente.

Tal y como se ha dicho anteriormente, 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, con el objetivo de incorporarles algún valor añadido y presentarlos a los usuarios finales.





**Ilustración 3. Arquitectura de OAI-PMH**

Si nos centramos en la arquitectura propia de cada uno de los elementos que componen esta arquitectura encontramos una segunda arquitectura propia para cada uno de ellos.

### 1.8.1 Proveedor de datos

#### 1.8.1.1 Prerrequisitos

- Soporte de almacenamiento de metadatos: Es recomendable que los metadatos se almacenen en una base de datos relacional a la que se pueda acceder mediante consultas SQL, es necesario que exista un único identificador para cada recurso. También se puede utilizar un sistema de ficheros tipo LDAP.
- Soporte Web: Un servidor web.
- Mantenimiento de un API para la posterior implementación del protocolo: Se puede utilizar cualquier lenguaje de programación que permita el desarrollo de aplicaciones web. Java, Perl, PHP son buenas elecciones a este respecto. El API debe contemplar el acceso a base de datos o al sistema de ficheros, si bien no es necesario implementar seguimiento de sesiones.
- URL: Mantenimiento de un identificador para el repositorio o revista único, basado en su URL.
- URI: Mantenimiento de un identificador único para cada recurso que forma parte del repositorio. Es una buena práctica la utilización de DOIs.
- Formato de los metadatos: Se pueden utilizar todos los formatos de metadatos que se desee. Como mínimo debe soportar *Dublin Core* sin cualificar.
- Sellos de tiempo: Se debe mantener un *datestamp* para la creación del recurso y otro para su modificación si es que es necesario.
- Soporte lógico para una jerarquía de conjuntos (*sets*): Permite el mantenimiento de clasificaciones de cara a una ordenación temática y/o conceptual de los recursos.
- Control de flujo: Se consigue a través de una implementación de la reanudación de la señal (*resumption token*) en la que se basa el diálogo entre el proveedor de metadatos y

el servidor de datos. Este requisito no es obligatorio pero sí muy recomendable en cualquier proveedor de metadatos.

### 1.8.1.2 Componentes

- Un analizador que valide los argumentos recibidos a partir de las peticiones OAI.
- Un generador de errores que cree respuestas XML para codificar los mensajes de error.
- Sistema de consulta a la base de datos o al sistema de ficheros para la recuperación de los metadatos.
- Generador de respuestas XML para la codificación de los metadatos que se enviarán como resultado de la petición de *Service Provider* o del *Harvester*.
- Control de flujo que realice una lista de secuencias incompleta cuando los repositorios sean muy grandes de cara a que el envío de la respuesta se haga poco a poco en función de la reanudación de la señal.

La interacción de estos componentes integran la arquitectura del *Data provider*, tal y como se muestra en la siguiente figura:

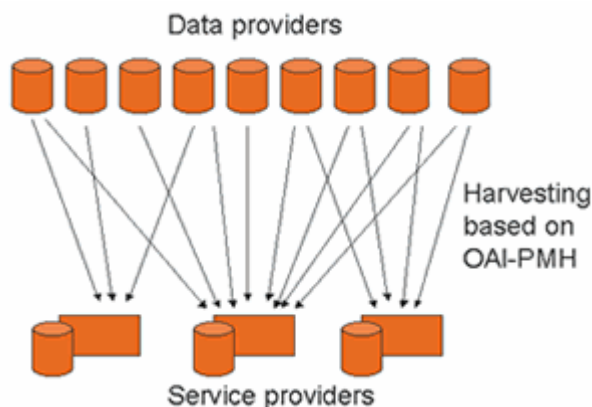


Ilustración 4. Arquitectura del *Data provider*

## 1.8.2 Proveedor de servicios

### 1.8.2.1 Prerrequisitos

- Soporte de almacenamiento de metadatos: Es recomendable que los metadatos sean almacenados en una base de datos relacional a la que se pueda acceder mediante consultas SQL. También se puede utilizar un sistema de ficheros tipo LDAP o bases de datos XML.
- Soporte Web: Servidor web tipo Apache, IIS, etc.
- Mantenimiento de una API para la posterior implementación del protocolo: Se puede utilizar cualquier lenguaje de programación que permita el desarrollo de aplicaciones web. Java, Perl, PHP son buenas elecciones a este respecto. La API debe contemplar el acceso a base de datos o al sistema de ficheros, si bien no es necesario implementar seguimiento de sesiones.

### 1.8.2.2 Componentes

- Gestión del archivo: Sistema de administración web que permita dar de alta nuevos repositorios (*Data Providers*) y gestionar la base de datos.
- Creador de peticiones: Se encarga de generar las peticiones para el proveedor de datos en función de los verbos OAI y de enviarlas a través de HTTP.
- Temporizador (*Scheduler*): Encargado de comprobar si se han hecho modificaciones en los *Data Provider* que forman parte de la iniciativa. Este proceso se puede hacer a mano o automatizar, a través de herramientas como *cron*.
- Control de flujo (*Flow control*): Implementación para la reanudación de la señal en el caso de tener que recibir el listado de recursos de un proveedor de datos en varias veces debido a que este sea demasiado grande.
- Mecanismo de actualización (*Update mechanism*): Mecanismo que se encargue de actualizar la información que se tiene en la base de datos sobre un proveedor de datos.
- Analizador XML (*Parser XML*): Se encarga de analizar las respuestas extraídas de los proveedores de datos, con validación incluida.
- Herramienta de normalización (*Normaliser*): Se encarga de transformar los distintos formatos de metadatos que se reciben de los proveedores de servicios a una estructura homogénea, armonizando su posterior presentación.
- Base de datos (*Database*): Base de datos para almacenar la información sobre los proveedores de datos.
- Comprobador de redundancias (*Duplication checker*): Se encarga de comprobar si dos proveedores de datos distintos repiten algún recurso y lo soluciona asignando un identificador distinto al recurso en función del proveedor de servicios al que pertenezcan.
- Modulo de servicios (*Service module*): Proporciona los servicios al usuario final, que únicamente interactúa con la base de datos que se ha creado después de recopilar los metadatos de los distintos *Data Providers*.

La interacción de estos componentes se materializa en la siguiente figura:

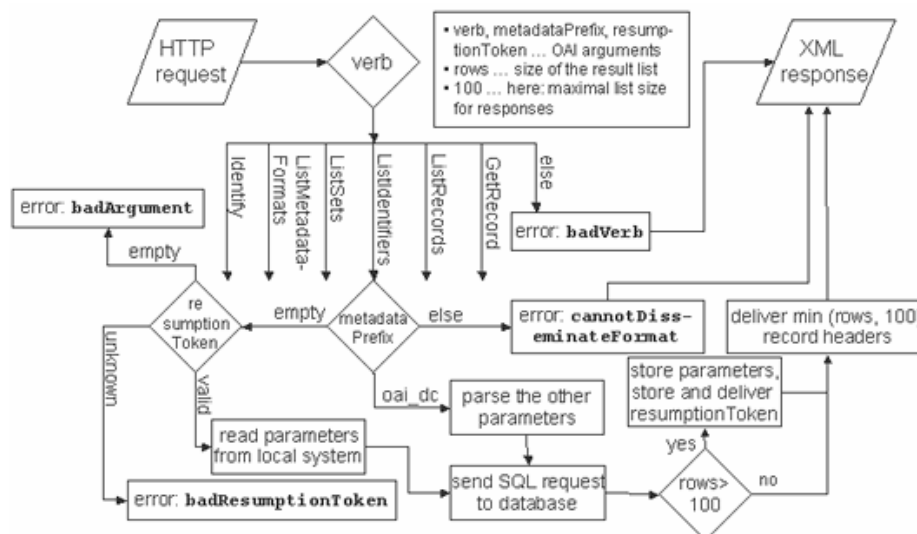


Ilustración 5. Interacción de los componentes del proveedor de servicios

### 1.8.3 Proveedores de datos y servicios

Desde enero de 2001 se ha mantenido un registro de todos los archivos que han implementado el protocolo OAI-PMH. El registro no es obligatorio por lo que se supone que son muchos los archivos que existen y no se han registrado. Entre ellos tenemos como más destacables: arXiv.org junto con el resto de iniciativas que mencionamos en el punto 2, CERN que recoge informes y prepublicaciones en el área de Física o *Citebase* que proporciona datos sobre citas recibidas por los *eprints* almacenados en varios archivos.

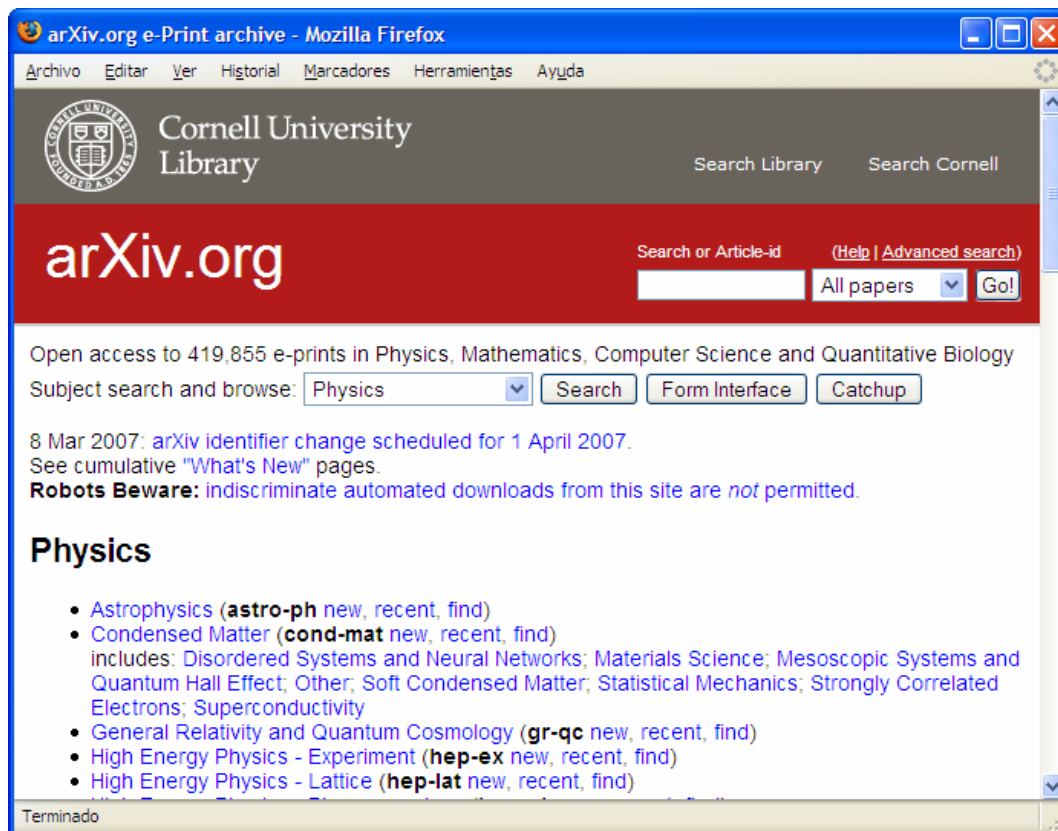


Ilustración 6. arXiv

Otros archivos de interés en el área de la biblioteconomía y documentación son por ejemplo: @rchiveSIC <<http://archivesic.ccsd.cnrs.fr/>>, DLIST (*Digital Library of Information Science and Technology*) <<http://dlist.sir.arizona.edu/>>, E-LIS (*Eprints in Library and Information Science*) <<http://eprints.rclis.org/>>.

Igualmente se mantiene un registro de los servicios creados utilizando los datos proporcionados por los anteriores. Tampoco aquí es obligatorio el registro por lo que es imposible saber cuántos servicios existen realmente en la actualidad. Sin embargo podemos destacar: ARC <<http://arc.cs.odu.edu/>>, OAIster <<http://oaister.umdl.umich.edu/>>, Perseus <<http://www.perseus.tufts.edu/>>.

## 2 Los sistemas de gestión de contenidos

### 2.1 Conceptos previos

#### 2.1.1 Consideraciones generales

Un Sistema de Gestión de Contenidos (CMS) es un elemento crítico para el éxito de prácticamente cualquier sitio web e *intranet*<sup>19</sup>, pero muchas organizaciones no están al corriente aún de esta tecnología.

Por norma general, un sitio web, así como una *intranet*, va creciendo a lo largo del tiempo, lejos de ser un sistema perfecto. En muchos casos, el contenido es anticuado e incluso inexacto, resulta muy difícil encontrar información concreta, la actualización de los contenidos es compleja y la apariencia se vuelve obsoleta. Peor todavía es que se pierden los cambios en las páginas del sitio además de sobrecargar con continuas tareas de actualización a los administradores de la web. Preguntas como: ¿Cuáles fueron las noticias publicadas el pasado mes? ¿Cuántas modificaciones se realizaron en el año pasado? No podrían ser respondidas. Es una consecuencia directa de la utilización de productos como *Dreamweaver* o *Frontpage*, ya que estas herramientas están más enfocadas a la creación que al mantenimiento de un sitio web.

Un Sistema de Gestión de Contenidos se diseña específicamente para solucionar estos y otros muchos problemas, así como para ofrecer numerosas ventajas en la ardua tarea de mantener un sitio web en un entorno distribuido y de colaboración.

El **objetivo principal** de un CMS es soportar la creación, mantenimiento, distribución y publicación de cualquier tipo de información deseada por una empresa. Cubre el ciclo de vida completo de una página del sitio web, desde la disposición de herramientas sencillas para la creación y modificación, pasando por la publicación y finalmente por el archivo. También proporciona la capacidad de manejar la estructura del sitio, el aspecto de las páginas publicadas y la navegación ofrecida a los usuarios. Los sistemas de gestión de contenidos pueden ser mucho más amplios y complejos, pero éste es el núcleo común de todos ellos.

#### 2.1.2 Definición

Un CMS es un sistema de gestión de contenidos que permite gestionar el ciclo de vida de todo el contenido de un sitio web. Debido a que su punto de actuación se centra fundamentalmente en la gestión de sitios web también se les conoce como gestores de contenido web (WCM, *Web Content Management*).

Técnicamente se trata de un sistema de información situado sobre un servidor web, apoyado habitualmente en uno o varios sistemas de gestión de base de datos.

Desde el punto de vista del usuario, se trata de gestionar, de forma uniforme, accesible y cómoda, un sitio web dinámico, con actualizaciones periódicas, y sobre el que pueden trabajar una o más personas, cada una de las cuales tiene una función determinada.

Desde el punto de vista del cliente, se trata de un sitio web dinámico, con apariencia e interfaz uniforme, con un diseño centrado en el usuario, y que permite llevar a cabo fácilmente las tareas para las que ha sido diseñado.

---

<sup>19</sup> Una intranet es una red de ordenadores de una red de área local (LAN) privada empresarial o educativa que proporciona herramientas de Internet, la cual tiene como función principal proveer lógica de negocios para aplicaciones de captura, reportes, consultas, etc. con el fin de auxiliar la producción de dichos grupos de trabajo; es también un importante medio de difusión de información interna a nivel de grupo de trabajo.

Por tanto, el sistema permite manejar de manera independiente el contenido por una parte y el diseño por otra. Así, es posible manejar el contenido y darle en cualquier momento un diseño distinto al sitio sin tener que darle formato al contenido de nuevo, además de permitir la fácil y controlada publicación en el sitio a varios editores.

### 2.1.3 Beneficios

Existe un amplio abanico de ventajas para el negocio que pueden obtenerse poniendo un CMS en la empresa:

- Proceso de autoría orientado a flujo
- Soporte de autoría distribuida y delegada
- Tiempo de respuesta rápido para la creación y modificación
- Mayor consistencia y flexibilidad
- Navegación más controlada y mejorada
- Incremento de la seguridad
- Reducción de información duplicada
- Gran capacidad de crecimiento y control del mismo
- Reducción de costes en el mantenimiento del sitio

Más allá de todas estas ventajas, el mayor beneficio de un CMS es soportar los objetivos y estrategias del negocio. Por ejemplo, el CMS puede ayudar a mejorar las ventas, aumentar la satisfacción del usuario así como incrementar la comunicación con los mismos.

### 2.1.4 Clasificación

Existe una gran variedad de sistemas de gestión de contenidos pudiéndose agrupar en las siguientes categorías:

- Sistema de tipo foros: Crea un foro de discusión en línea donde la gente se puede reunir y discutir temas en los que están interesados.
- Sistema de tipo *wiki*<sup>20</sup>: Permite al usuario editar una página y guardar los cambios realizados.
- Sistema de gestión de bitácoras o *weblogs*: Permiten gestionar un sitio o sitios web con una serie de características comunes: la organización cronológica y el archivo de las historias antiguas.
- Sistema de tipo portal: Incluyen, aparte de lo anterior, otra serie de características como la gestión de información estática y documentación, encuestas, foros y mensajería.
- Sistema genérico: Permiten gestionar información de cualquier tipo, y son generalmente configurables y ampliables con nuevas funcionalidades.

También nos podemos encontrar otras clasificaciones como por ejemplo atendiendo a se trata de soluciones comerciales o del lado del software libre (Tramullas J., 2005).

---

<sup>20</sup> La palabra *wikiwiki* (o *awiwī*) significa, *rápido* en hawaiano.

### 2.1.5 Objetivos del CMS

Éstos son algunos de los puntos más importantes que hacen útil y necesaria la utilización de un CMS en la organización:

- **Inclusión de nuevas funcionalidades en la web.** Esta operación puede implicar la revisión de multitud de páginas web y la incorporación del código que aporta dichas funcionalidades. Con un CMS eso puede ser tan simple como incluir un módulo realizado por terceros, sin que suponga muchos cambios en la web. El sistema puede crecer y adaptarse a las necesidades futuras.
- **Mantenimiento de gran cantidad de páginas.** En una web extensa hace falta un sistema para distribuir los trabajos de creación, edición y mantenimiento con permisos de acceso a las diferentes áreas. Esta forma de trabajar convierte al CMS en un sistema con funcionalidades distribuidas y que permite delegación de autoría. También se tienen que gestionar los metadatos de cada documento o contenido del sitio, las versiones, la fecha de publicación y caducidad de la información, así como los enlaces rotos<sup>21</sup>.
- **Reutilización de objetos o componentes.** Un CMS permite la recuperación y reutilización de páginas, documentos, y en general de cualquier objeto publicado o almacenado.
- **Páginas interactivas.** Las páginas estáticas llegan al usuario exactamente como están almacenadas en el servidor web. En cambio, las páginas dinámicas no existen en el servidor tal como se reciben en los navegadores, sino que se generan según las peticiones y perfiles de los usuarios. De esta manera cuando por ejemplo se utiliza un buscador, el sistema genera una página con los resultados. Para conseguir esta interacción, los CMS conectan con una base de datos que hace de repositorio central para todos los datos de la web.
- **Cambios del aspecto visual de la web.** Si no hay una buena separación entre contenido y presentación, un cambio de diseño puede provocar la revisión de muchas páginas web para su adaptación. Los CMS facilitan los cambios con la presentación, por ejemplo, a través de la incorporación de estándares como el CSS (*Cascading Style Sheets* u hojas de estilo en cascada). De esta manera se consigue una independencia entre presentación y contenido.
- **Consistencia de la web.** La consistencia en una web no quiere decir que todas las páginas web sean iguales, sino que hay un orden (visual). Un usuario nota enseguida cuándo una página no es igual que el resto de las de la misma web por su aspecto, la disposición de los objetos o por los cambios en la forma de navegar. Estas diferencias provocan sensación de desorden y dan a entender que la web no ha sido diseñada por profesionales. Los CMS pueden aplicar un mismo estilo en todas las páginas (con el mencionado CSS) y aplicar una misma estructura mediante plantillas.
- **Control de acceso.** Controlar el acceso a una web no consiste simplemente en permitir la entrada a la web, sino que comprende gestionar los diferentes permisos aplicados a grupos o individuos en cada área de la web.

---

<sup>21</sup> Hipervínculos a páginas no existentes en la red.

### 2.1.6 Aplicación en el e-Learning<sup>22</sup>

El *e-Learning* tiene unas necesidades específicas que un CMS general no siempre cubre, o si lo hace, no da las mismas facilidades que una herramienta creada específicamente para esta función.

En general, las herramientas que dan soporte a esta estrategia de educación, conocidas como sistemas de gestión del aprendizaje (*Learning Management Systems* o LMS) son, por un lado, utilidades para la presentación de contenidos (textos, documentos, animaciones, imágenes, etc.) y por otro, facilitan la comunicación entre alumnos y entre los tutores y alumnos. Además, aportan herramientas para el seguimiento y valoración de los estudiantes. Es decir, facilitan una translación del modelo real en el mundo virtual.

Un buen ejemplo de plataforma de tele-formación es el sistema de gestión de cursos con licencia de código abierto denominado *Moodle*. Estos sistemas son diferentes a los CMS, tanto por el objetivo final como por las características. No obstante comienzan a incluir capacidades de los sistemas de gestión de contenidos.

Con la integración de las dos herramientas nace un nuevo concepto, los sistemas de gestión de contenidos para el aprendizaje (*Learning Content Management Systems*, LCMS).

Son muchas las ventajas que se obtienen de la incorporación de este tipo de herramientas en la educación:

- El uso de las TIC para la mejora de la actividad docente.
- Técnicas para producir material complementario en línea optimizando la formación y rendimiento académico de los estudiantes.
- Un acceso más rápido y eficaz de docentes y estudiantes a la información, reduciendo el grado de obsolescencia a la información que consume, y aprovechando y utilizando de una forma más eficiente los recursos de información que los organismos ponen a su alcance.
- Formación adecuada al profesorado en aspectos relacionados con la enseñanza virtual, en general, y en la adecuada estructuración de los contenidos para su digitalización, en particular.

## **2.2 Descripción funcional de un CMS**

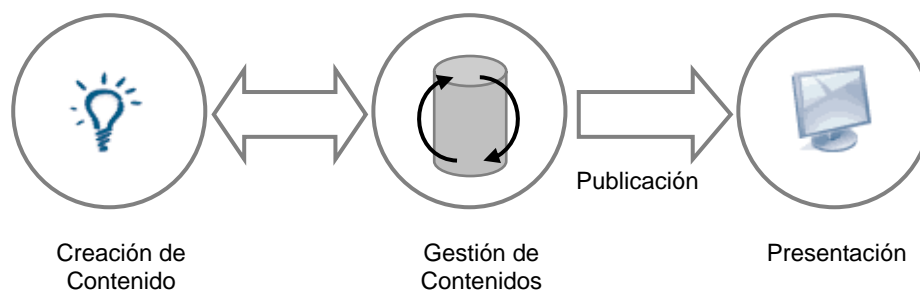
James Robertson (Robertson J., 1998) propone una división de la funcionalidad de los sistemas de gestión de contenidos en cuatro categorías:

- creación de contenido,
- gestión de contenido,
- publicación
- y presentación.

---

<sup>22</sup> Aprendizaje asistido por tecnologías de la información





**Ilustración 7. Funcionalidad de un CMS**

### 2.2.1 Creación de contenido

Un CMS aporta herramientas para que los creadores sin conocimientos técnicos en páginas web puedan concentrarse en el contenido. Lo más habitual es proporcionar un editor de texto WYSIWYG, en el que el usuario ve el resultado final mientras escribe, al estilo de los editores comerciales, pero con un rango de formatos de texto limitado. Esta limitación tiene sentido, ya que el objetivo es que el creador pueda poner énfasis en algunos puntos, pero sin modificar mucho el estilo general del sitio web.

Hay otras herramientas como la edición de los documentos en XML, utilización de aplicaciones ofimáticas con las que se integra el CMS, importación de documentos existentes y editores que permiten añadir marcas, habitualmente HTML, para indicar el formato y estructura de un documento.

Un CMS puede incorporar una o varias de estas herramientas, pero siempre tendría que proporcionar un editor WYSIWYG por su facilidad de uso y la comodidad de acceso desde cualquier ordenador con un navegador y acceso a Internet.

Para la creación del sitio propiamente dicho, los CMS aportan herramientas para definir la estructura, el formato de las páginas, el aspecto visual, uso de patrones, y un sistema modular que permite incluir funciones no previstas originalmente.

### 2.2.2 Gestión de contenido

Los documentos creados se depositan en una base de datos central donde también se guardan el resto de datos de la web, cómo son los datos relativos a los documentos (versiones hechas, autor, fecha de publicación y caducidad, etc.), datos y preferencias de los usuarios, la estructura de la web, etc.

La estructura de la web se puede configurar con una herramienta que, habitualmente, presenta una visión jerárquica del sitio y permite modificaciones. Mediante esta estructura se puede asignar un grupo a cada área, con responsables, editores, autores y usuarios con diferentes permisos. Eso es imprescindible para facilitar el ciclo de trabajo (*workflow*) con un circuito de edición que va desde el autor hasta el responsable final de la publicación. El CMS permite la comunicación entre los miembros del grupo y hace un seguimiento del estado de cada paso del ciclo de trabajo.

### 2.2.3 Publicación

Una página aprobada se publica automáticamente cuando llega la fecha de publicación, y cuando caduca se archiva para futuras referencias. En su publicación se aplica el patrón definido para toda la web o para la sección concreta donde está situada, de forma que el resultado final es

un sitio web con un aspecto consistente en todas sus páginas. Esta separación entre contenido y forma permite que se pueda modificar el aspecto visual de un sitio web sin afectar a los documentos ya creados y libera a los autores de preocuparse por el diseño final de sus páginas.

#### 2.2.4 Presentación

Un CMS puede gestionar automáticamente la accesibilidad del web, con soporte de normas internacionales de accesibilidad como WAI, y adaptarse a las preferencias o necesidades de cada usuario. También puede proporcionar compatibilidad con los diferentes navegadores disponibles en todas las plataformas (Windows, Linux, Mac, Palm, etc.) y su capacidad de internacionalización lo permite adaptarse al idioma, sistema de medidas y cultura del visitante.

El sistema se encarga de gestionar muchos otros aspectos como son los menús de navegación o la jerarquía de la página actual dentro del web, añadiendo enlaces de forma automática. También gestiona todos los módulos, internos o externos, que incorpore al sistema. Así por ejemplo, con un módulo de noticias se presentarían las novedades aparecidas en otro web, con un módulo de publicidad se mostraría un anuncio o mensaje animado, y con un módulo de foro se podría mostrar, en la página principal, el título de los últimos mensajes recibidos. Todo eso con los enlaces correspondientes y, evidentemente, siguiendo el patrón que los diseñadores hayan creado.

### 3 MyCMS v1.0: una experiencia en la ESO<sup>23</sup>

Desde el *Grupo de Investigación en Interacción y e-Learning* (GRIAL) de la Universidad de Salamanca (<http://grial.usal.es>), se inicia un proyecto innovador con la idea de incorporar las ventajas de la gestión de contenidos y del *e-Learning* a la Enseñanza Secundaria Obligatoria, con el fin de obtener una educación de mayor calidad.

Tras una etapa previa de análisis y estudio sobre las características fundamentales de los actuales gestores de contenido de software libre y comerciales, el proyecto culmina con la construcción e implantación de una herramienta para la gestión de contenidos, denominada *MyCMS v1.0* (*My Content Management System*) en el **Instituto de Enseñanza Secundaria Fray Luis de León** de la ciudad de Salamanca.

La idea fundamental del proyecto es que MyCMS haga las funciones de *intranet* o portal interno de comunicación y difusión de información entre alumnos, profesores y resto del personal vinculado con el centro de enseñanza. De esta forma, en general, gran parte de las disciplinas de la enseñanza secundaria se verán notablemente reforzadas, y en particular una de ellas, la educación a distancia.

#### 3.1 **Conceptos previos**

Según James Robertson (2007, abril) para que la *intranet* sea realmente eficaz, deberá centrarse entorno a tres objetivos fundamentales y sobre los cuales el CMS deberá ofrecer una gestión completa:

- Contenido
- Comunicación
- Actividad

---

<sup>23</sup> Enseñanza Secundaria Obligatoria

MyCMS pretende convertirse en un gestor de contenidos con capacidad para gestionar la *intranet* a estos tres niveles. No obstante, podemos ofrecer ésta y otras muchas herramientas similares, pues sin la complicidad del propio personal del centro el proyecto fracasará y todos los esfuerzos en la implantación estos sistemas de colaboración habrán sido en vano.

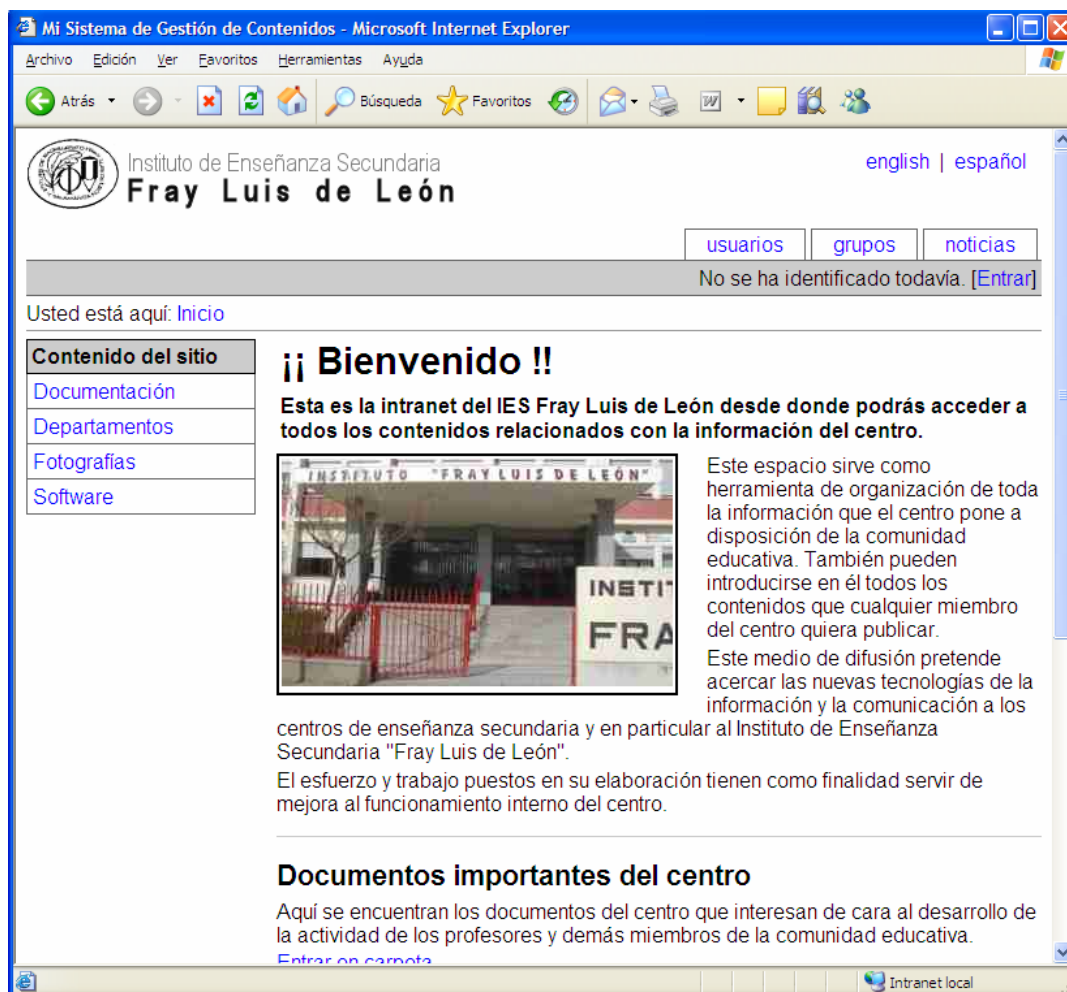


Ilustración 8. MyCMS

### 3.1.1 Contenido

El papel de la *intranet* como repositorio de información es claro. Disciplinas como la usabilidad (Nielsen, J., 1999) y la arquitectura de la información han contribuido a mejorar las capacidades de la *intranet* para ofrecer sitios mejor estructurados y en los que localizar información concreta resulta una tarea sencilla.

El esfuerzo depositado se centra fundamentalmente en la gestión de la *intranet* considerada como un repositorio de información, que se sustenta sobre un sistema de gestión de contenidos en el que se establecen diferentes estrategias de delegación de autoría sobre los contenidos de la misma.

Mientras que la *intranet* juega un papel fundamental en el desarrollo de contenido para la corporación, éste no es el aspecto fundamental que contribuye al éxito del sitio. Si concebimos la *intranet* como "un lugar al que ir a leer cosas", ésta sólo será consultada cuando se necesite, normalmente cada semana, mes o trimestre. Lo que realmente contribuye al éxito, desde el

punto de vista del contenido, es la veracidad, exactitud y continua actualización de la información disponible en línea así como de la archivada. Es precisamente esta gestión del contenido conseguida con los CMS la que contribuyen al éxito de la *intranet*.

### 3.1.2 Comunicación

La *intranet* puede alcanzar también un papel importante como mecanismo de comunicación hacia y entre sus usuarios. Esto se ve tipificado, por ejemplo, a través de la sección de *Noticias* localizada en la página principal del sitio utilizada para comunicar noticias clave para el organismo.

Sin embargo, la *intranet*, no es un medio de empuje eficaz considerada como un simple canal de comunicación de noticias. En otras palabras, el personal no viene a la intranet solamente para comprobar si se han añadido noticias nuevas. El éxito de la *intranet* pasa por la necesidad de que sus usuarios la consideren también una herramienta útil. Esto, en la mayor parte de los casos, se refleja en la disponibilidad de otras capacidades relevantes para los usuarios como por ejemplo: la consulta de calificaciones, las faltas de asistencia, la consulta de la nómina, los permisos laborales, etc.

En definitiva, sin una razón clara para ir al sitio, la intranet no triunfa y por lo tanto fracasa en el objetivo de difusión de información.

### 3.1.3 Actividad

Más allá de ver la *intranet* como un almacén de documentos, también puede ser un sitio en el que se pueda operar y realizar *actividades* que contribuyan a la realización de las tareas cotidianas del personal.

Algunas de estas tareas pueden ser:

- Aplicaciones en línea para auto-servicios de Recursos Humanos (datos personales, permisos vacacionales, datos fiscales, etc.).
- Aplicación en línea para auto-matrícula.
- Informes y formularios en línea personalizados.
- Herramientas para la formación en línea.
- Otras herramientas de colaboración.

Estos instrumentos permiten al personal completar parte de su trabajo cotidiano a través del sitio y cambian así la concepción de la *intranet* como un simple sitio para la consulta de noticias.

## 3.2 **Características de MyCMS v1.0**

La idea fundamental de todo sistema de gestión de contenidos es que el usuario de lo único que debe preocuparse es de la información que desea publicar, ya que el sistema gestionará todos los detalles técnicos y administrativos necesarios.

### 3.2.1 Características generales del sistema

- **Organización del sitio web:** MyCMS está preparado para organizar eficientemente los contenidos de su sitio en secciones y categorías, lo que facilita la navegabilidad para los usuarios y permite crear una estructura sólida, ordenada y

sencilla. Desde el gestor de contenidos de MyCMS usted podrá crear, editar y borrar las secciones y categorías de su sitio de una manera sencilla y práctica.

- **Arquitectura de la información:** MyCMS estructura el contenido en base a dos tipos básicos de información: *elementos sencillos* (contenidos básicos de información) y *elementos contenedor* (contenidos con información propia y que pueden englobar otros elementos de información: sencillos y contenedor). Dada la importancia de esta característica más adelante hablaremos en detalle.
- **Publicación y gestión de los contenidos:** Con MyCMS podrá crear secciones y categorías ilimitadas y editarlas desde un sencillo editor que permite formatear los textos con los estilos e imágenes deseados. Los contenidos son totalmente modificables.



**Ilustración 9. Publicación y gestión de los contenidos**

- **Administración de usuarios y grupos:** MyCMS le permite almacenar datos de usuarios registrados. La administración de usuarios es jerárquica, y los distintos grupos de usuarios poseen diferentes niveles de facultades/permisos dentro de la gestión y administración del sitio.
- **Diseño y aspecto de la interfaz del sitio:** Es posible cambiar todo el aspecto del sitio web fácilmente a través del panel del Administrador del sitio, gracias al sistema de plantillas CSS disponibles en MyCMS.
- **Sistema de navegación y de menús sencillo:** Totalmente modificables desde el gestor de contenidos de MyCMS.
- **WAI-AA:** Nivel Doble-A de Conformidad con las Directrices de Accesibilidad para el Contenido Web 1.0 del W3C<sup>24</sup>.
- **Internacionalización:** Soporte a múltiples idiomas.

<sup>24</sup> El Consorcio World Wide Web (W3C) desarrolla tecnologías inter-operativas (especificaciones, líneas maestras, software y herramientas) para guiar la Red a su potencialidad máxima a modo de foro de información, comercio, comunicación y conocimiento colectivo.

### 3.2.2 Características de la publicación de contenidos

- **Automatización en la publicación:** Los contenidos de MyCMS pueden programarse con fecha de publicación y de caducidad. De esta forma, un contenido de información puede programarse para que se publique automáticamente al llegar una determinada fecha, y más tarde ocultarse también de forma automática en la fecha programada.

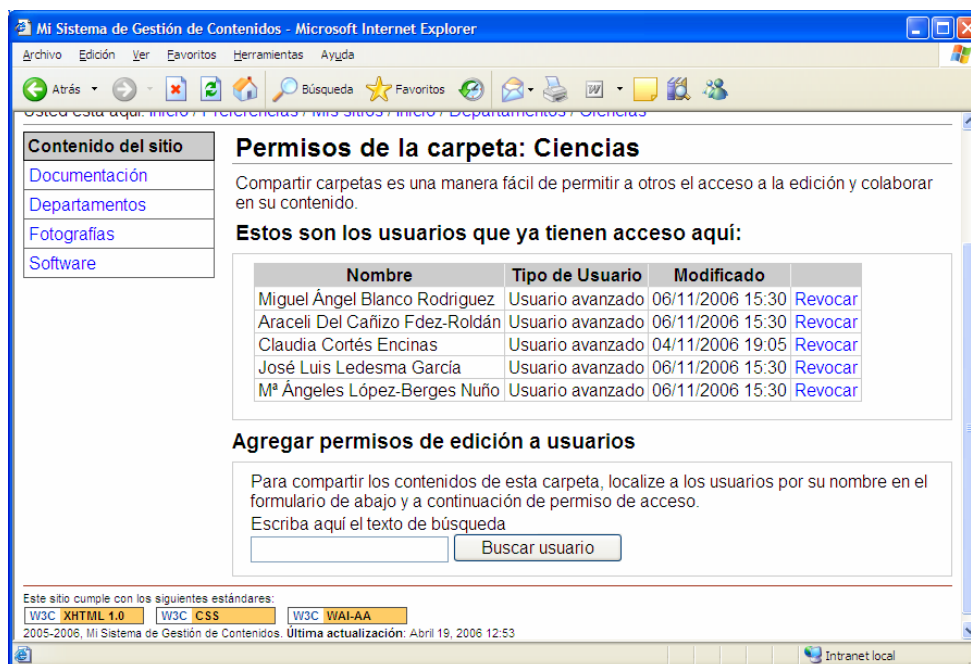


Ilustración 10. Automatización de la publicación

- **Delegación de autoridad:** Con MyCMS usted podrá delegar autoridad, en la gestión de las secciones que desee, sobre alguno de los usuarios registrados en el sistema. A través de esta característica evitamos la sobrecarga de trabajo del administrador en la publicación de contenidos y dejamos dicha tarea realmente en el/los propietario/s de la información.
- **Control de versiones:** Todas las modificaciones realizadas sobre los contenidos publicados generan una nueva versión del contenido que queda guardada con la fecha completa y autor que realiza la modificación.
- **Niveles de visibilidad:** Los contenidos en MyCMS se publican con un nivel de visibilidad acorde con la comunidad de usuarios hacia la que va enfocada. En este sentido MyCMS define cuatro grados de visibilidad: ninguna (el contenido no se muestra), baja, media y alta. Estos niveles de visibilidad están asociados a los perfiles de usuario: sin registrar, alumno, profesor y administrador respectivamente.
- **Archivo:** Los contenidos caducados o publicaciones que hayan perdido vigencia siguen formando parte del sistema sin necesidad de tener que borrarlos. Esto permite también a los gestores de la posibilidad de poder volver a disponer de ellos en cualquier momento favoreciendo así la reutilización de contenidos.

### 3.2.3 Arquitectura de la información

El término "Arquitectura de la Información" (AI) fue utilizado por primera vez por Richard Saul Wurman en 1975, quién la define como: “*El estudio de la organización de la información con el objetivo de permitir al usuario encontrar su vía de navegación hacia el conocimiento y la comprensión de la información*”.

Si nos ceñimos exclusivamente a la AI en el campo de la Web, una de las definiciones que Louis Rosenfeld y Peter Morville ofrecen en su libro "Information Architecture for the World Wide Web" (Rosenfeld, L., y Morville, P., 1998), puede darnos de más fácil comprensión: “*La AI se enfoca sobre el diseño de la organización, los mecanismos de indexación y catalogación y el sistema de navegación a fin de facilitar la navegación y búsqueda de información a los usuarios*”.

El concepto "Arquitectura de la Información" no sólo engloba la actividad de organizar información, sino también el resultado de dicha actividad. Por ello, es muy importante diseñar un mecanismo ágil y sencillo para organizar la información en los CMS y facilitar la navegación y las formas de encontrar fácilmente la información solicitada por los usuarios.

El diseño de la estructura y mecanismo de navegación están estrechamente relacionados con los elementos de información disponibles en el gestor de contenidos. Así pues, cuando un editor desee publicar alguna información, escogerá del gestor de contenidos el elemento de información que mejor se ajuste o sirva para representar a dicha información. Cualquier elemento de información se puede incluir en cualquier parte del mapa de navegación del sitio web.

En MyCMS todos los objetos de información o tipos de contenido disponen de los siguientes atributos: *oid* (identificador del objeto), *versión* (número de cambios realizados sobre el objeto), *título largo*, *título corto*, *cabecera*, *cuerpo*, *fecha de comienzo*, *fecha de caducidad*, *visibilidad*, *autor* y *fecha de la versión*.

MyCMS propone cinco tipos de contenido para representar y publicar cualquier información en el sitio con la apariencia de una página web. Estos cinco tipos de contenido son:

- **Artículo:** posee información descriptiva de un escrito, pero además permite anexar otros tipos de contenido del tipo *archivo* o *imagen*. Los contenidos de este tipo podrán ser mostrados (si el editor lo juzga oportuno) en la sección de noticias del sitio web.
- **Archivo:** posee información descriptiva de un documento y un hipervínculo a la localización exacta del fichero para su posterior descarga.
- **Imagen:** posee información descriptiva de una imagen y un hipervínculo a la localización exacta de la imagen para su visualización o posterior descarga.
- **Carpeta:** posee información descriptiva del contenido de una carpeta y, además, permite agregar cualquier otro tipo de contenido (inclusive otras *carpetas*).
- **Galería:** es un caso particular del tipo *carpeta*, a la que sólo se pueden agregar *imágenes*, con el fin de dar una apariencia especial a la visualización del objeto en cuestión.

### 3.2.4 Estructura de navegación: el objeto *carpeta*

La estructura de navegación utilizada por MyCMS hace que la información se organice de forma arborescente. Todos los nodos (independientemente del tipo) contienen información que será mostrada en forma de página web.

Podemos diferenciar dos tipos de nodos:

- **Nodos contenedor** (*carpetas, galería, artículo*) que nos permiten acceder a un nuevo menú o rama del árbol en donde podemos encontrar (o no) otros nodos hijo que cuelgan de él.
- **Nodos hoja** (*archivo, imagen*) que hacen de nodos terminales, es decir, nodos de los cuales no cuelga ningún otro nodo.

Desde el punto de vista semántico, podemos pensar que el objeto *carpeta* permitirá entrar en él y mostrar los objetos que engloba y así sucesivamente con otros objetos *carpeta* que se encuentren en su interior. De esta manera iremos explorando todo el esquema o mapa de navegación creado para el sitio web. Sin embargo, esto mismo no ocurre con los objetos *galería* y *artículo*, ya que están limitados a contener únicamente *archivos* o *imágenes*.

Para llevar a buen término este esquema conceptual de navegación, en MyCMS todos los objetos disponen de un atributo adicional obligatorio denominado *oid\_parent*. Este atributo hace referencia al objeto o nodo padre del cual cuelgan dentro del mapa de navegación del sitio web.

MyCMS muestra el mapa de navegación en todo momento de dos formas:

- Textualmente en una barra de situación localizada en la cabecera de toda página web que puede identificarse a través del literal: “*Usted está aquí: ...*”. Esta barra únicamente mostrará la ruta completa para llegar al nodo actual.
- Y gráficamente a través de un menú de opciones en el margen izquierdo. Este menú desplegará el árbol completo hasta el nodo actual.



Ilustración 11. Usuarios



Como se puede observar, se trata de un mecanismo de navegación que en muchos aspectos nos recuerda al comportamiento básico del protocolo *gopher*<sup>25</sup>, un servicio de Internet también consistente en el acceso a la información a través de menús y considerado como el predecesor de la Web.

### 3.3 Arquitectura de MyCMS

MyCMS, como todo sistema de gestión de contenidos web, es una aplicación apoyada en la tecnología de Internet y como tal, resulta apropiado presentarla dentro del marco de las aplicaciones web multi-capa modernas. Damos pues, un salto en las aplicaciones basadas en arquitecturas cliente-servidor propias de la familia de los protocolos de Internet<sup>26</sup>.

Por norma general, el usuario interactúa con las aplicaciones web a través del navegador. Como consecuencia de la actividad del usuario, se envían peticiones al servidor, donde se aloja la aplicación, que normalmente hace uso de una base de datos en la que almacena toda la información. El servidor procesa la petición y devuelve la respuesta al navegador que se encarga de presentar los resultados al usuario. En la mayoría de los casos, el navegador suele limitarse a presentar los datos (cliente *ligero*) y no lleva a cabo ninguna tarea de la lógica de negocio. No obstante, es común la utilización de tecnologías como los *applets*<sup>27</sup>, lenguajes interpretados o DHTML<sup>28</sup> para mejorar la presentación e interacción con el usuario. Esto sitúa al cliente en un punto intermedio entre cliente *ligero* y *pesado*. En el caso concreto de MyCMS el usuario trabajará con este tipo de clientes intermedios para lo cual se apoya en algunos de estos estándares.

#### 3.3.1 Arquitectura multi-capa distribuida

La arquitectura de las aplicaciones web modernas, y en particular MyCMS, presenta un esquema de tres capas en el que se separa la lógica de presentación de la lógica de negocio. Estos niveles son:

- La *capa cliente*: representa la interfaz de la aplicación de cara al usuario. Está en contacto directo con el usuario, quien a través del navegador, envía las acciones al servidor y recibe los resultados.
- La *capa servidor* o *de aplicación*: contiene toda la lógica que modela los procesos de negocio y es donde se realiza todo el procesamiento necesario para atender las peticiones de los usuarios. Por norma general está representada por un servidor web<sup>29</sup> y un servidor de aplicaciones<sup>30</sup> con el código de negocio.

---

<sup>25</sup> El protocolo *gopher* está definido en la RFC1436.

<sup>26</sup> También conocido como *conjunto de protocolos TCP/IP* (dos de los protocolos más importantes de que componen la arquitectura de Internet: *TCP*, Protocolo de Control de Transmisión e *IP*, Protocolo Internet).

<sup>27</sup> Programa escrito en el lenguaje Java, con la característica que normalmente se ejecuta en el contexto de otro programa, en este caso, dentro del propio navegador.

<sup>28</sup> HTML Dinámico (del inglés *Dynamic HTML*), conjunto de técnicas que permiten crear sitios web interactivos utilizando una combinación del lenguaje HTML estático, un lenguaje interpretado del lado del cliente (como JavaScript), el lenguaje de hojas de estilo en cascada (CSS) y la jerarquía de objetos de un DOM.

<sup>29</sup> El servidor Web, es el programa que implementa la parte del servidor, del protocolo HTTP (*HyperText Transfer Protocol*), dentro de una típica arquitectura C-S. Su función elemental consiste en atender las peticiones en este caso del cliente o navegador Web.

<sup>30</sup> El servidor de aplicaciones es el dispositivo software que contiene los programas de los clientes.

- La *capa de datos*: garantiza la persistencia, el suministro y el almacenamiento de la información al nivel de aplicación. Normalmente está representada por un Sistema de Base de Datos. En MyCMS además se ha añadido una fuente de datos externa para el almacenamiento de archivos con el fin de no sobrecargar excesivamente al SGBD.

Esta separación hace que la lógica de negocio sea independiente de cómo y dónde se presenta. De esta forma, la lógica de negocio (en la capa central) no necesita conocer qué tipo de cliente muestra los datos. Las tres capas permiten una alta portabilidad, además de balancear las solicitudes de los clientes entre varios servidores. La seguridad es más fácil de implementar al estar el código de negocio y los datos fuera del cliente y las futuras actualizaciones del software son mucho más rápidas ya que implican únicamente la actualización de la capa de aplicación.

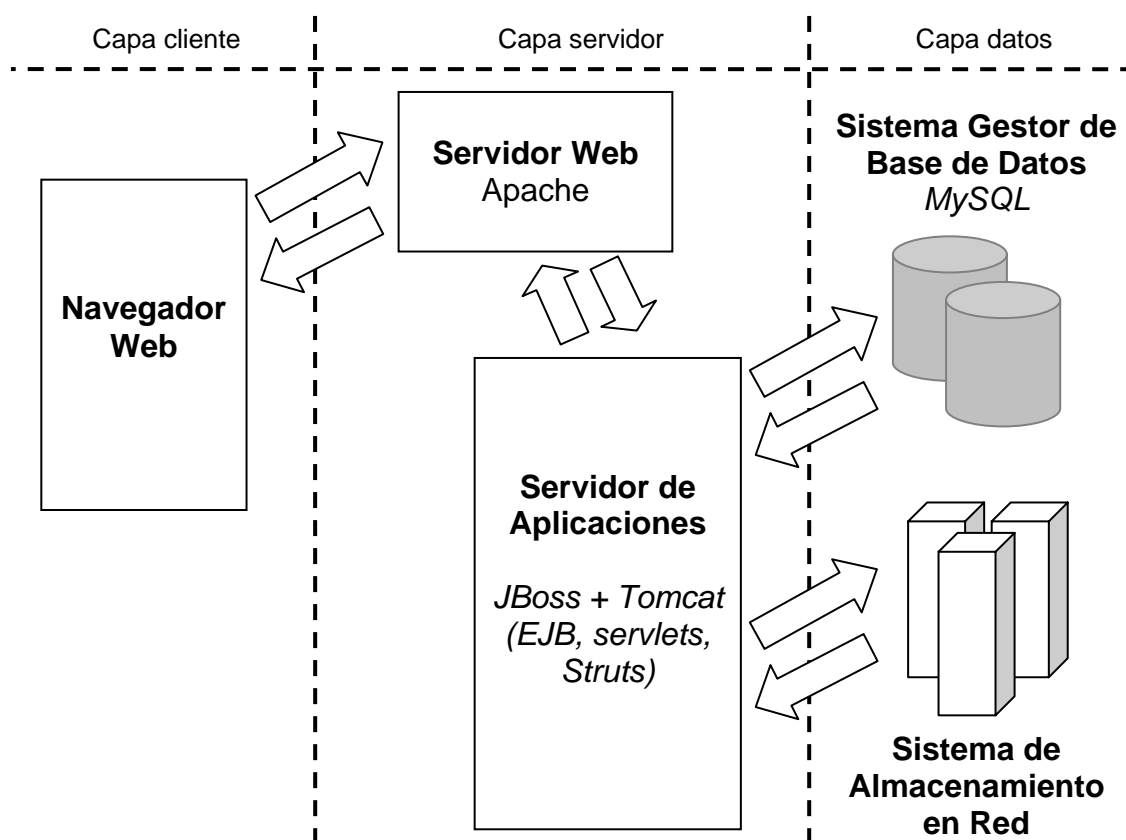


Ilustración 12. Arquitectura de MyCMS

En lo referente a la capa cliente, MyCMS dispone de un nivel de presentación perfectamente adaptado a los navegadores Web actuales. Así pues, productos como Mozilla e Internet Explorer 6.0 son totalmente compatibles con el nivel de presentación de MyCMS. En lo que afecta a la capa de servidor, MyCMS está implementado íntegramente bajo el conjunto de tecnologías soportadas por la arquitectura J2EE<sup>31</sup>. Recomienda el uso de JBoss Application Server como servidor de aplicaciones y Tomcat 5.5 (también llamado Jakarta Tomcat o Apache Tomcat) como contenedor de servlets, desarrollado bajo el proyecto Jakarta en la Apache

<sup>31</sup> Acrónimo de Java 2 Enterprise Edition, plataforma de programación, para desarrollar y ejecutar software de aplicaciones escritas en el lenguaje de programación Java.

Software Foundation. Finalmente, en la capa de datos, la propuesta es utilizar como sistema gestor de base de datos indistintamente MySQL ó PostgreSQL en sus versiones más recientes. Por otro lado para el sistema de almacenamiento en red se propone una solución de bajo coste basada en el almacenamiento del propio sistema operativo, no obstante existen soluciones de mayor envergadura y por supuesto de un coste más elevado, como la implantación un sistema de almacenamiento en red<sup>32</sup> (SAN).

Como se puede observar, el planteamiento de la arquitectura de MyCMS, permite una gran escalabilidad, flexibilidad y portabilidad a cualquier otra plataforma, a la vez que reduce considerablemente los costes y esfuerzos empleados en la implantación de nuevas versiones del producto.

### 3.3.2 Patrones de diseño

Los patrones del diseño (Alur D., Crupi J. y Malks D., 2001) tratan los problemas del diseño que se repiten y presentan en situaciones particulares, con el fin de proponer soluciones. Por lo tanto, son soluciones exitosas a problemas comunes. Existen muchas formas de implementar patrones de diseño. A los detalles de las implementaciones se les denomina estrategias.

Un patrón de diseño es una abstracción de una solución en un nivel alto. Los patrones solucionan problemas que existen en muchos niveles de abstracción. Hay patrones que abarcan las distintas etapas del desarrollo; desde el análisis hasta el diseño y desde la arquitectura hasta la implementación.

#### 3.3.2.1 El Modelo-Vista-Controlador

Uno de los patrones fundamentales a la hora del diseño de MyCMS ha sido el Modelo-Vista-Controlador (en adelante MVC) de tipo 2. Este patrón propone la separación en distintos componentes de la interfaz de usuario (vista), el modelo de negocio y la lógica de control.

En la arquitectura MVC de tipo 2 las diferentes páginas que el usuario ve lanzan acciones utilizando un único controlador, que despacha las diferentes peticiones a un conjunto de acciones previamente registradas en el controlador. Un único *servlet*<sup>33</sup> es llamado desde el cliente, quedando visible un único punto de entrada al controlador.

Esta arquitectura contrasta con el tipo 1, en la cual, la lógica de control de la aplicación (acciones que hay que llamar, vistas que se deben generar y control de navegación entre páginas) va integrada en la capa de presentación.

Para la implementación en MyCMS de la arquitectura MVC de tipo 2, debido a su estrecha relación con el mundo Java, se ha utilizado el *framework* de *Struts* (<http://struts.apache.org>) integrado dentro del conjunto de iniciativas *open source*. Este entorno se nutre de otra serie de tecnologías de programación como: Enterprise JavaBeans<sup>34</sup>, *servlets*, JavaServer Pages (páginas web dinámicas), Tiles (plantillas) y TagLibs (bibliotecas de etiquetas personalizadas) entre otras, que lo convierten en uno de los más potentes del mercado.

---

<sup>32</sup> SAN, del inglés *Storage Area Network*.

<sup>33</sup> Pequeño programa que se ejecuta en el contexto de un contenedor de *servlets* (*Apache Tomcat*) e incluso de un servidor de aplicaciones.

<sup>34</sup> También conocido como EJBs, forma parte del estándar de construcción de aplicaciones J2EE de Sun Microsystems y sirve para el aprovisionamiento de objetos desde el lado del servidor.

### 3.3.2.2 Otros patrones de diseño

Con la aparición de la plataforma J2EE, ha surgido todo un nuevo catálogo de patrones de diseño. En este sentido se plantean diferentes patrones de diseño (Alur D., Crupi J. y Malks D., 2001) en base a tres capas de la arquitectura de aplicaciones J2EE (presentación, negocios e integración).

En MyCMS se han utilizado los siguientes patrones de diseño:

- Capa de negocios
  - *Business Delegate*: Un objeto que reside en la capa de presentación y en beneficio de los otros componentes de la capa de presentación llama a métodos remotos en los objetos de la capa de negocios.
  - *Value Object / Data Transfer Object / Replicate Object*: Un objeto serializable para la transferencia de datos sobre la red.
- Capa de integración
  - *Data Access Object Service Activator*: Consiste en utilizar un objeto de acceso a datos para abstraer y encapsular todos los accesos a la fuente de datos. El DAO maneja la conexión con la fuente de datos para obtener y almacenar datos.

## 3.4 Presente y futuro

En la actualidad, a pesar de todos los esfuerzos empleados desde las direcciones de educación y demás entes gubernamentales, queda mucho por hacer llegar a la enseñanza secundaria todo este conjunto de avances tecnológicos que se suceden continuamente. En algunos casos la falta de formación de personal en las TIC<sup>35</sup>, el rechazo a lo nuevo o simplemente el miedo a lo desconocido dificultan la integración en los centros.

La experiencia realizada con MyCMS v1.0 y su implantación en el I.E.S. Fray Luis de León son un ejemplo más de la necesidad de este tipo de soluciones en este tipo de centros de educación. El resultado ha sido muy bueno, entre otras cosas, porque el grado de implicación y entrega del personal ha sido elevado.

Haciendo un breve balance del proyecto:

- Se ha conseguido implantar una intranet administrativa, con toda la documentación necesaria para el centro estructurada y disponible en línea para toda la comunidad de usuarios.
- Se dispone de un tablón de noticias y de un directorio actualizado de usuarios, grupos.
- Tanto el personal docente (a título individual) como los departamentos (en su conjunto) disponen de un espacio virtual en el que pueden publicar toda la información vinculada con su docencia (temarios, bibliografía, ejercicios, enlaces de interés, etc.).
- En este sentido se delega autoridad de edición sobre zonas del sitio web perfectamente definidas y se distribuye la carga de creación y publicación de contenidos.

---

<sup>35</sup> TIC, acrónimo de Tecnologías de la Información y de la Comunicación.

- A su vez MyCMS permite la recogida de trabajos de los alumnos facilitando así una comunicación más directa y bidireccional con sus profesores o tutores.

Desde el punto de vista técnico y funcional, los CMS amplían sus capacidades por medio de la incorporación de estándares que mejoran la compatibilidad de componentes, facilitan el aprendizaje al cambiar de sistema y aportan calidad, escalabilidad y estabilidad.

Algunos de estos estándares son el CSS, que permite la creación de hojas de estilo; XML, un lenguaje de marcas que permite estructurar un documento; XHTML, que es un subconjunto del anterior orientado a la presentación de documentos vía web; WAI, que asegura la accesibilidad del sistema; y RSS, que permite syndicar contenidos del tipo noticia.

También las aplicaciones que rodean a los CMS tienden a convertirse en estándar (de facto), como los servidores web *Apache* e *ISS*; los lenguajes de programación *PHP*, *Perl* y *Python*; y los sistemas gestores de bases de datos *MySQL* y *PostgreSQL*. La disponibilidad para los principales sistemas operativos de estas aplicaciones y módulos, permite que los CMS puedan funcionar en diversas plataformas sin muchas modificaciones.

Sobre el futuro de los CMS, Robertson (2003a, enero) apunta que:

- Los CMS se convertirán en un artículo de consumo, cuando los productos se hayan establecido y más soluciones lleguen al mercado. Eso provocará una disminución de los precios en los productos comerciales y una mayor consistencia en las funcionalidades que ofrecen.
- En este entorno, muchas empresas que implementan sitios web tendrán que cerrar.
- Muchos proyectos fracasarán por no ajustarse a los estándares y no entender conceptos como usabilidad o accesibilidad, arquitectura de la información, gestión del conocimiento y contenido.
- El campo de los gestores de contenido madurará hasta conseguir un alto grado de consistencia y profesionalidad.
- Se adoptarán estándares en el almacenamiento, estructura y gestión del contenido.
- Se producirá una fusión entre gestión de contenidos, gestión de documentos y gestión de registros.

También se puede añadir la incorporación de sistemas de *e-Learning* y gestión del conocimiento, y en los entornos de *intranet* corporativa, la posibilidad de acceder a otras fuentes de datos como por ejemplo sistemas de soporte de decisiones (*Decision Support Systems*, DSS).

## 4 Referencias

Díaz Hermoso L. M (2001, junio). Estrategia de informatización de la Biblioteca de Ciencias Económicas y Empresariales. Disponible en: <http://www.um.es/cugio/tesis/lianaa.pdf>.

Sharon T, Frank A. J. (2000). Bibliotecas digitales en Internet. 66th IFLA Council and General Conference. Israel: 2000. Disponible en: <http://www.ifla.org/IV/ifla66/papers/029-142s.htm> [2006, noviembre].

Saffady W. (1995). Digital library concepts and technologies for the management of library collections: an analysis of methods and costs. *Library Technology Reports* 1995;31(3):223-238.

Edward A. Fox (1993). *Source Book on Digital Libraries*.

Edward A. Fox (1990-2001). DL Self-Study: Definitions 1999-2001. Disponible en: <http://ei.cs.vt.edu/~dlib/def.htm> [2006, noviembre].

Gary Marchionini and Edward A. Fox (1999, mayo). "Progress toward digital libraries: augmentation through integration", pp. 219-225, guest editors' introduction to "Progress Toward Digital Libraries", eds. Gary Marchionini and Edward A. Fox, Special Issue, *Information Processing & Management*.

Edward A. Fox (1999, julio). According to 5S Framework.

Barrueco J. M. (2003). Open archives initiative. Protocol for Metadata Harvesting (OAI-PMH): descripción, funciones y aplicación de un protocolo. *El Profesional de la Información*, vol. 12, no. 2, pp. 99-106.

Paepcke A., Chen-Chuan K. Chang, García-Molina H., and Winograd T. (1998). Interoperability for Digital Libraries Worldwide. *Communications of the ACM*, 1998/Vol. 41, No. 4.

Hedstrom M. (1998), Digital Preservation: a Time Bomb for Digital Libraries. University of Kentucky [en línea]. Disponible en: <http://www.uky.edu/~kiernan/DL/hedstrom.html> [2006, diciembre]

Lassila O. (1998). Web Metadata: A Matter of Semantic. *IEEE Internet Computing*, July/August 1998, 2, 4, 30-37.

Rust G. (1998, junio). Metadata: the Right Approach. *D-Lib Magazine*: July/August. Disponible en: <http://www.dlib.org/dlib/july98/rust/07rust.html> [2007, enero].

Robertson, J. (2002, abril). Centralised or decentralised authoring?. Step Two [en línea]. Disponible en: [http://www.steptwo.com.au/papers/kmc\\_decentralised/index.html](http://www.steptwo.com.au/papers/kmc_decentralised/index.html) [2007, enero].

Robertson, J. (2002, agosto). What are the goals of a CMS?. Step Two [en línea]. Disponible en: [http://www.steptwo.com.au/papers/kmc\\_goals/index.html](http://www.steptwo.com.au/papers/kmc_goals/index.html) [2007, enero].

Robertson, J. (2003a, enero). Looking towards the future of CM [en línea]. Step Two [en línea]. Disponible en: [http://www.steptwo.com.au/papers/cmb\\_future/index.html](http://www.steptwo.com.au/papers/cmb_future/index.html) [2007, enero].

Robertson, J. (2003b, enero). Putting metadata to work. Step Two [en línea]. Disponible en: [http://www.steptwo.com.au/papers/cmb\\_metadata/index.html](http://www.steptwo.com.au/papers/cmb_metadata/index.html) [2007, enero].

Robertson, J. (2003, junio). So, what is a content management system?. Step Two [en línea]. Disponible en: [http://www.steptwo.com.au/papers/kmc\\_what/index.html](http://www.steptwo.com.au/papers/kmc_what/index.html) [2007, enero]

Robertson, J. (2007, abril). Three fundamental purposes of an intranet. Step Two [en línea]. Disponible en: [http://www.steptwo.com.au/papers/cmb\\_threepurposes/index.html](http://www.steptwo.com.au/papers/cmb_threepurposes/index.html) [2007, enero]

Cuerda X. y Minguillón J. (2004, noviembre). Introducción a los Sistemas de Gestión de Contenidos (CMS) de código abierto. Mosaic [en línea]. Disponible en: <http://mosaic.uoc.edu/articulos/cms1204.html> [2007, enero].

Nielsen, J. (1999). Designing Web Usability: The Practice of Simplicity. New Riders Publishing.

Rosenfeld, L., y Morville, P. (1998), Information Architecture for the World Wide Web (2a. ed.). Sebastopol: O'Reilly.

John Shiple (2007). Information Architecture Tutorial [en línea]. Webmonkey, The Web Developer's Resource. Disponible en: <http://www.webmonkey.com> [2007, enero]

James J. (2002, marzo). A visual vocabulary for describing information architecture and interaction design. "jjg.net" [en línea]. Disponible en: <http://www.jjg.net/ia/visvocab/>

Alur D., Crupi J. y Malks D. (2001). Core J2EE Patterns: Best Practices and Design Strategies (2a. ed.). Upper Saddle River (New Jersey) : Prentice Hall PTR.

Spielman S. (2003). The Struts Framework: Practical Guide for Java Programmers. San Francisco: Morgan Kaufmann Publishers.

Tramullas J. (2005, mayo). Herramientas de software libre para la gestión de contenidos. "Hipertext.net" [en línea]. Núm. 3, 2005. Disponible en: <http://www.hipertext.net> [2006, diciembre]