

Archivos en la era digital: problema (y solución) de los recursos electrónicos

Por María José de Acuña y Xavier Agenjo

Resumen: Son muy numerosos los artículos publicados acerca de lo que suponen, y pueden suponer, los recursos digitales para el mejor funcionamiento de las llamadas instituciones de memoria. Pero la mayoría de ellos se refieren a los recursos digitales como algo que se consume y se pone a disposición de los usuarios, rara vez como algo que se produce y se pone también a su disposición. A partir de las funciones propias de cualquier tipo de archivo, se propone crear recursos digitales acordes con una normativa dada (estructuras de metadatos, OAI-PMH, web semántica, etc.) con la finalidad de incrementar la capacidad de recuperación de la información y la visibilidad de ésta, tanto para especialistas como para cualquier tipo de usuario. Europa debe marcarse como objetivo prioritario desarrollar nuevos sistemas de recuperación de la información y nuevos productos basados en estándares y protocolos que le proporcionen una más competitiva posición en la sociedad de la información.

Palabras clave: Recursos digitales, Metadatos, Web semántica, Archivos, Protocolo OAI-PMH.



María José de Acuña es licenciada en ciencias de la información y en geografía e historia. Desde octubre de 2000 forma parte de la empresa Digibis Producciones Digitales en la que coordina proyectos de digitalización. Trabaja en cibermetría y en implantación de sistemas integrados de gestión.

Xavier Agenjo es licenciado en filosofía y letras y pertenece al cuerpo facultativo de archiveros, bibliotecarios y arqueólogos (en excedencia). Ha sido jefe del Servicio del Catálogo Colectivo del Patrimonio Bibliográfico y de la Unidad de Coordinación Informática de la Biblioteca Nacional, director de su Departamento de Acceso a la Información y al Documento así como director de la Biblioteca Menéndez Pelayo. Desde enero de 2002 es director de proyectos de la Fundación Ignacio Larramendi, donde se encarga de la coordinación de los trabajos de la Bibliotecas Virtuales Larramendi. Ha participado en numerosos proyectos nacionales e internacionales y es autor de un centenar de publicaciones en el campo de la biblioteconomía, la bibliografía y la automatización y digitalización de procesos documentales.



Title: Archives in the digital age: the problem (and solution) of electronic resources

Abstract: Many articles have been written about what digital resources represent –or could represent– for improving the performance of ‘memory institutions’. But the majority of them refer to digital resources as things to be consumed and made available to users, but rarely as something to be produced and later made available to them. Starting with the functions of archives, the article proposes the creation of digital resources according to standards (metadata structures, OAI-PMH, semantic Web, etc.) in order to enhance resource discovery and information visibility, for specialists as well as for other types of users. Europe must give top priority to the development of new information retrieval systems and new products based on emerging standards and protocols in order for European countries to achieve a more competitive standing within the information society.

Keywords: Electronic resources, Metadata, Semantic web, Archives, OAI-PMH protocol.

Acuña, María José de; Agenjo, Xavier. “Los archivos en la era digital: el problema (y la solución) de los recursos electrónicos”. En: *El profesional de la información*, 2005, noviembre-diciembre, v. 14, n. 6, pp. 407-413.

1. Las dificultades del nuevo escenario

Recientemente, la *Secretaría de Estado* del ramo criticaba la tendencia del sector público y privado español a utilizar las partidas presupuestarias disponibles

para las tecnologías de la información y las comunicaciones simplemente adquiriendo recursos (hardware, software, licencias, permisos) producidos fuera de España. Ése era justamente el camino contrario al que se

Artículo recibido el 10-09-05
Aceptación definitiva: 11-10-05

debería emprender para crear una industria informática y cultural europea (siguiendo la *Estrategia de Lisboa*) y, en particular, española, verdaderamente competente.

Se trata, pues, de apostar por una industria europea y adquirirla o implementarla por el mero hecho de haberse desarrollado en Europa. Se habla de realizar productos de igual o superior calidad con los que se consiga neutralizar la marcada dependencia existente en la actualidad de los grandes centros de información digital norteamericanos. Alcanzar en 2010 los objetivos fijados en la *Estrategia de Lisboa* y hacer de Europa un área geográfica más competitiva y dinámica desde el punto de vista económico exige experimentar un salto tecnológico suficientemente importante como para lograr contrarrestar el terreno perdido frente a EUA y algunos países de Asia.

«La tendencia a recopilar la información y a ponerla (previo pago) a disposición de todo el mundo desde los EUA, no sólo no se ha modificado sino que se ha visto intensificada»

La llamada de atención que ha venido a suponer el proyecto de *Google* para crear una biblioteca digital con 15 millones de obras digitalizadas, ha alarmado a algunos países europeos que se ven forzados o a responder a esa iniciativa o a permitir que la mayoría de los recursos digitales referentes a Europa acaben siendo consultados en bases de datos creadas en los EUA. Esta circunstancia terminará dándose también en los archivos.

No puede decirse que la situación descrita sea nueva. Ya los antiguos ‘terminalistas’, solían conectarse a *Dialog*, en Palo Alto, California, de forma prácticamente exclusiva, con algunas excepciones francesas o inglesas. La tendencia a recopilar la información y a ponerla (previo pago) a disposición de todo el mundo desde los EUA, no sólo no se ha modificado sino que se ha visto intensificada. Y también ha ocurrido lo mismo con las fuentes de información, recursos y bases de datos de acceso gratuito, siendo el número de recursos en la web procedentes de EUA incomparablemente mayor que la producción europea.

Si se parte de la idea de que las instituciones de memoria¹ pueden y deben ser productoras de recursos y contenidos digitales, en el caso que nos ocupa (los archivos) los identificaremos como generadores de información atendiendo a las funciones que le son inherentes. O dicho de otra manera: los archivos atesoran y conservan una enorme cantidad de información que puede ser transformada en contenidos accesibles a tra-

vés de la web para diferentes tipos de usuarios. No se trata aquí de si son creados a partir de la digitalización de documentos en papel o si han nacido ya digitalmente; ni de la forma en que deben ser tratados a lo largo de la vida de una institución, como evidencia o prueba de una actividad; o de los pasos y procesos que deben seguir los documentos hasta su conservación en un archivo histórico, si es el caso. Cada organismo, cada archivo, seguirá los procedimientos que estén establecidos en la normativa profesional, pero en todos ellos existe el potencial de generar contenidos para la web². Por una parte, al describir los documentos el archivero crea bases de datos; por otra, el archivo custodia documentos generados en diferentes períodos históricos, por distintas instituciones y en varios soportes (de manera creciente en formato electrónico). Finalmente, el archivo facilita el acceso a la información poniéndola al alcance de los diversos usuarios, incluidas las copias digitales, cuando la ley y los recursos disponibles lo permiten.

Resulta evidente que los archivos han profundizado en la aplicación de las tecnologías de la información a un ritmo espectacular durante los últimos veinticinco años. Sin duda ha influido notablemente la aparición de normativa internacional para la descripción archivística –*ISAD(G)* e *Isaar(CPF)*– y para la transmisión de esta información en formatos estructurados como las definiciones de tipo de documento *EAD (Encoded Archival Description)* y *EAC (Encoded Archival Context)*. Igualmente, la importancia de la gestión archivística en cualquier institución puede verse también en la publicación de normas como la *ISO 15489 Information and documentation–records management (2001)* o la *ISO 23081 Information and documentation–records management processes–metadata for records (2004)*. Sin apartarse de las funciones tradicionales apuntadas, los archivos desempeñan un papel fundamental en todo lo relacionado con la gestión integral de cualquier tipo de información durante todo su ciclo de vida, desde su nacimiento hasta su expurgo o conservación permanente. En este sentido los archiveros han desarrollado técnicas y procedimientos con los que han logrado mejorar la difusión de su competencia profesional, de manera tal que hoy día este colectivo recibe por parte de la sociedad una consideración (que, dicho sea de paso, siempre merecieron) como piezas claves del motor de las instituciones³. Esta apreciación aumentará con toda seguridad en los próximos años, a medida que estas instituciones difundan a través de la web las ingentes cantidades de información contenidas en los documentos primarios que atesoran.

Esta difusión sin embargo planteará, como ya lo ha hecho en otros entornos como el bibliotecario o el académico, problemas para su localización, búsqueda y

recuperación en el entorno distribuido de la web actual, y mucho más en la futura web semántica. Estos retos han de solventarse con la creación de un nuevo tipo de estructuras de información y aplicaciones informáticas que no sólo sean capaces de solucionar eficazmente los procesos archivísticos tradicionales, sino de añadir nuevas funcionalidades basadas fundamentalmente en el acceso y la difusión inteligente de la información. De esta forma será posible crear/producir (y no sólo adquirir/consumir) nuevos recursos electrónicos que estén al servicio de todos los usuarios y del archivo mismo.

2. Incremento de la visibilidad de los recursos digitales⁴

A partir del planteamiento de la conveniencia de crear contenidos accesibles en la web a partir de la información conservada en archivos, debemos enfrentarnos a las limitaciones que desde el punto de vista funcional presentan los buscadores para encontrar recursos digitales en general, y archivísticos en particular. Hasta hace poco tiempo, su utilidad se circunscribía a páginas estáticas html, aunque ya se ha incorporado la búsqueda en el contenido de documentos en pdf, Word, y otro tipo de formatos. Desde luego, las bases de datos permanecen inaccesibles para ellos, lo cual representa un enorme reto a la hora de aumentar la visibilidad los recursos digitales que produzcan los archivos. La solución que está fomentando la comunidad internacional se basa en el uso de la recolección de metadatos⁵ (*metadata harvesting*), especialmente en el protocolo *OAI-PMH*⁶ (*Open Archives Initiative-Protocol for Metadata Harvesting*), sobre lo que existe en España cierta literatura⁷ que ha permitido un conocimiento relativo de sus características. Es de esperar que esta experiencia se concrete en ejemplos legibles en la web. En la actualidad existen en España 19 repositorios *OAI*, pertenecientes en su mayoría al ámbito de las universidades y surgidos del entorno de las bibliotecas universitarias.

En nuestra opinión, éste debe ser un protocolo a implantar en aquellos archivos e instituciones de memoria, que quieran no sólo crear contenidos para la web sino también implantar aplicaciones que favorezcan su localización. Si bien *OAI-PMH* surgió en el seno de la comunidad académica y científica para la búsqueda y recuperación de textos electrónicos, es perfectamente aplicable en cualquier contexto, incluido por supuesto el archivístico. Supone una alternativa a la visión del mecanismo de búsqueda y recuperación propuesto por el modelo distribuido de Z39.50. Se aleja de éste por la complejidad de su aplicación y por su falta de precisión (tan bien conocida por los usuarios habituales de este procedimiento cuando consultan más de un servidor a la vez) a causa de los diferentes

grados de aplicación de la norma y porque Z39.50 no es un protocolo web.

Las características básicas de *OAI-PMH* son:

—Flexibilidad para adaptarse a cualquier ámbito: puede ofrecer información sobre todo tipo de recurso, tanto físico como digital.

—Gran facilidad de implantación. Los componentes del modelo *OAI* están basados en un recolector de metadatos o *harvester*, en una interfaz de búsqueda y recuperación y un repositorio común formado por la recolección de repositorios individuales, así como un conjunto de peticiones y respuestas que se realizan a través de http. Las especificaciones para definir la estructura que permite diseñar, tanto el recolector⁸ como los repositorios⁹ pueden encontrarse en el sitio web de *Open Archives Initiative*.

Sin embargo, es importante señalar que —aunque, lógicamente, es necesaria la figura de un analista programador para llevar a cabo el desarrollo y poner en funcionamiento el software— su estructura es absolutamente inteligible para un archivero, para un bibliotecario o para un museólogo. La primera impresión, como ocurre ante cualquier lenguaje de marcado, parece presuponer una codificación difícil, pero no lo es mucho más de lo que pueda serlo *EAD* y es perfectamente asumible por los profesionales de las instituciones de memoria para impulsar el desarrollo de los sistemas de recuperación basados en este protocolo.

«La capacidad del protocolo de recolección de metadatos OAI para fomentar el acceso a la información conservada en los archivos puede verse en OAIster»

Muestra de que *OAI* no es sólo un protocolo para la búsqueda y recuperación de artículos científicos o de material bibliográfico en la web es que puede transmitir y presentar la información en varios formatos, los básicos para los profesionales de las instituciones de memoria, estando entre ellos *EAD* y siendo el formato mínimo y común para cualquier entorno *Dublin Core* sin calificar.

Es de señalar la importantísima sinergia que ha surgido entre *OAI* y los protocolos de recuperación de información en bases de datos distribuidas en el entorno http, dentro de *ZING* (*Z39.50 International Next Generation*) y denominados *SRW/U* (*Search/Retrieve Web Service*)¹⁰. Se denominan así porque en el primer caso (*SRW*) el protocolo es operativo mediante *SOAP* mientras que *SRU* lo es a través de url. Esta correlación entre *OAI* y *SRW/U*¹¹ era previsible puesto que en

último término ambos protocolos tienen como objetivo facilitar la búsqueda y recuperación de la información, aunque afrontándola desde diferentes perspectivas. Por otra parte, se sustentan en la utilización de estructuras de metadatos específicas, en las que los archivos están representados por medio de *EAD*.

Son ya millones los recursos digitales accesibles en la red mediante *OAI* en los distintos recolectores que existen. *OAIster*¹², servicio creado por la *Universidad de Michigan*, impulsora del protocolo *OAI*, es uno de los más populares, pero *Scirus*¹³, y otros, son también increíblemente potentes. En el momento de dar la última redacción a ese trabajo (29 de julio de 2005) *OAIster* reunía 5.704.392 registros procedentes de 510 instituciones. La capacidad del protocolo de recolección de metadatos *OAI* para fomentar el acceso a la información conservada en los archivos puede verse en el servicio de búsqueda y recolección que proporciona *OAIster*. Dentro de las instituciones que ofrecen sus bases de datos en forma de repositorios *OAI* recolectables, se encuentran dos importantes muestras de cómo los archivos pueden aplicar este protocolo y difundir sus colecciones: *Archives in London and the M25 Area* y *Princeton University Library*.

Todo ello nos lleva a la necesidad de mantener una política sistemática de asignación de metadatos, tanto de los instrumentos de descripción como de los documentos digitales, para que puedan ser recuperables en la web y alcanzar un máximo de visibilidad. Consideramos necesario destacar que estamos hablando de aquellos metadatos que faciliten la recuperación de la información por todos los usuarios y que podrán ser derivados, extraídos o convertidos a partir de los procesos de asignación o captura propia de la gestión documental o archivística. Es indiscutible que sólo mediante una adopción sistemática de una política de descripción que tenga en cuenta la normativa archivística, de gestión documental y los esquemas de transmisión de los protocolos de búsqueda y recuperación de información será posible aprovechar el esfuerzo en la creación de metadatos para facilitar el uso de la información a sus usuarios finales.

Los archivos custodian la información, pero no siempre es accesible para el usuario final; en general facilitan su búsqueda, pero están aún muy lejos de disponer de herramientas para el usuario que les permita consultar los documentos eficazmente. No se trata sólo de ofrecer, mediante los instrumentos de descripción, la información que se tiene reunida, sino de crear y desarrollar aplicaciones más potentes para ofrecer toda la información no reservada. La búsqueda y recuperación de información en bases de datos archivísticas presenta unas peculiaridades específicas; en ellas se agrupan fondos, secciones, colecciones documenta-

les, etc. con una fuerte estructura jerárquica de los diferentes niveles de descripción y en las que el contexto de un documento dentro de esa jerarquía constituye una información fundamental. En último término, las descripciones elaboradas según la *ISAD(G)* [o *Isaar(CPF)*], o mejor aún, formateadas en *EAD* o en *EAC*] no dejan de ser sino documentos secundarios que facilitan el acceso a los contenidos, pero no son los propios documentos primarios.

«Los archivos custodian la información, pero no siempre es accesible para el usuario final»

Lo que aquí se propone es que la difusión de la información contenida en los archivos y los objetos digitales asociados (bien sea por digitalización o por tratarse de documentos electrónicos), incorporen siempre sus metadatos conforme a la normativa expuesta para que sea posible el desarrollo de repositorios abiertos en los que descargar sistemáticamente los registros. De esta forma, podrán ser recogidos por recolectores desarrollados por terceros y ofrecer a los usuarios tanto nuestros propios recursos digitales o electrónicos como el compendio de información producida por terceros. Es fundamental añadir que existe una inmensa cantidad de información todavía no digitalizada y, siguiendo las líneas expuestas, uno de los aspectos estratégicos del proceso global de digitalización consiste en llevarla a cabo no sólo para funciones de preservación, o incluso de consulta simple, sino para edificar en torno a ellas repositorios y recolectores basados en *OAI*.

Ésta es una propuesta para animar a la creación de contenidos para la web actual y facilitar su difusión en la misma, pero el reto que se nos avecina es afrontar los condicionantes de la web semántica. El siguiente apartado se centrará en la tecnología que permitirá construir a medio plazo una constelación de sitios conforme a la metodología definida por la normativa del *W3C* y que tiene como objeto ir construyendo la futura web semántica. Hacer factible la visibilidad de los recursos digitales existentes y mejorar su búsqueda y recuperación se ha convertido en el objetivo fundamental de los organismos que, a escala internacional, están impulsando su desarrollo.

3. La web semántica

Sobre esta nueva concepción de la red hay tantas definiciones que quizá lo más práctico sea remitir al interesado a la página del *W3C*¹⁴ y permitir que allí se beba en las fuentes directas¹⁵. En último término, es un nuevo modelo de estructuración, interrelación y recuperación de la información, gracias al uso de una serie

de normas y estándares ya existentes, como xml o rdf, o algunas nuevas creadas al efecto como OWL¹⁶, así como determinados protocolos de comunicación y estructura de la información que configuran una solución informática, completamente nueva, original y sumamente potente. Con frecuencia, aquellos profesionales que se sienten interesados por las perspectivas que parecen ofrecer las ontologías para las instituciones de memoria, lamentan no encontrar entre las muchas que se están desarrollando en los últimos años ninguna que encaje exactamente con su campo de competencia. Se podría mencionar, tal vez, por su interés en un entorno europeo, dos iniciativas específicas: *EuroStory.net*¹⁷ y *Vicodi*¹⁸, que sí participan, y de forma avanzada, de este nuevo modelo conceptual e informático.

Lo que se pretende es superar las carencias e insuficiencias que la world wide web histórica ha ido presentando a medida que crecía exponencialmente el número de sitios en la Red. La web semántica tiene como objetivo sustentar los sitios y servicios de la web en unas nuevas estructuras de información constituidas por una red organizada de conceptos y relaciones semánticas pertenecientes a un dominio del conocimiento a las que se denomina 'ontologías'. La suma de conceptualizaciones de los diferentes dominios del conocimiento concretadas en ontologías, junto con los agentes de software intermediarios son sus elementos constitutivos, de igual modo en que el conjunto de los sitios actuales componen en la actualidad la web que conocemos¹⁹.

«A través de la estructura semántica que componen las ontologías, tanto los usuarios como los agentes de software recuperan información verdaderamente pertinente»

La web actual puede caracterizarse, entre otras muchas maneras, porque los recursos y la información que contienen están relacionados únicamente mediante enlaces simples (urls). Los usuarios navegan entre los referidos recursos por medio de estos enlaces, sin embargo, el contenido del sitio, recurso o página, y por tanto, su interés o pertinencia es sólo inteligible para el usuario. Los sistemas informáticos, aplicaciones y agentes de software de la web actual sólo son capaces de leer la dirección, pero no pueden interpretar el contenido. A pesar de las nuevas funcionalidades que ofrecen los buscadores, no logran superar las limitaciones de la recuperación de la información; por un lado, la mayor parte de la información, la contenida en bases de datos, permanece oculta. Su patrón de búsqueda está formado por comparaciones de cadenas de

caracteres más o menos afinadas, pero no disponen de una estructura semántica que permita la navegación entre significados y contenidos. Por último, también es frecuentísimo el exceso de información, el denominado 'ruido'. Tal y como previó en su día **Claude Shannon**²⁰ en su *Teoría matemática de la comunicación* (1948), los ruidos y silencios son inherentes a los sistemas de información automatizada.

Y es obvio que es en este entorno en el que tienen que desplegar su actividad los archivos. Ni que decir tiene que a medida que estas instituciones se incorporan a la web difundiendo sus bases de datos, sus metadatos y documentos digitales, participarán plenamente de estos problemas, por lo que creemos que no está de más aventurar algunas de las soluciones que se están proponiendo en las tecnologías de la información.

3.1. Características de la web semántica

Dado que el nacimiento de las ontologías y de la web semántica es tan reciente, son muchas las interpretaciones y diseños que actualmente se están discutiendo. Pero si vamos estableciendo una contraposición con la web actual, se podría decir que presenta estas características:

—Los recursos no están relacionados únicamente mediante enlaces simples, sino a través de relaciones que tienen un significado, relaciones semánticas, como por ejemplo: "los presupuestos municipales se generan en Hacienda" o "Hacienda elabora los presupuestos municipales".

—A diferencia del tradicional modelo web, el contenido de las páginas, previamente estructurado, es inteligible no sólo para el usuario sino también para las aplicaciones, programas y agentes de software. El entramado de relaciones semánticas lo proporcionan las ontologías y por medio de ellas se puede navegar a significados más amplios, más restringidos o relacionados.

—A través de la estructura semántica que componen las ontologías, tanto los usuarios como los agentes de software recuperan información verdaderamente pertinente. Los sistemas informáticos son capaces de leer la información utilizando las relaciones extensibles de las ontologías y pueden proporcionar respuestas ajustadas y pertinentes, incluso no previstas por el usuario.

3.2. La construcción de la web semántica

Cualquier dominio del conocimiento o de la actividad puede ser estructurado en forma de ontología, por lo que cualquier institución histórica o actual, sus funciones, actividades y procedimientos, sus departamentos, unidades componentes, cargos y miembros,

sus reglamentaciones, relaciones, etc., pueden presentarse de esta forma como elemento nutriente de la web semántica. Es decir, podemos pensar en que un archivo, sea del tipo que sea, atesora no sólo la información contenida en sus documentos sino que también guarda en sus metadatos el conocimiento sobre un determinado área y que éste podrá ser general o sectorial.

Pongamos el ejemplo del patrimonio documental, aunque podrían ser muchos otros; en las guías, inventarios, instrumentos de descripción, tesauros, etc., existen conceptos y relaciones entre los mismos que pueden ser extraídos y estructurados en forma de ontología para compartir el conocimiento. De los metadatos de archivos digitales y virtuales, así como en otros repertorios y fuentes de información complementarios (necesarios por otra parte para la elaboración de cualquier instrumento de descripción), se pueden extraer relaciones de nombres de lugares, históricos y actuales, profesiones, tipologías documentales, instituciones, profesionales, actividades comerciales, etc., interrelacionándolas conforme a un modelo de entidad/relación²¹.

No sería difícil extender esta propuesta a una suma de ontologías del patrimonio documental, bibliográfico e histórico (por seguir con el ejemplo) que podrían conformar una posible ontología del patrimonio cultural de una determinada zona. De hecho, el *Comité Internacional para la Documentación del Consejo Internacional de Museos (ICOM-Cidoc)* viene trabajando desde hace 10 años en la elaboración de una ontología base para el intercambio de información sobre patrimonio cultural en el entorno de la web semántica. Este modelo, el *Conceptual Reference Model*²², inicialmente pensado para el mundo de los museos, está diseñado para que pueda ser utilizado por cualquier institución de memoria.

Es claro que el desarrollo de la web semántica dependerá de la disponibilidad de contenidos y del marcado sintáctico y semántico de estos contenidos, así como del desarrollo de herramientas que permitan el análisis del conocimiento. Para ello, se deberán adoptar lenguajes de representación del conocimiento, como rdf u *OWL*, para lo cual las instituciones de memoria se encuentran en una situación privilegiada respecto a otras áreas de conocimiento. En efecto, estas organizaciones estructuran la información que poseen en instrumentos tales como inventarios, catálogos, tesauros, clasificaciones y todo tipo de taxonomías. Una gran mayoría de esos instrumentos se encuentra ya automatizada²³ y un porcentaje elevado en estructuras normalizadas. Ya son menos los que han utilizado lenguajes de marcado como xml para estructurar esta información y menos aún las que emplean metadatos pa-

ra atribuir significados mediante una codificación precisa.

Sin embargo, los dos últimos pasos no presentan dificultades especiales y así, pasar de *ISAD(G)* a *EAD* no entraña una especial problemática. Sencillamente los responsables de archivos tienen que conocer y tomar la oportuna decisión para realizar las transformaciones al nuevo etiquetado. El siguiente paso que se propone es extraer los conceptos y relaciones semánticas de los metadatos y crear ontologías como procedimiento para estar presentes en la futura web semántica. Todo ello habrá de hacerse basándose en las correspondientes aplicaciones informáticas y participando en el diseño de las mismas. En nuestra opinión, se dará este paso si se percibe la importancia de cumplir las funciones archivísticas en el entorno tecnológico actual, lo que se puede alcanzar mediante recolectores y ontologías.

En la actualidad *Digibis*, con la colaboración de la *Fundación Ignacio Larramendi*, está llevando a cabo, con la financiación del *Programa Profit*, una iniciativa de este tipo, denominada “ontología de polígrafos” (*FIT-350200-2004-38*²⁴), en la que se incluye la transformación de contenidos típicos como archivos personales y epistolarios. Confiamos en que pueda tener una utilidad no sólo para nuestros propios fines sino también para las instituciones de memoria en general.

4. Conclusiones

Todas las recomendaciones, normativas y estructuras de información expuestas en este texto tienen como objetivo animar a que los archivos, y también el resto de instituciones de memoria, encaminen su actividad, en el marco de las funciones que le son propias, hacia el entorno de la creación de contenidos digitales y su difusión en la web por los mecanismos y tendencias más actuales de la misma. En definitiva se trataría de:

a. Hacer accesible internacionalmente el patrimonio cultural español:

—Crear un *corpus* orgánico digital del patrimonio documental al que en la actualidad es difícil acceder, o que no es accesible al público en general.

b. Generar contenidos en los idiomas de España para la web:

—Digitalización de documentos, piezas, etc., conversión de bases de datos existentes, conversión de fuentes de referencia consolidadas, etc.

c. Implementar sistemas de información adaptados a la normativa de la comunidad internacional vincula-

da a todo tipo de instituciones de memoria y a las tendencias de las tecnologías de la información en la web.

d. Conocer, mediante el desarrollo y uso de las correspondientes aplicaciones y sistemas de información como la recolección de metadatos o la creación de ontologías, el dominio Patrimonio Cultural para contribuir a su difusión, mantenimiento y estudio.

Notas

1. En este sentido puede leerse el artículo de **Hernández, Francisca; Wert, Carlos; Recio, Ignacio; Aguilera, Begoña; Koch, Walter; Bogensperger, Martin; Linde, Peter; Günter, Georg; Mulrenin Bob; Agenjo, Xavier; Yeats, Robin; Bordoni, Luciana; Poggi, Fabrizio**. "Xml for Libraries, archives, and museums: the project Covax. En: *Applied artificial intelligence*, 2003, v. 17, n. 8-9, pp. 797-816.

Este artículo describe una completa metodología, realizada por europeos dentro del programa *IST*, acerca de cómo es perfectamente posible analizar y diseñar las soluciones técnicas precisas para permitir el acceso, a través de internet, a descripciones de documentos de archivos, bibliotecas y museos codificadas de forma homogénea, basándose en la aplicación de sgm/xml, y crear bases de datos, utilizando tecnología avanzada. El proyecto *Covax (Contemporary Virtual Archives in Xml)* se inició en 2000 y concluyó en 2002.

<http://www.covax.org>

2. Un ejemplo de ello es *Einstein Archives Online* del *California Institut of Technology*. En este archivo están disponibles tanto los documentos digitalizados como el inventario del archivo y una base de datos con los documentos descritos individualmente.

<http://www.alberteinstein.info/>

3. Lo cual puede apreciarse en la norma *ISO 15489 Records management* dirigida a la gestión de documentos en todo tipo de instituciones. La correspondiente norma *UNE* se encuentra en fase de información pública.

4. Este artículo presenta concreciones sobre los archivos a partir de los contenidos más genéricos presentados en la ponencia "Recursos digitales: un reto para las bibliotecas nacionales". En: *Jornadas sobre bibliotecas nacionales: las bibliotecas nacionales del siglo XXI*, 2005.

<http://bv.gva.es/documentos/Ponencias/Agenjo.pdf>

5. **Weibel, Stuart L.** "Border crossings reflections on a decade of metadata consensus building". En: *D-Lib magazine*, 2005, July/August, v. 11, n. 7/8.

<http://www.dlib.org/dlib/july05/weibel/07weibel.html>

6. En general, toda la información que presenta la página principal se caracteriza por estar perfectamente ordenada y ser sumamente inteligible. Es una página muy viva que presenta con frecuencia nuevas funcionalidades o versiones más actualizadas de los distintos protocolos y estructuras de información que contiene.

<http://www.openarchives.org/OAI/openarchivesprotocol.html>

7. En este artículo, leído en abril de 2005 y redactado en el otoño de 2004, se hacían referencias a esas limitaciones: **Agenjo Bullón, Xavier; Hernández Carrascal, Francisca**. "La recolección de metadatos (metadata harvesting) y su aplicación en España". En: *IX jornadas españolas de documentación*, 2005. Madrid: Fesabid, 2005, 586 pp. Isbn 84-930335-5-3.

Además del texto de la comunicación puede verse la presentación que tuvo lugar el viernes 19 de abril.

http://www.fesabid.org/madrid2005/descargas/presentaciones/comunicaciones/fernandez_francisca.pps

8. <http://www.openarchives.org/OAI/openarchivesprotocol.html>

9. La creación de repositorios, conviene insistir, no supone necesariamente el desarrollo de un *harvester*, sino que es sencillamente una forma de roturar el campo para que terceros recolectores recuperen la información.

<http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm>

10. <http://www.loc.gov/z3950/agency/zing/srw>

11. Es muy interesante la lectura del artículo aparecido en febrero de 2005: **Sanderson, Robert; Young, Jeffrey; LeVan, Ralph**. "SRW/U with OAI: expected and unexpected synergies". En: *D-lib magazine*.

<http://www.dlib.org/dlib/february05/sanderson/02sanderson.html>.

12. Es conveniente visitar con frecuencia este sitio pues el incremento constante del número de repositorios a los que accede el recolector permite, o puede permitir, que una búsqueda que no había obtenido éxito en una consulta anterior lo consiga en una segunda sesión.

<http://oaister.umdl.umich.edu/oaister>

13. *Scirus* nació con una vocación específicamente científica y, aunque no recoge, al menos en la actualidad, tantos repositorios como *OAIster*, sí plantea una filosofía y una metodología completamente distinta, que es muy recomendable conocer.

<http://www.scirus.com/srsapp>

14. Esta dirección es la versión española del *W3C*, lo que siempre facilita la consulta, aunque con frecuencia haya que acudir a la fuente original.

<http://www.w3c.es>

15. Aunque cerrado el año pasado (2004) contiene el más amplio conjunto de menciones y declaraciones sobre la ontología. Lo considero de lectura imprescindible.

<http://www.w3.org/2001/sw/WebOnt>

16. Esta página, muy bien traducida al español, no sólo es un buen conjunto de explicaciones sobre *OWL* sino que también aclara algunos puntos sobre el concepto y la función de la ontología.

<http://www.w3c.es/Traducciones/es/SW/2005/owlfaq>

17. <http://www.eurohistory.net/Index.do>

18. Muy clarificadora resulta en este sentido la siguiente presentación:

http://www.museumscomputergroup.org.uk/meetings/1_2005_docs/A-Beginner's-guide-to-the-Semantic-Web.ppt

<http://www.vicodi.org/about.htm>

19. Sobre la bibliografía en español acerca de la ontología y la web semántica me remito a la comunicación titulada "La web semántica y las ontologías en el mundo de los archivos" (actualmente en prensa) presentada en el *VI Congreso de la Asociación de Archiveros de Castilla y León (ACAL)*, celebrado en Burgos, en noviembre de 2004.

20. Una información completa, entre las aproximadamente 64.500 páginas dedicadas a **Claude Shannon**, figura básica de nuestro tiempo, es:

<http://www-groups.dcs.st-and.ac.uk/%7ehistory/Mathematicians/Shannon.html>

21. **Agenjo, Xavier; Hernández, Francisca**. "Del catálogo colectivo a la biblioteca virtual: la Biblioteca Virtual del Patrimonio Bibliográfico". En: *I Jornadas sobre patrimonio bibliográfico en Castilla-La Mancha*, 2003.

22. En esta misma página se ofrece una codificación del modelo (versión 3.4.9) en *RDFS*, directamente importable a programas de edición de ontologías como *Protégé*. La versión 3.4.9 se corresponde, desde octubre de 2003, con la *ISO Draft International Standard ISO/DIS 21127*. Desde el 5 de julio, la versión 3.1 de este extendido editor de ontologías incluye un potente y útil editor de *OWL*.

<http://protege.stanford.edu/>

<http://protege.stanford.edu/plugins/owl/>

http://cidoc.ics.forth.gr/official_release_cidoc.html

23. **Escalona, María José; Mejías, M.; Torres, Jesús; Cordero, Juan M.; Romano, M. G.** "Aplicación integrada de la biblioteca digital del patrimonio histórico andaluz" En: *Jbidi 2000: primeras jornadas de bibliotecas digitales*, 2000. Isbn 84-8448-066-6, pp. 295-298.

http://imhotep.unizar.es/jbidi/jbidi2000/31_2000.pdf

24. Resolución de 2 de marzo de 2005, de la *Dirección General para el Desarrollo de la Sociedad de la Información*, por la que se publican las ayudas concedidas en el año 2004 del *Programa Investigación y Desarrollo*.

María José de Acuña, Digibis Producciones Digitales, Claudio Coello, 123. 28006, Madrid.

mariajose.acuna@digibis.com

Xavier Agenjo, Fundación Ignacio Larramendi.

xavier.agenjo@larramendi.es