



INTRODUCCIÓN

El presente trabajo tiene como objetivo principal analizar el estado actual de las colecciones patrimoniales en la red, así como los desafíos que se plantean en el entorno web. Se exploran las estrategias de acceso y normalización que permiten garantizar la preservación y el uso de estos recursos digitales.

M<sup>a</sup> Luisa Alvite Díez

---

**COLECCIONES PATRIMONIALES EN LA RED: ACCESO Y NORMALIZACIÓN EN EL ENTORNO WEB**

En el contexto de la digitalización de los recursos culturales, el acceso y la normalización de las colecciones patrimoniales en la red se convierten en factores clave para garantizar su preservación y su disponibilidad para el usuario. Este artículo aborda los desafíos que plantea el entorno web y propone estrategias para superarlos.

El primer desafío es la fragmentación de los recursos, que dificulta el acceso a la información completa. Para superar este problema, es necesario implementar estrategias de agregación y normalización de los datos, que permitan integrar y organizar los recursos de manera coherente y accesible.

Además, es fundamental garantizar la interoperabilidad de los sistemas de acceso, lo que requiere el uso de estándares internacionales y protocolos de comunicación. Esto permitirá a los usuarios acceder a los recursos desde diferentes dispositivos y plataformas, mejorando así la experiencia de usuario.

Por último, es necesario establecer mecanismos de preservación digital que aseguren la integridad y la disponibilidad de los recursos a largo plazo. Esto implica la implementación de estrategias de backup, migración y emulación, que permitan salvaguardar los recursos frente a los cambios tecnológicos y a la obsolescencia de los formatos.

## INTRODUCCIÓN

Las nuevas Tecnologías de la Información y Comunicaciones (TIC) aplicadas a los contenidos culturales han provocado una transformación drástica en la creación de nuevos productos documentales directamente relacionados con el patrimonio cultural, en su almacenamiento, en la descripción o estructuración de estos objetos informativos y, por último, en su difusión, desprovista de limitaciones temporales o espaciales.

Las instituciones tradicionalmente proveedoras de contenidos: bibliotecas, archivos y museos, actúan en este nuevo espacio de información no sólo como intermediarias sino como creadoras de información con valor añadido. En este contexto, estas entidades culturales han de asignar al objeto informativo una serie de características para facilitar la identificación, localización y recuperación de dicho objeto. Se trata, en última instancia, de priorizar el acceso frente a la propiedad.

La red constituye una fuente de conocimiento inimaginable que nos acerca a la utópica biblioteca universal, y para las instituciones responsables de la conservación y difusión del patrimonio cultural caminar con el paradigma de Internet constituye un elemento sustancial en el cumplimiento de las expectativas y la dinámica de la sociedad del conocimiento. Lo primordial, no obstante, continúa siendo el objeto informativo y no el formato o el medio en que éste se difunde.

En 1994 se creó la *ECPA (European Commission on Preservation and Access)* para actuar como plataforma de estudio y cooperación para las organizaciones culturales en las áreas de la preservación y el acceso, su intención es intensificar el conocimiento sobre estos problemas y subrayar la urgencia de la situación ante la opinión pública. Para la ECPA preservación y acceso son aspectos complementarios y no opuestos en la gestión de las colecciones, se podría hablar de preservación para el acceso.

En este ámbito resulta representativo el programa *Memory of the World* mantenido por la UNESCO con el fin de salvaguardar el patrimonio documental, entendido éste como reflejo de la diversidad de lenguas, pueblos y culturas. Igualmente interesante es el proyecto UNESCO *Guidelines for Establishing Digitization Programmes in Libraries and Archives* en colaboración con la IFLA (*International Federation of Library Associations and Institu-*

tions) y el ICA (*International Council on Archives*), su objetivo es desarrollar una guía que cubra aspectos tales como criterios de selección, métodos y técnicas de conversión, metadatos, aspectos legales, calidad, presupuestos, planificación, necesidades técnicas, desarrollo y mantenimiento de interfaces, etc.

La universalización de la información gracias a las nuevas tecnologías y, en concreto el desarrollo de Internet como medio de difusión y formación, constituye asimismo una clara preocupación para organismos de la importancia de la Comisión Europea, que concibe Internet como un instrumento clave para promover el desarrollo de la sociedad de la información. Entre sus actuaciones cabe citar: GÉANT, Internet de alta velocidad en la línea del Proyecto Internet2 de Estados Unidos, el programa IST (*Information Society Technologies*) encuadrado en el V Programa Marco (1998-2002) de investigación y desarrollo tecnológico en el que se inscriben programas como *Digital Heritage and Cultural Content*, uno de los cinco programas temáticos del V Programa Marco entre cuyas prioridades se halla el asegurar el acceso integrado a las colecciones y materiales de bibliotecas, archivos y museos. De gran trascendencia son programas como *eContent (European Digital Content on Global Networks)*, programa sustituto de INFO2000, dedicado a la financiación de proyectos relativos a la creación, uso y distribución de contenidos en redes globales, *CULTIVATE-EU (Cultural Heritage Application Network European Union)*, nacido con la pretensión de incrementar el nivel de conocimiento del Programa IST para el desarrollo y utilización de las aplicaciones relacionadas con el patrimonio cultural, *PROMISE (Promoting the Information Society in Europe)*, sin olvidar la importancia de proyectos ya finalizados como *Aquarelle (The information Network on Cultural Heritage)*, *EUAN (European Union Archive Network)*, etc. Fuera del Programa Marco se encuentra *ISIS (Information Society Initiative in Standardization)*, iniciativa de la Comisión europea dedicada a reforzar la normalización en el terreno de las tecnologías de la información y comunicaciones, aspecto este fundamental como veremos en los proyectos en curso centrados en la difusión del patrimonio cultural.

En los últimos años, los formatos digitales se han erigido como soporte idóneo no sólo para la preservación de información impresa tradicional, sino también como los medios ideales para lograr la difusión y el acceso a la información de forma ilimitada. Digitalizar una colección no es una mera tarea informática, han de resolverse aspectos tan esenciales como la previa selección de fondos susceptibles de digitalización, los derechos de reproducción y explotación, la conjunción de intereses cien-

tíficos y culturales con la explotación comercial, la obvia necesidad de establecer políticas de cooperación, etc. y, por último, el aspecto en el que centramos nuestro estudio, se precisan herramientas de acceso, herramientas para estructurar y describir la información de forma que esta sea recuperable y, para ello, como requisito previo es precisa la normalización.

En estos momentos son más habituales en la red documentos referenciales que el texto completo de los mismos, si bien, las previsiones conducen hacia un aumento de originales disponibles en la red: bases de datos selectivas de fondos históricos con la imagen digitalizada del documento completo, corpus de textos electrónicos, bibliotecas digitales, archivos virtuales, etc. Su desarrollo supondrá, sin duda, un pilar esencial en la infraestructura de la información dirigida a la investigación y a la comunidad educativa.

Nos proponemos en este trabajo, atendiendo a la unánime consideración de la world wide web como interfaz única, realizar, de un lado, una aproximación a las iniciativas normalizadoras desarrolladas para gestionar y recuperar información relacionada con el patrimonio cultural vía web, y, de otro, recoger proyectos en desarrollo en los que se articulan diversas soluciones para lograr este cometido.

## LOS LENGUAJES DE MARCAS Y LA DISTRIBUCIÓN DE LOS CONTENIDOS

La popularización de la World Wide Web, entendida como colección de recursos multimedia distribuidos, guarda una relación directa con el soporte de los documentos que la integran: el lenguaje de marcas HTML (*Hypertext Markup Language*).

En los últimos años se ha hecho incomprensible el tratamiento de recursos informativos sin la web y es, en este contexto, donde juega un papel esencial la denominada *familia* de normas SGML responsable de garantizar la consistencia en la creación, procesamiento, tratamiento y distribución de documentos, así como en la flexibilidad de la presentación y formato, incidiendo, por tanto, en la mejora del control y la recuperación de la información (Peis; Moysa, 2000).

Un lenguaje de marcado viene determinado por un conjunto de reglas que permiten fijar el tipo de marcas que se utilizarán, las marcas permitidas en cada una de las partes del documento, la forma de distinguir el texto del documento de las marcas y, por último, la gramática y sintaxis que rigen el empleo de las mismas.

Como señalan Martín Galán y Rodríguez Mateos (2000) es posible

utilizar el marcado de documentos con fines documentales permitiendo separar elementos estructurales del documento o realizar abstracciones para extraer metainformación del mismo. Distinguen así dos tipos de lenguajes de marcado: los orientados a la presentación de los documentos y aquellos dirigidos a la descripción formal y de contenido de los documentos.

A finales de los años sesenta se iniciaron las investigaciones que dieron lugar al nacimiento de SGML (*Standard Generalized Markup Language*), metalenguaje de etiquetado de texto convertido en norma ISO 8879 en 1986. Las particularidades de SGML frente a otros lenguajes de marcas pueden cifrarse en que se trata de un lenguaje no propietario, independiente de aplicaciones y plataformas, posibilita la aplicación de un riguroso control sobre el contenido de los documentos, permite la independencia entre contenido, estructura y presentación y, por último, asienta el concepto de DTD (*Document Type Definition*) empleada para definir un tipo de documento estableciendo su estructura, los elementos de los que va a constar, el orden de los mismos, los valores que deben recoger, la posibilidad de ser repetidos y los elementos que pueden contener a otros.

Para Hípola y Eito (2000) SGML es el formato de edición digital que explota con mayor profundidad lo que los autores denominan "opción semántica", es decir, el formato que introduce etiquetas en el contenido del documento no sólo con el fin de distinguir partes significativas y explicitar su estructura sino asignando un valor semántico a cada una de sus partes.

A pesar de la enorme potencialidad que ofrece SGML, en particular, en lo que a recuperación de información se refiere y debido a su complejidad, a los enormes costes que supone su utilización en sistemas de información y al escaso interés que ha despertado en las grandes empresas de software, su uso ha quedado relegado a la publicación, gestión e intercambio de documentos electrónicos en grandes instituciones.

Sin embargo, ha sido HTML, una aplicación del lenguaje SGML desarrollada inicialmente por Tim Berners-Lee, que indica cómo se han de codificar los documentos para su distribución en la web, es decir un sencillo modelo de marcado hipertextual, el lenguaje con mayor presencia en la red y como señalan De la Rosa y Senso (1999) HTML ha supuesto la "socialización de la información", cualquier persona es capaz de poner contenidos en Internet empleando un lenguaje sencillo.

El exponencial desarrollo de la web, convertida en la fundamental herramienta de aprovechamiento de la potencialidad de Internet y las presiones ejercidas por las principales compañías de software hicieron

necesaria la creación del W3C (*World Wide Web Consortium*) encargado de la normalización y el control de los estándares para la web. A pesar de que este organismo logró convertir, a partir de la versión 3.2, en una DTD para los documentos HTML las limitaciones de este lenguaje eran patentes: carencia de sintaxis, escasas posibilidades de estructurar la información, insuficiente potencialidad para la recuperación de la misma, etc.

Se han diseñado alternativas para resolver estos problemas: DHTML (*Dynamic HTML*), HTML-HELP, XML (*eXtensible Markup Language*), XHTML (*eXtensible Hypertext Markup Language*) y, paradójicamente, se ha llegado ya a la versión 4.01 de HTML, si bien para el W3C las posibilidades de HTML parecían agotadas tras HTML 4.0, más aun, el lenguaje HTML se ha convertido en estándar internacional: ISO/IEC 15445:2000.

Paralelamente, sin embargo, algunos estudiosos han acuñado la denominación de segunda generación de la web para referirse a la propiciada por el desarrollo de XML, e incluso se habla de la propia sociedad de la información como germen de XML y es que si HTML supuso el verdadero motor de la web en estos momentos parece ser una rémora para su desarrollo.

XML comenzó su andadura en 1996 respaldado por el W3C con la intención de diseñar un lenguaje de marcas capaz de integrar la simplicidad de HTML con la potencia de SGML. Constituye, por tanto, una versión abreviada de SGML viable en el entorno web, elimina parte de las operaciones sintácticas de SGML pero proporciona estructura a la información. El W3C publicó la versión 1.0 como recomendación en 1998, actualmente contamos con la 2ª edición revisada, recomendación de 6 de octubre de 2000.

Entre las características básicas de XML especialmente interesantes para la gestión de la información, podemos señalar: carácter extensible que posibilita la creación de nuevas marcas atendiendo a los distintos tipos de documentos; orientado a objetos, siguiendo la tendencia actual en el mundo de la programación; incrementa la eficiencia en la interconexión de sistemas dado que facilita la independencia de los datos respecto de las aplicaciones; las etiquetas XML marcan la estructura y la semántica del documento; carácter modular que permite organizar internamente la información como se desee (establecer referencias, guardar partes del documentos en ficheros diferentes); distingue la presentación del documento de la información sobre la semántica y la estructura del mismo (permite indicar cómo se deben formatear, imprimir, visualizar los distintos elementos de un documento a través de especificaciones

XSL (*eXtensible Style Language*); permite una gestión verdaderamente eficaz de los hiperenlaces empleando el estándar XML Linking Language (XPath, Xlink y Xpointer) y, dentro del ámbito de la recuperación de la información, cuenta con especificaciones para facilitar la consulta a grandes colecciones a texto completo: XML-QI (*A Query Language for XML*).

Como señalan De la Rosa y Senso (1999) XML se ha creado con el fin de intercambiar documentos, sean estos referenciales o a texto completo, a través de la web, dado que las opciones existentes no eran viables a este efecto. HTML no permite estructurar la información de forma estricta, de forma que no es posible la recuperación de datos de manera adecuada y SGML es enormemente complejo para su implementación, de ahí que algunos de los estudios más recientes traten de XML/SGML como conceptos inseparables y así son tratados en diversos proyectos que serán expuestos más adelante.

Java, como lenguaje de programación y como plataforma de desarrollo de aplicaciones estándar y aplicaciones web ha sido una de las primeras tecnologías en dar soporte al lenguaje XML, como señala Aladro (2001) ambos lenguajes ofrecen un medio neutral para el desarrollo de aplicaciones que pueden ser ejecutadas en cualquier sistema. A esto hemos de añadir el entusiasmo que XML ha despertado en los máximos representantes del ámbito del software cliente para la web y la compatibilidad de XML tanto con SGML como con HTML todo lo cual ha desembocado en el hecho de que en estos momentos las principales empresas desarrolladoras ya soportan XML: IBM, Oracle, Microsoft, Interleaf, Software AG y existen un gran número de DTDS que emplean este nuevo lenguaje: RDF (*Resource Description Framework*), CDF (*Channel Definition Format*), MathML (*Mathematical Markup Language*), XML News, etc.

En el ámbito de la Ciencia de la Documentación el gran interés de XML reside en la potencialidad de este metalenguaje para describir la estructura lógica y la semántica del contenido de los distintos tipos documentales (sin necesidad de emplear modelos de metadatos), permitiendo el intercambio y recuperación de la información así como la interoperabilidad entre sistemas. El sistema de intercambio de datos más cercano es, sin duda, el formato MARC (*Machine Readable Cataloguing*) basado en la estructuración e intercambio de datos como si se tratase de texto plano, muy lejos de las actuales tendencias que vienen considerando el documento como objeto, de ahí sus limitaciones y escasa flexibilidad.

Para Senso y De la Rosa (1999) una de las disciplinas donde con mayor facilidad puede implementarse XML es la Documentación, teniendo

en cuenta que XML ordena la información en estructuras arbóreas, creando diversas jerarquías para facilitar su gestión y que los productos documentales, en general, presentan una estructura implícita que facilita este tipo de planteamiento. En esta línea, la National Library of Medicine americana ha elegido XML como formato para la distribución de la base de datos MEDLINE, punta de lanza en el mundo documental, creando una DTD específica y planea, asimismo, una versión XML para los Medical Subject Headings; en el terreno de la documentación histórica, trabajos recientes en nuestro país proponen el uso de XML para la identificación, modelado y representación de protocolos notariales como parte integrante de un sistema basado en el conocimiento en el campo de las ciencias históricas, Serrano Tenllado, Calvo y Romero (2001); resulta relevante la existencia en el seno del W3C de un grupo de trabajo denominado XML Query desarrollando especificaciones para la interacción entre el mundo de la web y el de las bases de datos.

## ESTÁNDARES Y PROYECTOS

La proliferación de bibliotecas digitales y, en general, de colecciones de documentos electrónicos ha hecho ostensible el grave problema de la localización y recuperación de este tipo de información en la web. Con el fin de articular soluciones que permitan la gestión y recuperación eficaz de los contenidos distribuidos a través de esta malla multimedia se han asentado dos líneas de investigación relacionadas con el desarrollo de modelos normalizados de aplicación en la web: de un lado, modelos de descripción y organización de los contenidos informacionales, metadatos o "datos estructurados sobre los datos que contienen los documentos" y, de otro, la introducción de metainformación directamente en documentos XML, intercalando etiquetas para asignar valores semánticos.

Consideramos los modelos de metadatos, como señalan Méndez y Merlo (2000) no como normas en el sentido amplio sino como proyectos de desarrollos de normas o estándares de facto para comunidades determinadas de usuarios. En el ámbito de las colecciones patrimoniales distribuidas vía web los modelos más significativos, si bien existen numerosas iniciativas más, son:

- *TEI (Text Encoding Initiative)* Proyecto internacional destinado al desarrollo de pautas para la preparación y el intercambio de textos electrónicos, fundamentalmente literarios. El proyecto se puso en marcha en 1987 como una investigación corporativa patrocinada por *ACH (Association for computers and the Humanities, ACL (Associa-*

*tion for Computational Linguistics*) y *ALLC (Association for Literary and Linguistic Computing)*. El mayor soporte al proyecto procede de NEH (U.S. National Endowment for the Humanities), la Comisión europea, la Andrew Mellon Foundation y el Social Science and Humanities Research Council de Canadá. En diciembre de 2000 se estableció un nuevo Consorcio para mantener y desarrollar el TEI. Se trata de una DTD de SGML con formidables facultades para la indización y el intercambio de información textual. A pesar de que en su origen se concibió como un estándar para proyectos lingüísticos sirve para estructurar todo tipo de textos electrónicos. En estos momentos está en proceso de revisión el texto P3 para asegurar el soporte en XML. Diversas instituciones han puesto en marcha proyectos empleando este estándar: *NOLA: Network of Literary Archives, Proyecto Perseo, Bodleian Library/Toyota City Imaging Project*, de la Universidad de Oxford, etc

- *EAD (Encoded Archival Description)*. Se gestó como proyecto de investigación de la Universidad de California en Berkeley en 1993. Se trata de una DTD SGML/XML mantenida en la actualidad por la Library of Congress de Estados Unidos y cuya responsabilidad recae en la Society of American Archivists. Su uso se ha generalizado para la descripción de documentos de archivo, con la posibilidad de que ésta se halle enlazada a su correspondiente imagen digital. Generalizando, podemos decir que EAD describe el contenido, la estructura y el contexto de una colección. La internacionalización de EAD se debe, por un lado, a la adaptación de su estructura a los elementos de ISAD(G), por otro, a la colaboración de los equipos de trabajo de EAD y TEI para procurar que EAD sea lo más cercano a TEI y, por último, la posibilidad de generar automáticamente un registro MARC a partir de EAD e importar un registro MARC a un instrumento de descripción de EAD. Entre otros proyectos en los que se aplica EAD destacamos por su relevancia: *The American Heritage Virtual Archive Project*, en el que colaboran la Universidad de California, Berkeley, La Universidad de Stanford, la Duke University y la Universidad de Virginia para la creación de una base de datos compartida, codificada con EAD para recoger, describir y dar acceso a las colecciones documentales sobre la historia y la cultura americana.
- *MARC DTDs (Machine Readable Cataloging Document Type Definitions)*. Proyecto desarrollado por la Network Development and MARC Standards Office de la Library of Congress. El propósito inicial

era crear DTDs de SGML para soportar la conversión de los datos catalográficos desde la estructura del formato MARC a SGML y viceversa sin perder información. El proyecto se inició en 1995 y en estos momentos está disponible en XML. Es posible descargar las DTDs desde la página correspondiente de la Library of Congress.

- *DC (Dublin Core)*, modelo de metadatos más extendido en el ámbito bibliotecario. Conjunto de 15 elementos para la descripción del contenido web atendiendo a tres tipos de información: *Contenido*: Título, Tema, Resumen, etc., *Propiedad intelectual*: Autor, Editor, Derechos, etc. e *Instancia*: Fecha de edición, Forma, Formato y Número internacional. Los elementos Dublin Core son opcionales, repetibles e independientes de la sintaxis del documento. Algunas de las críticas recibidas recaen en su carácter demasiado genérico. Proyectos como *DESIRE*, *BIBLINK* o *ExLibris Special Collections Directory* hacen uso de este modelo, sin duda el empleado con mayor profusión.
- *RDF (Resource Description Framework)*, estándar desarrollado por el W3C para el intercambio de información entre varios sistemas de metadatos. Emplea XML como marco para la interoperabilidad entre distintos modelos de descripción de contenido. La especificación del esquema 1.0 candidato a Recomendación es de 27 de marzo de 2000. Autores como Peis y Fernández Molina (2001) proponen el modelo de metadatos RDF para la gestión de documentos electrónicos y para Méndez Rodríguez (1999) RDF se convertirá en el componente esencial de las bibliotecas digitales. Empleado en proyectos relacionados con el mundo de la Documentación como *CORC: Cooperative Online Resource Catalog*, *eLib*, etc

Con el propósito de escudriñar el complejo panorama de los estándares de metadatos en febrero de 2000 se puso en marcha el proyecto *SCHEMAS (Forum for Metadata Schema Implementers)* para proporcionar un foro para los diseñadores de modelos de metadatos involucrados en proyectos desarrollados bajo el programa IST y en iniciativas nacionales europeas. Entre sus objetivos figura guiar y formar sobre los esquemas de metadatos (jornadas de trabajo, bases de datos web, FAQs, etc.) y promover directrices para la adaptación de múltiples estándares o módulos de metadatos para uso local en esquemas personalizados.

Las líneas de trabajo son ostensibles en multitud de proyectos y sistemas puestos en funcionamiento haciendo uso de tecnología web, lenguajes de marcas avanzados combinados en muchas ocasiones con es-

tándares señalados anteriormente de descripción de contenidos. Reseñaremos sólo algunos de los que consideramos más significativos.

- *DELOS (A network of Excellence on Digital Libraries)* punto de referencia para los proyectos asentados en el programa IST europeo relacionados con las bibliotecas digitales. Su finalidad es estimular el intercambio de experiencias, proporcionar mecanismos para la evaluación de modelos, contribuir a la definición de estándares, etc. Sus actividades se organizan alrededor de cinco foros y están dirigidas al ámbito de las bibliotecas, archivos, patrimonio cultural, publicaciones electrónicas, etc. Uno de sus foros se halla dedicado a la estandarización y su eje central se dedica a las cuestiones relacionadas con el despliegue de esquemas estándar de metadatos e infraestructuras relacionadas con la publicación, registro y esquemas de aplicación conectados.
- *EVA (European Visual Archive)*. Proyecto llevado a cabo dentro de INFO2000 y concluido en febrero de 2001 cuyo sistema continuará desarrollándose. Se ha creado una aplicación accesible en entorno web para proporcionar acceso a colecciones fotográficas del London Metropolitan Archives y del Stadsarchief Antwerpen. Los pilares del proyecto se sustentan en la elaboración de una DTD basada en XML, la utilización del modelo de metadatos Dublin Core para la descripción de contenidos y la implementación de un sistema de búsqueda multilingüe.
- *COVAX (Contemporary Culture Virtual Archives in XML)*. Proyecto europeo desarrollado dentro del programa europeo de IST con la intención de construir un sistema global de información para la búsqueda y recuperación de documentos de archivos, bibliotecas y museos y proporcionar acceso a los documentos primarios con independencia de la ubicación física de los mismos. La dirección del proyecto recae en la Residencia de Estudiantes, formando parte igualmente del consorcio, entre otros, la Universitat Oberta de Catalunya, la Biblioteca Menéndez Pelayo, London and South Eastern Library Region, etc.

Sus objetivos primordiales son: la difusión del patrimonio cultural europeo facilitando el acceso a los ciudadanos europeos a las fuentes del patrimonio intelectual, cultural y científico, la explotación a través de Internet de las infraestructuras culturales existentes, la aplicación de normas en el campo de la estructura y la recuperación de la información y la interoperabilidad entre sistemas.

Para su desarrollo se han registrado tres objetivos específicos:

1. Aplicación de XML y de las DTDs empleadas en la descripción de materiales bibliográficos (MARC DTD), archivos (EAD), materiales de museos (AMICO DTD) y versiones electrónicas de textos culturales (TEI lite). Para ello, los socios se comprometen a la selección de una muestra de registros y documentos procedentes de bibliotecas, archivos y museos, se procederá al análisis de las DTDs existentes para pasar a proponer la estructura de la información común para todos los tipos de documentos, con posterioridad, cada socio convertirá los registros seleccionados a la correspondiente DTD. El proceso de conversión es medular en el proyecto.
2. Formación de una red de repositorios a modo de bases de datos distribuidas a las que pueda acceder cualquier usuario, independientemente del sistema, como si se tratase de una sola base de datos. COVAX actuará como una "metabiblioteca" ofreciendo acceso a un enorme repositorio de cultura contemporánea. Cada socio mantendrá su propio repositorio XML, se desarrollará un motor de búsqueda para acceder a todos los archivos distribuidos
3. Definición y desarrollo de interfaces comunes y aplicaciones de software para facilitar el acceso, la recuperación y la presentación de la información, las interfaces permitirán diferentes búsquedas, globales y específicas, por tipos de documentos o instituciones. La usabilidad del sistema se verificará atendiendo a un amplio abanico de usuarios: docentes, archiveros, bibliotecarios, estudiantes, etc.

Uno de los apartados fundamentales del proyecto está cimentado en la agrupación y búsqueda en bases de datos distribuidas mediante el uso de estándares abiertos sustentados en XML, se ha empleado en el proyecto XMLSpy para manejar y simplificar la labor de marcado y edición de documentos, soporta la sintaxis, la validación, conformidad y correcta construcción de XML, la definición de DTDs, la definición de esquemas, el manejo de XSL, etc. Para la implementación de los repositorios XML se han empleado los paquetes de software *Tamino* de Software AG y *TeXiML* de IXIASoft. Otra innovación esencial viene determinada por

EXPLAIN base de datos en XML basada en el mismo concepto de la norma Z39.50.

Partiendo de las aportaciones de COVAX se ha puesto en marcha el proyecto *Kulturerbe: online* con la intención de crear una infraestructura nacional de servicios, tanto organizativos como técnicos con la finalidad de proporcionar un catálogo virtual del patrimonio cultural de Austria.

- *MALVINE: Manuscripts and Letters via Integrated Networks in Europe*. Proyecto desarrollado por el HIT (*The Humanities Information Technologies Research Program*) de la Universidad de Bergen. La coordinación del proyecto corresponde a la Staatsbibliothek de Berlín y entre los socios, por parte española, figura la Biblioteca de la Universidad Complutense. Su objetivo es habilitar un acceso adecuado a diferentes colecciones de cartas y manuscritos modernos conservados y catalogados en bibliotecas, archivos, centros de documentación y museos europeos con la intención de construir una red que conecte este tipo de instituciones en Europa, independientemente de la plataforma empleada, y accesible universalmente vía web como si se tratase de una base de datos homogénea y unificada.

Proporcionará una interfaz de usuario multilingüe y ofrecerá copias en formato digital de documentos relevantes contribuyendo no sólo a la difusión de este valioso patrimonio sino también a la protección y conservación de los documentos originales.

Dentro del proyecto se ha desarrollado una herramienta de conversión para traducir los catálogos que se encuentran en varios formatos locales a un formato común de intercambio, empleando para este propósito XML o SGML y EAD. Los catálogos se convierten, por tanto, en sistemas abiertos no propietarios facilitando la búsqueda, recuperación e intercambio de datos, la posible publicación en diversos formatos y su almacenamiento en soportes de larga duración.

MALVINE trabaja conjuntamente con el proyecto *LEAF (Linking and Exploring Authority Files)* con el fin de generar una DTD de XML para la normalización de la información bibliográfica. El sistema desarrollará un modelo de arquitectura para un sistema de búsqueda distribuida recogiendo las Autoridades existentes con independencia de su creación en bibliotecas, archivos o museos de su nacionalidad. Establecerá un fichero europeo común de Auto-

ridades —nombres de persona y entidades— con enlaces a las organizaciones que proporcionan información sobre la referida autoridad, así como a los items de la misma. El modelo LEAF será aplicable a todos los proyectos sobre patrimonio cultural.

- *COLLATE (Collaboratory for Annotation, Indexing and Retrieval of Digitised Historical Archive Material)*. Proyecto de la Unión europea con el propósito de implementar un entorno de trabajo vía web para los archiveros, investigadores y usuarios finales interesados en el patrimonio cinematográfico europeo de los años veinte y treinta. La colección está centrada en la documentación filmica histórica, incluyendo archivos censurados, fotos y fragmentos de películas. El fondo será indizado y se le añadirán comentarios realizados por expertos para proporcionar así un valor añadido a la propia documentación. El sistema explota metadatos empleando documentos procesados en XML para aprovechar su potencialidad en la cooperación informacional.
- *Lacito (Linguistic data archiving project)*. Proyecto del CNRS francés (*Centre National de la Recherche Scientifique*) para el archivo de documentos lingüísticos asociando sonido y texto en un mismo formato. El acceso al sistema es independiente de la plataforma y el software empleado y se ha empleado XML para el marcado de los documentos.
- *OAI (Open Archives Initiative)*, desarrolla y promueve la interoperabilidad de estándares con la aspiración de facilitar la difusión de los contenidos. Su origen radica en la idea de intensificar el acceso a publicaciones electrónicas en el ámbito académico, si bien, en estos momentos, el marco tecnológico y los estándares surgidos de esta iniciativa se extienden a un abanico enorme de materiales digitales. El respaldo a esta iniciativa se asienta en la *Digital Library Federation*, *Coalition for Networked Information*, la *National Science Foundation* y la *Defense Advanced Projects Agency*. La OAI ha definido el *Open Archives Metadata Harvesting Protocol* como mecanismo para proporcionar y proveer un marco de interoperabilidad independiente de aplicaciones que pueda ser usado por distintos organismos dedicados a la publicación de contenidos en la web. El protocolo soporta distintos modelos de metadatos, emplea XML para codificar los registros y hace uso de URI (Uniform Resource Identifiers) como clave para extraer los metadatos de un determinado item.
- *CIMI Consortium*. Consorcio de instituciones y organismos relacio-

nados con el patrimonio cultural fundado en 1990 y una de cuyas principales áreas de interés se halla centrada en el fomento del uso de estándares y su difusión en la comunidad museística. En estos momentos trabaja en el *Proyecto MIDIIS (Museum Initiative for Digital Information Interchange Standards)* entre cuyos proyectos de investigación destacan: mda SPECTRUM XML DTD adaptada como estándar web para la descripción de objetos de museos, se investiga en la interoperabilidad de productos XML de museos con trabajos semejantes de archivos, bibliotecas y centros de investigación. No se trata tanto de trabajar en un nuevo estándar como de facilitar la migración y conversión de datos de un sistema a otro reduciendo los costes actuales. Dentro de MIDIIS se asienta un proyecto de tratamiento de metadatos para adaptar los registros empleados bajo el modelo Dublin Core a la metodología de OAI. Gran interés presenta, por último, el proyecto *Harmony* para buscar la interoperabilidad entre distintos vocabularios de metadatos.

- *RLG Cultural Materials Initiative*. Iniciativa del Research Library Group, (USA) orientada a promover el acceso a materiales culturales. Se trata de una corporación sin ánimo de lucro constituida por alrededor de 160 universidades, bibliotecas nacionales, archivos, sociedades históricas y otras instituciones con colecciones relevantes para la investigación y la enseñanza. La complejidad del proyecto reside en el hecho de tratar-aunar diferentes tipos de instituciones, que tratan diferentes tipos de información y que emplean diferentes prácticas descriptivas. El acceso se efectúa vía web empleando como interfaz de búsqueda y recuperación el sistema Eureka. Un apartado esencial dentro de RLG lo constituyen las colecciones archivísticas, la base de datos de RLG reúne instrumentos de descripción creados utilizando EAD así como en otros formatos (HTML o ASCII) y posibilita que éstos se hallen asociados potencialmente a la fuente primaria digitalizada.
- *RSLP Collection Description*. Proyecto incluido en el Research Libraries Programme de UKOLN (UK Office for Library and Information Networking). En la elaboración del modelo se ha contado con el apoyo presupuestario de OCLC, la intención es que sea aplicable a colecciones físicas y digitales de todo tipo, incluyendo bibliotecas, colecciones de arte y materiales museísticos. Tiene la intención de contribuir al desarrollo en el panorama de la descripción de colecciones y en el debate sobre los diversos modelos de metadatos. El proyecto desarrollará un modelo de metadatos y

empleará como sintaxis asociada RDF. Parece consolidada la propuesta de que en un futuro se creen herramientas de conversión del esquema de descripción RSLP a ISAD(G)/EAD y viceversa.

- *CHIOS (Cultural Heritage Interchange Ontology Standardization)*. Programa liderado por la *Foundation for Research and Technology Hellas* y en el que participan instituciones de la relevancia del *Germanisches Nationalmuseum* de Nuremberg, la *Museum Documentation Association*, el *National Museum* de Dinamarca o el *Research Library Group* americano. El objetivo del proyecto es asegurar que el modelo CIDOC CRM, ya aceptado por ISO/TC46, se constituya en estándar internacional. El CRM es la primera ontología formal cuyo propósito es facilitar el intercambio e interoperabilidad de la información cultural entre museos, bibliotecas y archivos. Ha sido desarrollado por el *ICOM/CIDOC*.
- *EGHD (European Chronicles On-line)*. Programa incluido en el IST del V Programa marco. El objetivo es el desarrollo de aplicaciones y de nuevos modelos de metadatos para películas con el fin de dar soporte al desarrollo de bibliotecas digitales de materiales audiovisuales. El proyecto intenta mejorar la accesibilidad, la recuperación y el acceso distribuido a grandes colecciones audiovisuales. Se construye sobre la base de los dos modelos de bibliotecas digitales audiovisuales: *Informedia Digital Video Library* y *Media Archive*. El proyecto comenzará a trabajar con las colecciones nacionales de archivos filmicos documentales de Alemania, Francia, Italia y Suiza, con la posibilidad de que otros archivos se unan más adelante. El proyecto definirá un modelo común de metadatos adaptado a la información filmica cuya asignación se realizará de forma semi-automática, diseñará una interfaz multilingüe, desarrollará técnicas de producción de sumarios visuales, etc.
- *MASTER (Manuscript Access through Standards for Electronic Records)*. Proyecto europeo iniciado en 1999 bajo el IV Programa Marco, liderado por el Centre for Technology and the Arts de De Monfort University (Leicester), participan la Biblioteca Real de los Países Bajos, (La Haya), el Instituto Arnamagnean (Copenhague), la Biblioteca Nacional checa (Praga), el IRHT (Institut de Recherche et d'Histoire des Textes, CNRS, París) y la Universidad de Oxford. Además, otras importantes bibliotecas e instituciones son miembros asociados del proyecto, el cual mantiene una estrecha relación con muchos de los proyectos sobre manuscritos desarrollados en Norteamérica (*EAMMS, Electronic Access to Medieval Manus-*

*cripts*) y *Digital Scriptorium*) así como con la iniciativa TEI. Destinado a crear un único catálogo en línea de manuscritos medievales custodiados en bibliotecas europeas. Se ha desarrollado un estándar único para la descripción de manuscritos legibles por máquina, la descripción de los registros se realiza conforme a SGML/XML. El IRHT ha construido una base de datos elaborada con Microsoft Access desde la cual los registros pueden ser exportados a otros proyectos construidos en formato SGML/XML. El catálogo contendrá asimismo imágenes de muchos de los manuscritos

Proyectos igualmente relevantes como el alemán del Archivschule Marburg de la Philipps\_Universität Marburg centrado igualmente en el uso de XML, PRESTO (*Preservation Technology for European Broadcast Archives*) y AMICITLA (*Asset Management Integration of Cultural heritage In The Interchange between Archives*) integrados en el IST están cimentados en el empleo de nuevas tecnologías y, en concreto en el uso de metadatos aplicados a archivos audiovisuales.

Como colofón, queremos señalar dos importantes proyectos inscritos dentro del programa IST europeo: *Renardus*, cuyos socios pertenecen a bibliotecas europeas e instituciones relacionadas con la información. Su objetivo es facilitar el acceso a recursos de calidad y de interés para las comunidades académica y de investigación mediante el desarrollo de directorios analíticos (*subject gateways*) que proporcionen a los usuarios un acceso integrado a través de una única interfaz a recursos distribuidos en Internet. En el enfoque del proyecto se aúna la investigación sobre estándares, aplicaciones técnicas y modelos de metadatos. Por último, el proyecto REGNET (*Cultural Heritage in REGIONal NETWORKs*) iniciado el 1 de abril de este mismo año, pretende conjugar la distribución de contenidos culturales con las posibles actividades comerciales potencialmente desarrolladas por estos organismos dirigidas a los usuarios/clientes (B2C) y a otras empresas (B2B). El proyecto contempla el uso de estándares propios de la descripción documental (UNIMARC, ISAD(G), CIDOC), el uso de Dublin Core y de otras DTDs, el almacenamiento de información estructurada en XML, el desarrollo de hojas de estilo Xsl, la integración de un sistema de búsqueda y recuperación de la información basado en el estándar Z39.50 o metodologías similares basadas en XML, distribución de contenidos para aplicaciones relacionadas con la telefonía móvil, etc. Así el proyecto cuenta entre sus bloques de trabajo con un subsistema de ontologías, dedicado a especificaciones de metadatos necesarias tanto en el ámbito de los contenidos relacionados con el pa-

patrimonio como en el dedicado al comercio electrónico.

## CONCLUSIONES

Se pretende la creación de sistemas potentes que permitan la gestión, recuperación, intercambio e integración de la información en un modelo abierto y distribuido como es la WWW. En este entorno digital, se ha producido una fructífera confluencia de intereses en bibliotecas, archivos y museos, emblemas del patrimonio cultural, compartiendo investigaciones sobre métodos de trabajo, estándares, interfaces, protocolos, técnicas de almacenamiento y recuperación, etc. Creemos que a pesar del aparente desconcierto que pueden causar el considerable número de proyectos emprendidos, no se trata de imposición de formatos sino de establecer posiciones comunes cooperativas que eviten la confusión, la duplicación de esfuerzos y se encaminen al intercambio internacional de información descriptiva. Obviamente, la naturaleza de sus colecciones, la diferente tipología institucional y los papeles históricamente desempeñados por cada una de ellas son bien diferentes y, paradójicamente, las nuevas tecnologías están actuando como una vía de integración en el nuevo prototipo de acceso a la información.

Del desarrollo de los proyectos aquí reseñados es posible coagular una serie de líneas de trabajo dirigidas al desarrollo de sistemas de información sustentados en la aplicación de modelos de metadatos, de eficacia ya demostrada, en conjunción, en algunos de los casos, o separadamente en otros, con la definición de estructuras lógicas de tipos documentales formalizadas, mayoritariamente, a través de DTDs de XML o Esquemas XML para incrementar las posibilidades de indización y recuperación; aparece asimismo como preocupación capital la convergencia de las tradicionales normas de descripción documental y los requisitos para su implementación en el entorno digital.

Importancia también cardinal poseen las investigaciones en el desarrollo de la interoperabilidad de sistemas, interoperabilidad entre distintos lenguajes de metadatos y diseño de interfaces de búsqueda y recuperación multilingües para satisfacer al nuevo usuario digital.

En el caso europeo cobran un especial dinamismo los proyectos centrados en archivos audiovisuales: archivos sonoros, colecciones fotográficas, cinematográficas, etc., tal vez la razón se halle en los materiales especiales de los que parten y sus particulares problemas de gestión y conservación.

El desarrollo de los contenidos culturales ha de venir de la mano de la estandarización, la cooperación internacional y el desarrollo de políticas

de información globales, es destacable, en este sentido, la Recomendación R (2000) 13 del Comité de Ministros del Consejo de Europa considerando que los archivos constituyen una parte esencial e irremplazable del patrimonio cultural asegurando la preservación de la memoria de la humanidad y garantizando la posibilidad de conocer de manera objetiva los elementos de la historia particular de cada país, por ello aconseja que los gobiernos de los Estados miembros adopten una legislación en materia de comunicación de archivos o ajusten la legislación existente a los principios enunciados en la recomendación.

Es observable como el despliegue de proyectos de investigación de los últimos años ha propiciado el asentamiento de vigorosos sistemas de información en funcionamiento, son ya propuestas reales que animan a continuar las líneas de trabajo emprendidas.

## BIBLIOGRAFÍA

- ALADRO, A. «XML y Java». *iWorld*, 2001, marzo, p 68-76.
- ALÒS-MONER, A.; COMAPOSADA, M.; SERRANO, J; AGUILERA, B. «El format XML obre la possibilitat d'accés a catàlegs bibliogràfics [en línia]» En: *Jornades Catalanes de Documentació (7.1999.Barcelona)* <http://www.cobdc.org/7es/6.pdf> [Consulta: 26 jul. 2001]
- CARIDAD SEBASTIÁN, M. (Coord.) *La sociedad de la información: política, tecnología e industria de los contenidos*. Madrid: Centro de estudios Ramón Areces, 2000.
- DEMPSEY, L. «Scientific, industrial, and cultural heritage [en línea]: a shared approach, a research framework for digital libraries, museums and archives» *Ariadne*, 1999, n. 22 <<http://www.ariadne.ac.uk/issue22/dempsey/>>. [Consulta: 27 jul. 2001]
- GUTIÉRREZ RODRÍGUEZ, A. ; MARTÍNEZ GONZÁLEZ, R. *XML a través de ejemplos*. Madrid: Ra-ma, 2001.
- HÍPOLA, P.; EÍTO BRUN, R. «Edición digital: formatos y alternativas». *El profesional de la información*, 2000, vol. 9, n. 10, p. 4-15
- LASO BALLESTEROS, I. «Arquitectura para un sistema transeuropeo entre archivos». *Boletín ACAL: Archivamos*, 2000, n. 38, p. 44-48.
- MARCOS, M.C: "Los archivos en la era digital". *El Profesional de la información*, 1999, vol. 8, n. 6 , p. 4-13.
- MARTÍN GALÁN, B. ; RODRÍGUEZ MATEOS, D. «Estructuración de la información mediante XML: un nuevo reto para la gestión documental». En: *Jornadas Españolas de Documentación (7. 2000. Bilbao)*. Bilbao: FESABID, 2000, p. 113-123.

- MÉNDEZ RODRÍGUEZ, E. «RDF [en línea]: Un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio». En: *Jornades Catalanes de Documentació* (7.1999.Barcelona) <http://www.cobdc.org/7es/1.pdf> [Consulta: 19 jul.2001]
  - MÉNDEZ RODRÍGUEZ, E.; MERLO VEGA, J.A. «Localización, identificación y descripción de documentos web: tentativas hacia la normalización». En: *Jornadas Españolas de Documentación* (7 .2000. Bilbao). Bilbao: FESABID, 2000, p. 221-231.
  - MUSCIANO, C. ; KENNEDY, B. *HTML y XHTML*. Madrid: Anaya Multimedia, 2001.
  - ORTEGO DE LORENZO-CÁCERES, M.P.; BONAL ZAZO, J.L. «Archivos en línea [en línea]: formatos de difusión de información archivística». *Biblit.* 2001, Año 2, n. 8 <http://bibliotecas.rcp.net.pe/biblios> [Consulta: 27 jul. 2001]
  - PEIS, E., FERNÁNDEZ-MOLINA, J.C. «Modelos metadatos para la gestión de documentos electrónicos». En: *Congreso ISKO-España (5.2001.Alcalá de Henares). La representación y organización del conocimiento: metodologías, modelos y aplicaciones*. Alcalá de Henares: Facultad de Documentación, 2001.
  - PEIS, E.; MOYA, F. «SGML y servicios de información» *El profesional de la información*, 2000, vol. 9, n. 6, p. 4-17.
  - RODRÍGUEZ MUÑOZ, J.V. ; DÍAZ ORTUÑO, P.M. «Arquitectura de la información: XML y web». En: *Anuari de Biblioteconomia, Documentació i Informació: BIBLIODOC 2000*. Barcelona: Col.legi Oficial de Bibliotecaris-Documentalistes de Catalunya, 2000 , p. 151-166.
  - ROSA, A. de la «Lenguajes de marcas aplicados a la transformación de estructuras documentales». *El Profesional de la información*, 2001, vol. 10, n.1-2, p. 4-22.
  - ROSA, A. DE LA; SENSO, J.A. «XML como medio de normalización y desarrollo documental». *Revista española de Documentación científica*, 1999, vol. 22, n. 4, p. 488-504.
  - SANTAMARÍA GALLO, A. «EAD: una norma para facilitar la distribución e intercambio de instrumentos de descripción en la Red». *Boletín ACAL: Archivamos*, 1999, n. 33, p. 28-30.
  - SANTAMARÍA GALLO, A. «Codificación de instrumentos de descripción archivísticos con SGML/XML: la norma EAD, v. 1.0, (Encoded Archival Description v.1.0)» *Tábula*, 1999, n. 4, p. 69-125.
  - SENSO, J.A.; ROSA, A. de la. «Especificaciones XML aplicadas a la Documentación». En: *Anuari de Biblioteconomia, Documentació i Informació: BIBLIODOC 1999*. Barcelona: Col.legi Oficial de Bibliotecaris-
-

- Documentalistas de Catalunya, 1999, p. 133-152
- SERRANO-TENILLADO, M.A.; CALVO, A. ROMERO, J.A. «La representación de documentos históricos: aplicación de esquemas y XML para modelar y representar la información y el conocimiento descrito en las fuentes de protocolos notariales». En: *Congreso ISKO-España (5º.2001.Alcalá de Henares). La representación y organización del conocimiento: metodologías, modelos y aplicaciones*. Alcalá de Henares: Facultad de Documentación, 2001.
  - SOCIETY OF AMERICAN ARCHIVISTS. *EAD: Descripción archivística codificada: directrices de aplicación, versión 1.0*. Madrid: Fundación Histórica Tavera, 2000.
  - SWEET, M. «The internationalisation of EAD (Encoded Archival Description)». *Journal of the Society of Archivists*, 2001, Vol. 22, n. 1, p. 33-38.

#### **DIRECCIONES WEB RELACIONADAS:**

- Archivo virtual de la Edad de Plata: 1868-1936.  
<http://www.archivovirtual.org>
- CHIOS <http://cidoc.ics.forth.gr/chios.html>
- CIMI Consortium <http://www.cimi.org>
- COVAX <http://www.covax.org>
- DELOS <http://www.iei.pi.cnr.it/DELOS>
- Dublin Core <http://dublincore.org>
- EAD <http://lcweb.loc.gov/ead/>
- ECHO <http://pc-erato2.iei.pi.cnr.it/echo>
- EVA: European Visual Archives <http://192.87.107.12/eva>
- ISO/IEC 15445:2000 Hypertext Markup Language (HTML)  
<http://www.cs.tcd.ie/15445/UG.html>
- LEAF <http://www.crxnet.com/leaf/>
- MALVINE <http://www.malvine.org>
- MARC DTDS <http://lcweb.loc.gov/marc/marcdtd/marcdtdback.html>
- MASTER <http://www.cta.dmu.ac.uk/projects/master/>
- OAI Open Archives Initiative <http://www.openarchives.org>
- RDF Resource Description Framework. Schema Specification 1.0  
<http://www.w3.org/TR/rdf-schema>
- REGNET <http://www.regnet.org>
- RLG Cultural Materials Initiative <http://www.rlg.org>
- Renardus <http://www.renardus.org>
- RSLP Collection Description <http://www.ukoln.ac.uk>
- SCHEMAS: Forum for Metadata Schema Implementers

<http://www.schemas-forum.org>

- TEI <http://www.tei-c.org>
- W3C HTML 4.01 Specification: W3C Recommendation 24 December 1999. <http://www.w3.org/TR/html401/>
- W3C XML 1.0 (Second Edition) W3C Recommendation 6 October 2000. <http://www.w3.org/TR/REC-xml>