

De las bibliotecas virtuales a las ontologías y *webs* semánticas

Xavier Agenjo

DIRECTOR DE PROYECTOS. FUNDACIÓN IGNACIO LARRAMENDI

Resumen

Una de las funciones, no la única, de las bibliotecas públicas es la de preservar el patrimonio bibliográfico y documental, una misión encomendada en régimen de exclusividad según la Ley de Patrimonio Histórico a las Comunidades Autónomas. Como consecuencia de la aplicación de esta ley y de su desarrollo, así como de la propia historia del Catálogo Colectivo del Patrimonio Bibliográfico se están produciendo importantes cambios en las implicaciones de este instrumento. Recientemente algunas Comunidades Autónomas están planteándose convertir sus catálogos colectivos en bibliotecas virtuales. El objetivo de esta comunicación es explicar cómo a partir de una biblioteca virtual puede configurarse una ontología. Se define qué es una ontología, se describe el estado de la cuestión de las ontologías especialmente a partir de las recomendaciones del W3C de febrero de 2004 en el cual se expone la posibilidad de establecer ontologías partiendo de bibliotecas virtuales, que a su vez se originan en catálogos colectivos basados en criterios tipo y topobibliográficos. Esta transformación, junto con la *ontologización* de la información propia de las colecciones locales de las bibliotecas públicas, permitiría el desarrollo de una ontología sobre el patrimonio cultural local aplicable a diferentes usos, entre ellos proyectos de desarrollo local sostenible.

Preliminar

Esta comunicación pretende orientarse especialmente al entorno de las bibliotecas públicas y encaminar sus recomendaciones y conclusiones a la Biblioteca Pública como institución perfectamente caracterizada dentro de lo que se denomina Instituciones de Memoria.

Parte de otra comunicación¹, presentada a las I Jornadas sobre Patrimonio Bibliográfico de Castilla-La Mancha y titulada *Del catálogo colectivo a la biblioteca virtual*, en la que se hacía un análisis del Catálogo Colectivo del Patrimonio Bibliográfico Castellano-Manchego como pretexto y ejemplo de lo que podría ser la transformación de un catálogo colectivo de fondo antiguo en una biblioteca virtual. En aquel trabajo se esbozaba la posibilidad de tomar como punto de partida una biblioteca virtual para construir una ontología. Por lo tanto, los puntos de desarrollo que se van a utilizar en la presente comunicación van a estar basados justamente en aquella primera idea, sometidos lógicamente al tamiz de la crítica y de experiencias posteriores.

También es importante comentar que por fechas muy similares, la Fundación Ignacio Larramendi² y la empresa DIGIBIS³ presentaron a la convocatoria de 2004 del programa PROFIT (Programa para el Fomento de la Investigación Técnica del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica [I+D+i] 2004-2007), un proyecto para la elaboración de una ontología para la *web* semántica basada en la Biblioteca Virtual de Polígrafos Españoles⁴, proyecto que con fecha de agosto de 2004, fue valorada positivamente por parte del organismo regulador. Además, los autores de esta comunicación están trabajando en otros proyectos de ontologías, menos perfilados que el que se sometió al

proyecto PROFIT, aunque ya muy avanzados y de los que seguramente se tendrá conocimiento público antes de que termine el año.

Las instituciones de memoria y la *web* semántica

En los últimos años se está produciendo un proceso de transformación de la *web* actual en la línea de superar algunas de sus deficiencias: el exceso de información desestructurada, su opacidad al contenido informativo disponible en bases de datos y el hecho de estar basado en mecanismos de búsqueda sobre palabras clave y no sobre conceptos. Esta transformación se ha reflejado en numerosas actividades destinadas a conformar lo que se denomina *web* semántica, basada en la tecnología de las ontologías⁵, en la navegación entre conceptos y relaciones entre conceptos y, sobre todo, en la creación de contenidos estructurados conceptualmente y legibles, procesables por máquinas.

Sin embargo, se han identificado una serie de retos para el avance de esta *web* semántica⁶ que pueden resumirse en que su éxito dependerá tanto de la facilidad de su uso como de la disponibilidad de contenidos variados y de calidad. A esto es imprescindible añadir la necesidad de que la futura *web* semántica disponga de contenidos en español. Para ello serán de singular importancia las bibliotecas, archivos y museos virtuales, y las bases de datos que mantienen, pues son los candidatos naturales a las transformaciones que plantea la Web Semántica.

Dicho de otro modo, la distancia entre las bibliotecas, archivos y museos virtuales y la *web* semántica es mucho menor que en ningún otro campo. Y esto es así por una serie de razones: se parte de bases de datos ya estructuradas regidas por una amplia normativa internacional (susceptibles de ser transformadas en ontologías) que representan un contenido informativo, visual y gráfico riquísimo, entrelazado, además, con las representaciones digitales de los propios documentos primarios que conservan. También contribuirá a facilitar esta transformación la familiaridad de los archiveros, bibliotecarios y museólogos con los conceptos básicos de las ontologías.

La aparición de la *world wide web*, síntesis, como es sabido, de Internet y de los lenguajes de marcado, ha transformado radicalmente el acceso a la información de todo tipo. El mundo de los archivos, bibliotecas y museos no sólo no se ha visto ajeno a esta transformación, sino que ha participado plenamente de ello redefiniendo y actualizando su normativa de intercambio. Es importante señalar cómo estos tres tipos de entidades, a los que se podría calificar genéricamente como Instituciones de Memoria, han sido capaces de adaptarse a los nuevos estándares partiendo de los ya existentes, lo cual es imprescindible si se considera el enorme volumen de registros codificados.

Para ello, las Instituciones de Memoria (archivos, bibliotecas y museos) han realizado un «mapeo» de sus estructuras de información a XML y a través de sus propios organismos de normalización han elaborado DTD y, posteriormente, *Esquemas* para ajustarse con precisión a la descripción e identificación de aquellos materiales que procesan técnicamente. Así, a partir del formato MARC se ha creado la *XML MARC DTD*⁷ y el *XML MARC Schema*⁸. En el mundo de los archivos ha nacido la *Encoded Archival Description (EAD)*⁹, también una DTD de XML, y en el mundo de los museos, después de algunos tanteos como *MUS-EAD* o *AMICO (The Art Museum Image Consortium)*¹⁰, se ha llegado al acuerdo en torno al *XML SPECTRUM*¹¹. Todo ello con el apoyo de organismos tales como la *IFLA (Federación Internacional de Asociaciones de Bibliotecas e Instituciones)*¹², la *CDNL (Conference of Directors of National Libraries)* y muy en particular con su iniciativa ICABS¹³, el *CIA (Consejo Internacional de Archivos)*¹⁴ y el *ICOM (Consejo Internacional de Museos)*¹⁵.

Se observa de forma pronunciada una convergencia de las normas pues, no en balde, se trata de DTD muy próximas entre sí, sino también un trabajo continuado en la normalización de los puntos de acceso a los datos. Los antiguos ficheros de autoridades y sistemas de indización ahora se recogen en proyectos tales como el *VIAF (Virtual International Authority File)*¹⁶ y el *EAC (Encoded Archival Context)*¹⁷.

De esta manera, archivos, bibliotecas y museos han convergido en torno a la normativa XML, no sólo para el intercambio de información de datos y metadatos, sino también para la estructuración de los mismos. Por ello, tal vez el paso decisivo que ya empiezan a plantearse las instituciones más avanzadas es el de la definición de ontologías para la *web* semántica¹⁸, basadas como el modelo prescribe en la normativa XML y RDF (Resource Description Framework), así como OWL (Ontology Web Language), y para las cuales las colecciones archivísticas, bibliográficas o museológicas están especialmente bien preparadas, ya que sus contenidos se encuentran previamente ordenados y clasificados conforme a criterios formales.

El desarrollo de las ontologías para archivos, bibliotecas o museos puede ser relativamente rápido, más que en otros sectores, por estas características tipológicas. La conversión de bases de datos bibliográficas, archivísticas y museísticas (o híbridas) pueden llegar a transformarse en ontologías para la *web* semántica si se sigue una metodología precisa¹⁹ de «mapeo» de la información conforme a una definición estructurada de aquella parte del mundo que la ontología quiere representar y aquellas preguntas que la ontología puede llegar a contestar.

Según las definiciones con las que el W3C²⁰ inicia su sitio sobre la Web Semántica, ésta «es la representación de datos en la WWW. Se trata de un esfuerzo en colaboración dirigido por el W3C con la participación de un gran número de investigadores e industrias. Se basa en el Resource Description Framework (RDF), que integra una gran variedad de aplicaciones que utilizan XML como sintáxis y URI para el nombrado.» «La Web Semántica es una extensión de la *web* actual en la cual la información se proporciona de una manera bien definida, para permitir el trabajo en cooperación de ordenadores y personas».²¹ (Tim Berners-Lee, James Hendler, Ora Lassila, «The Semantic Web», *Scientific American*, May 2001)

En el estado actual de la tecnología los desarrollos relacionados con la *web* semántica están mayoritariamente circunscritos a las universidades y centros de investigación, con un cierto número de experiencias en el proceso de creación de ontologías específicas de dominio y con desarrollos relacionados con el proceso completo de la *web* semántica. La investigación actual puede seguirse a través del sitio *web* del W3C para la Web Semántica o del sitio *Semantic Web Activity: Advanced Development for Europe*²².

Esta comunicación considera, sin embargo, la extracción (y mantenimiento) de una ontología a partir de una base de datos relacional. Las ontologías de dominio son «repositorios conceptuales» muy próximos al «modelo de datos conceptual» de las bases de datos. Ambos consisten en entidades, atributos y relaciones entre estos objetos. Además, una ontología incluye una serie de reglas y condiciones que en el mundo de la gestión de información se denomina *reglas de negocio*. Desde este punto de vista, para algunos autores²³, cuando se desarrollan modelos de datos conceptuales en esencia se están desarrollando ontologías.

El modelo de datos de la *web* semántica está directamente conectado con el modelo de las bases de datos relacionales. El mapeo es casi directo, un registro es un nodo RDF, un campo es un RDF propertyType, y un campo de un registro es un valor. Incluso, uno de los elementos conductores de la Web Semántica ha sido expresar en la *web* la gran cantidad de información contenida en bases de datos relacionales de una forma que pueda ser procesada por ordenadores²⁴.

Objetivo

Antes de precisar la razón y justificación de diseñar y construir una ontología en una biblioteca pública convendría intentar una breve definición preliminar de lo que es una ontología.

Podríamos decir que un sitio *web* como puede ser una biblioteca, que permite su acceso al catálogo y que contiene algunas informaciones adicionales, puede estar perfectamente estructurado mediante una ontología, que de esa manera se convierte en un lugar de referencia dentro de la red para todos los contenidos que puedan existir en la red acerca de determinada materia. Por eso se denominan ontologías. Así, de un portal temático pueden extraerse sus contenidos y estructurarse la información de una for-

ma ordenada y determinada, conforme a las normas de construcción de ontologías, convirtiendo la información en conocimiento.

Una metáfora, que no reúne ni con mucho todas las características, sería la diferencia que existe entre un libro que estuviera escrito en capítulos, pero que careciera de un índice. Una ontología vendría a ser como un conjunto amplio de índices de un libro y en el que los índices no sólo estuvieran formados por términos que remitieran a las páginas del libro, sino que mediante procedimientos de referencias cruzadas se interrelacionaran entre sí. El ejemplo es quizá un poco sencillo, pero nos puede permitir una primera aproximación a lo que sería la ontología. Más adelante daremos ya una definición más canónica.

Si entramos en el detalle de la ontología que se proponía para una biblioteca virtual sobre patrimonio bibliográfico, lo cual nos sirve de idea matriz a su vez para la constitución de una ontología para una biblioteca pública, vemos que en aquel momento recomendábamos que dentro de la información que podía ser recogida en el sitio *web*, cuyo elemento fundamental es la propia base de datos bibliográfica, se produjese un proceso de transformación de dicho sitio para superar algunas de sus deficiencias o carencias en la recuperación de información, que se reflejan siempre en una insatisfacción del lector, y alcanzar un mayor índice de éxito.

Estas carencias se deben fundamentalmente a que estamos trabajando con una información parcialmente estructurada y sin conexión explícita, que oculta gran parte del contenido informativo y de las relaciones conceptuales de las que se dispone en la base de datos, y sobre todo, al hecho de que la gran mayoría de los sistemas de búsqueda y recuperación de información están basados en la comparación de patrones de palabras, palabras clave, y no sobre conceptos. Por todo ello, se ha llegado a un proceso de transformación de la actual *web*, reflejado en numerosas y distintas iniciativas, que tiene como objetivo la creación de ontologías que sustenten una *web* semántica, es decir, una nueva *web* que permita la navegación entre conceptos relacionados, y en la creación de contenidos estructurados conceptualmente, legibles y procesables por máquinas. La definición que ya daremos más adelante de la *web* semántica es muy precisa.

¿Qué contiene una base de datos de fondo antiguo? Contiene descripciones de registros bibliográficos realizados conforme a una normativa bibliográfica que está en pleno proceso de cambio. Gracias a los FRBR, recientemente traducidas al español²⁵, poseemos un método de análisis radicalmente distinto de los Principios de Catalogación de París o de las ISBD, que han sustentado durante décadas las reglas de catalogación, y que definen lo que debería contener un registro bibliográfico.

En esas descripciones bibliográficas están incluidas implícitamente determinadas relaciones de las que hacemos un uso parcial en las búsquedas. Por ejemplo, podemos saber, a base de analizar determinados registros, si un impresor desarrolló su actividad en uno o varios lugares, si un autor escribió una determinada obra, si una obra fue escrita por un determinado autor, etc. Podemos llegar a saber que existen relaciones entre los autores a través de las relaciones entre registros de autoridad, pero esta información no está suficientemente calificada, sólo nos remiten a otros nombres de un autor o a otros autores relacionados, y no nos indican el tipo de relación. En una ontología tanto los conceptos como sus relaciones son explícitos, la información está estructurada para que así sea, y permiten establecer tantos conceptos y relaciones como sean necesarios para un dominio específico del conocimiento.

Podríamos preguntar a una ontología quién tenía una imprenta en determinado lugar, cómo se llamaba esa imprenta, cuántos años estuvo establecida, qué obras imprimió, quiénes eran sus autores, quiénes eran sus traductores, de qué idioma fueron traducidas, cuánto costaban los libros que allí se producían, qué tirada tenían, qué repercusión alcanzaban, etc. Es decir, en lugar de proporcionar unos datos fijos y estáticos fruto de la descripción bibliográfica, a través de la cual el investigador posteriormente elabora una tesis, su objetivo es sencillamente conseguir que un sistema sea capaz de contestar a esas preguntas porque esas relaciones se pueden establecer desde el principio en forma de ontología.

Hay que tener en cuenta que hablando del fondo antiguo, y seguimos en el marco de una biblioteca virtual de patrimonio bibliográfico, el conocimiento del dominio está disperso entre muchas fuentes, no sólo bases de datos bibliográficas o de autoridades, sino también otras fuentes de información sobre el libro antiguo, artículos, manuales, estudios, tipobibliografías, imágenes de las propias obras, imágenes de las marcas de agua, etc. Lo que se propone aquí es extraer este contenido de las fuentes de información reconocidas y transformarlo en una ontología.

Efectivamente, el conocimiento del dominio está contenido en esas fuentes. Realmente cuando se repasan la gran cantidad de estudios que se llevan haciendo sobre distintos aspectos de autores, títulos, obras, impresiones, ediciones, etc., se puede ver que todos ellos son constitutivos de una categoría conceptual que se interrelaciona con otras; es decir, determinado autor se relaciona con otros autores, bien porque es el coautor de un libro, o bien porque es el prologuista de otro libro, o bien porque es el censor, o bien porque ese libro está en poder de él ya que figura en el inventario de su biblioteca. Podemos ver que una obra tuvo sucesivas expresiones y manifestaciones, según los términos acuñados por los FRBR, ya que la obra puede estar traducida, puede estar adaptada por otro autor, puede haberse introducido o publicado en determinado periodo de tiempo, lo que constituirían nuevos parámetros de categorización. Igualmente, se puede hacer saber, y extraer, que en determinado lugar de impresión y durante determinado periodo se publicaban fundamentalmente libros de tal o cual materia porque podemos combinar la información del campo de edición/impresión con la información de materias, podemos, así mismo, trazar los datos referentes a las imprentas de una localidad precisa y ver qué tipos de libros se escribían, qué tipo de libros se traducían, qué libros eran adaptados, con qué editores o costeadores se relacionaban, durante qué periodo trabajaban, etc.

Es decir, podemos establecer relaciones semánticas entre todos los aspectos que configuran esa información. Lo mismo podemos decir de los grabados que ilustran libros, xilográficos o calcográficos, cuál fue su procedencia, quién vendió el papel, quién vendió el grabado, quién lo compró, etc., estableciendo así los circuitos de lectura de las obras. Los mapas geográficos de distribución de títulos por bibliotecas darían verdaderas sorpresas sobre la difusión de los libros, sobre todo si además se toman no sólo en su conjunto sino que se distribuyen por materias, por editores, por impresores, por lugares de impresión, por autores, etc. Podría observarse que en determinados lugares puede existir una predominancia de obras de autores de un periodo, o predominar obras de una materia concreta, o bien que los libros producidos por un impresor aparecen de forma preeminente en un lugar.

Todos los datos reflejados en los campos de fondos y localizaciones permiten que el catálogo colectivo se convierta en una impresionante fuente de información sobre la sociología de la lectura, la historia de la imprenta, de la edición, e incluso de la cultura en términos generales. Esto nos lleva a una conclusión muy clara, podemos multiplicar considerablemente la capacidad informativa de nuestra base de datos, de nuestra biblioteca virtual (que incorpora además de la descripción el propio documento primario en forma de objeto digital), y podemos además explotar toda la información que contienen las descripciones bibliográficas, agrupándola por conceptos interrelacionados mediante los mecanismos informáticos referidos de clases conceptuales, para poderlos recuperar de una manera más eficaz. Así, el catálogo colectivo se convierte no sólo en una biblioteca virtual sino que puede transformarse en una ontología en torno al amplio campo del libro. Exactamente, el libro, la imprenta, su difusión, la lectura, es decir, conceptos básicos de investigación de la sociología material y del estudio de las corrientes espirituales, de la Gestalt, a partir de una información que habitualmente se utiliza para identificar, describir y catalogar, y que está dispuesta para controlar la circulación de las obras de un catálogo y poco más.

Transformar esta información implícita en explícita supone una revolución metodológica, puesto que al introducir los datos se nos pedirá la cumplimentación de las relaciones que se han definido previamente en la ontología, permite definir para cada obra, manifestación, expresión e ítem, no sólo los datos de identificación propia, sino fundamentalmente las relaciones entre ellas.

Naturalmente, la transformación de una biblioteca virtual, o un catálogo colectivo, sobre patrimonio bibliográfico en una ontología definiría el ámbito temático de ésta y su objeto sería, evidentemente, el libro español impreso desde los inicios de la imprenta en Aguilafuente, provincia de Segovia, hasta 1958 en que el depósito legal se convierte en un instrumento seguro de control bibliográfico.

Dadas estas especulaciones podríamos entrar en otras paralelas sobre cuál podría ser la ontología que se trazase a partir de una base de datos, quizá ya convertida en una biblioteca digital, de una biblioteca pública cualquiera. Lo característico de una biblioteca pública, para el asunto que aquí estamos tratando, lo que la convierte en un punto de referencia dentro de la futura red semántica sustentada en ontologías, es la concentración especialmente elevada del fondo local. Esta concentración de fondo local se debe fundamentalmente a dos circunstancias, ser receptoras del depósito legal y haberse constituido para recibir los fondos procedentes de la desamortización. Es fácil observar que las obras impresas en una determinada zona se encuentran más fácilmente en las bibliotecas de la propia zona que en otras de puntos más alejados. Esto incluye el importante periodo de la historia española que va desde la desamortización hasta 1958.

Estamos hablando de un periodo de 125 años aproximadamente en los cuales el control bibliográfico es menor, que coincide, por tanto, con una serie de carencias importantes en cuanto a repertorios bibliográficos coetáneos, pero también, y por desgracia, en cuanto a repertorios bibliográficos actuales. Son muy frecuentes las tipobibliografías de los XVI-XVII, en algunos casos del XVIII, pero del siglo XIX, con alguna excepción que merece la pena citar como *A imprenta en Galicia: século XIX*, publicado en la colección *Bibliofilia de Galicia* bajo la dirección de Ignacio Cabano, son particularmente escasos. Si a esto le unimos que es en este periodo cuando se empiezan a producir los libros con papel de pasta de madera, invento realizado hacia 1840, y que, por lo tanto, los impresos adquieren con el tiempo un pH ácido que hace que se destruyan paulatinamente, resultará que las esperanzas de llegar a conocer un periodo sustancial de la historia de España son cada vez menores. Existen proyectos de digitalización como el de la Prensa Histórica²⁶, que lleva a cabo la Dirección General del Libro, Archivos y Bibliotecas, que, sin duda, supone una salvaguardia, pero que necesita de una inversión superior a la actual para evitar la destrucción, por efecto del papel ácido, de una masa importantísima de información. Por otra parte, ya sabemos que hay títulos irremisiblemente perdidos²⁷. Por otro lado, este periodo viene a coincidir, y aquí entraríamos en un aspecto más político, con la creación de la identidad cultural de la España contemporánea, que está basado, como muy bien mostró Caro Baroja, no tanto en ancestrales fuentes documentales históricas, sino más bien en la interpretación de las mismas que se llevó a cabo en el siglo XIX o, incluso en la propia redacción de las mismas en el siglo XX.

Así pues, ¿cuál sería el objeto de las ontologías que podrían realizar las bibliotecas públicas, ya sean las Bibliotecas Públicas del Estado, o bien otras bibliotecas públicas o municipales? No vamos a ser tan ingenuos como para creer que cualquier biblioteca pública o municipal puede dedicarse a establecer una ontología, pero sí que la utilización de las reglas metodológicas de una ontología puede ayudar a evaluar y conformar los recursos bibliográficos de la propia colección y de las propias obras de referencia que contiene incluso una biblioteca mediana.

Una de las ventajas de las ontologías trazadas conforme a las normas es que son autoconsistentes, lo que nos llevaría a que desde el mismo momento de su diseño nos daríamos cuenta de las lagunas que pudieran existir sobre cualquier aspecto de la información local en nuestra biblioteca tradicional, automatizada o virtual. Se detectarían las faltas de diccionarios biográficos, con una serie de datos que no suelen incluir, los diccionarios geográficos, los recursos patrimoniales, las tipobibliografías, que como ya se ha dicho podrían ser junto con el catálogo colectivo de patrimonio bibliográfico el elemento básico para la formación de una ontología concreta sobre el libro. En esa conformación de la ontología realizaríamos un cruce entre lo que las fuentes de información recogen y los recursos que verdaderamente tenemos. Evidentemente, la solución para aquellos recursos que conocemos bibliográficamente y

que, sin embargo, no poseemos en nuestras bibliotecas sería la digitalización de esos fondos para incorporarlos a los recursos propios y hacerlos accesibles a través de la ontología, o bien la biblioteca virtual, en los cuales la información no es poseída por la biblioteca sino como una referencia electrónica. De esta manera, las bibliotecas tendrían una metodología abierta, pero muy concreta, muy limitada por un conjunto de estándares que permitirían una redefinición de la biblioteca.

Claramente el objeto es el fondo local, los recursos patrimoniales y bibliográficos de esa localidad. Como se planteaba en un proyecto europeo que no llegó a ver la luz, *Genius Loci*, es decir El Genio del Lugar, se trata de las fuentes bibliográficas, hemerográficas, tan importantes y tan en franco retroceso, las fuentes archivísticas y museológicas, repertoriadas en inventarios y catálogos o accesibles directamente mediante la digitalización. Se trataría de configurar una ontología con la reunión en una base de datos de los componentes, clases, atributos y relaciones, de las fuentes bibliográficas y documentales sobre el patrimonio de todo tipo de una zona. En realidad, con todo ello se forma una especie nicho, no digamos ya ecológico, que también, patrimonial en el cual los diferentes tipos de patrimonio, bienes muebles, inmuebles, etnografía, ecología, etc., la cultura de un lugar en suma, estarían integrados en una ontología que recogería en una base de datos, diseñada específicamente para la *web* semántica, todas las informaciones locales. Los datos contenidos en esta ontología servirían no sólo como herramienta para las búsquedas semánticas, o para evaluar las colecciones de las secciones locales de las bibliotecas públicas como se ha dicho, sino que serían al mismo tiempo un mecanismo de ayuda para la toma de decisiones en diferentes tipos de actuaciones de patrimonio y territorio, incluso en proyectos de desarrollo sostenible de los que poco a poco se van dando casos a lo largo de toda Europa, itinerarios culturales, turismo cultural, recuperación de oficios tradicionales, etc. Todo ello requiere de una fuente de información perfectamente recopilada y estructurada, es decir de una ontología, que sustente e informe este tipo de actuaciones.

Metodología

El mero planteamiento de una ontología en las circunstancias actuales de las bibliotecas públicas, incluso aunque no sea más que para allegar los materiales para ello, nos puede ser enormemente útil. Puede que una biblioteca pública no sea capaz por sí misma de realizar el esfuerzo, además de las otras tareas que tiene asignadas, que supone la creación y alimentación de una ontología, pero, además, por las propias características de las ontologías no tiene sentido su desarrollo en parcelas estancas, ya que se trata de la conceptualización de un dominio común, motivo por el cual, y sobre una estructura común —quizá auspiciada por niveles administrativos superiores—, se pueden ir juntando los esfuerzos de pequeñas redes de bibliotecas locales para tener reunidas y estructuradas en una ontología todas las informaciones sobre ese lugar, por las razones antedichas. Recopilan materiales bibliográficos de temática local, o los reciben a través del depósito legal, y tienen una mayor facilidad para adquirir la información local, por la herencia de las adquisiciones de muchos años: la suscripciones a periódicos locales, que no suelen aparecer en ninguna otra localidad; la conservación de materiales efímeros, como grabados, carteles, tan ricos informativamente; la incorporación de tesinas y tesis que sobre esas localidades se llevan a cabo en las universidades de la zona y que contienen información de primera mano, que circula muy escasamente, sobre los fondos específicos de un lugar concreto. Es decir, una ontología viene a ser el lugar perfecto de reunión de todos los recursos de los que se dispone para constituir una base de datos avanzada sobre el patrimonio que no sólo marca la ley, sino que son necesarios para un futuro desarrollo sostenible.

Instrumentos y herramientas

Para la construcción de este tipo de ontologías existen una serie de instrumentos y herramientas que resultan particularmente apropiadas y cuya familiaridad y manejo a la hora de definir una ontología son muy útiles porque el diálogo con el instrumento nos proporciona una gran capacidad de análisis y

definición de la realidad. La división en átomos informativos que a su vez pueden interconectarse entre sí e interrelacionarse mediante los lenguajes citados previamente es mucho más fácil de realizar con estas herramientas. Proyectos como Protégé²⁸, que contiene una gran cantidad de funcionalidades, o Kaon²⁹ pueden suministrar información sobre cómo edificar una ontología³⁰. Siempre teniendo en cuenta que cuando se construye una ontología no se parte en modo alguno de cero, sino que se parte de la base de datos bibliográfica, del propio catálogo, que puede haberse transformado en una biblioteca virtual con recursos que no se encuentran en la propia biblioteca, o de esa pequeña red de bibliotecas locales.

Existen todo tipo de casos, pero es evidente que determinada prensa de un lugar concreto puede no encontrarse necesariamente en su localidad, sino que puede localizarse en la Biblioteca Nacional, en la Hemeroteca Municipal de Madrid, en la Casa del Arcediano de Barcelona o en la riquísima biblioteca que para el siglo XIX supone el Servicio Histórico Militar, y toda una serie de instituciones de carácter civil como son ateneos, círculos de recreo, círculos de lectura, instituciones económicas, casinos, cámaras de comercio; o las inmensas fuentes de información que son las bibliotecas y archivos de los obispados, y que no podrán formar parte de la biblioteca pública que lleva a cabo la ontología más que de una forma virtual, es decir formando parte de un catálogo colectivo que haya digitalizado justamente esos fondos y que los haya referenciado.

También formarán parte de esa ontología todas aquellas publicaciones que sobre el objeto de la ontología se hayan publicado. Como en este caso el objeto de la ontología es el propio lugar, es la propia colección local, serán esenciales todas las publicaciones académicas o no que sobre el lugar se hayan producido. Es decir, toda la recopilación de la colección local, de forma directa o de forma indirecta, se llevará a cabo no con el objetivo de almacenar o conservar en la biblioteca los propios materiales, sino para ser introducidos en una base de datos concebida según los principios de la ontología y convertirse así en el punto de referencia esencial para la localidad específica en la cual esté situada la biblioteca pública en cuestión y desde donde se lleve a cabo la actividad específica de esa biblioteca.

Hay que tener en cuenta que una biblioteca pública, aparte de ser el punto de referencia para la información de la localidad a la que sirve, debe llevar a cabo esta acción porque supone la adopción de una metodología y de unas pautas de trabajo claras a la hora de la recopilación de información, y porque además un porcentaje muy elevado, en torno a un tercio, de los usuarios de una biblioteca usan la colección local. Esta información es utilizada por el característico erudito local o los escolares que llevan a cabo estudios de conocimientos del medio, la historia o literatura de un lugar. Si constituimos una ontología verdaderamente eficaz también puede ser utilizada en proyectos de desarrollo basados en la cultura y en el patrimonio que encontrarán, justamente, en las ontologías realizadas con esas premisas y con esos objetivos toda la información necesaria.

Normalización

Ya se han citado las líneas directrices de la normalización que está llevando a cabo el W3C a través de sus recomendaciones. Sin embargo, no estaría de más mencionar el hecho de que la ISO (International Standardization Office) está preparando una ontología de referencia para el intercambio de información sobre patrimonio cultural. En efecto, el Comité 46, al cual se le deben tantísimos avances en el campo de la tecnología de la información, está transformando en norma ISO el Conceptual Reference Model³¹ elaborado por el ICOM-CIDOC (The International Committee for Documentation of the International Council of Museums), que se encuentra en estos momentos en la fase Draft (ISO/CD 27127). Se publicó en agosto del 2002, para formalizar el dominio «patrimonio cultural» y pretende cubrir todos los conceptos que son relevantes para esta información, especialmente aquellos que se necesitan en el amplio campo del intercambio de información entre museos, bibliotecas y archivos.

El CIDOC Conceptual Reference Model proporciona tanto definiciones como una estructura formal que permita la descripción de los conceptos implícitos y explícitos, así como sus relaciones, utilizados en patrimonio cultural. Es un marco semántico para cualquier tipo de información sobre patrimonio cultural y para que cualquier tipo de información sobre este tema pueda ser convertida a este modelo de referencia. Ya hemos dicho que la información que se necesita cubrir, para tener disponible una ontología preparada por una biblioteca pública con la información procedente de la colección local o sobre el lugar o territorio en el que se encuentra esa biblioteca y ejerce su influencia y labor, estará complementada por la información recogida por los museos y archivos de la misma localidad, o incluso por otros centros como los centros de interpretación, que se dedican a elaborar todo tipo de información sobre el lugar. Por supuesto, estamos hablando de la necesidad de establecer correspondencias entre Dublin Core, Spectrum, MARC, y el modelo de referencia que es el CRM.

El Conceptual Reference Model es, por tanto, uno de los documentos más importantes, cuya traducción al español se va a emprender próximamente, y que tendrá una mayor influencia en un futuro próximo. Conviene citar que existe un comité conjunto entre IFLA e ICOM que pretende llevar a cabo una armonización del FRBR y el CRM, puesto que sus objetivos no son tan diversos. Hay que estar atentos a los trabajos que vaya produciendo el mencionado Comité 46 y a sus grupos de trabajo 4 y 9. Por otra parte, se está preparando, a través de otro subcomité, el aspecto de la interoperabilidad técnica entre las aplicaciones informatizadas en información y documentación en colaboración con el CIDOC.

Este estándar pretende cubrir todos los conceptos relevantes para la información patrimonial cultural, pero sobre todo aquellos que se refieren al intercambio de información entre museos, bibliotecas y archivos. Se cree que sólo mediante extensiones de este estándar, dada la diversidad de tipos de materiales, se podrá llegar a una norma de amplio alcance. Es muy importante hacer hincapié en la gran calidad del CRM, en el gran número de entidades conceptuales, atributos y relaciones, y en que esas entidades están definidas en la jerga informática dentro del concepto entidad-relación. El CRM posee, además de estas características, otras que definen a una buena ontología, cubre el dominio completo, es ampliable y ofrece un marco de referencia fiable elaborado por una entidad con autoridad en la materia.

Excurso: el *Web Harvesting*

Cuando hemos elaborado una serie de criterios para la recogida de información local como puede ser los fondos que proceden de la desamortización, o bien aquellos fondos que proceden del depósito legal, o bien fruto de una política de adquisición de obras de referencia, o por medio de las relaciones con otras instituciones que pueden generar información sobre el lugar³², nos encontramos con que existe un procedimiento que debe ser implementado de forma paralela, o en fases sucesivas, a la ontología que es el *Web Harvesting*.

El *Web Harvesting* es uno de los procedimientos de rastrear la red y recuperar información, no ya de páginas estáticas, sino de bases de datos que cumplen determinados protocolos. Conviene tener presente que desde que se constituyó ICABS el *Web Harvesting* es uno de los procedimientos promovidos por este organismo, en concreto forma parte de su objetivo 3, acción 3³³. En último término, se trata de implementar piezas de *software* por medio de las cuales se podría cosechar o recolectar la información existente en otras bases de datos sobre el lugar en cuestión. La OAI-MHP (Open Archives Initiative- Metadata Harvesting Protocol)³⁴ y la tecnología Dublin Core muestran que existen ya importantes repositorios, crecientes en número (véase el reciente acuerdo³⁵ del Museum, Libraries and Archives Council (MLA) inglés para rastrear las bases de datos que están accesibles en la red y que tienen implementado a nivel de *software* de aplicación el OAI-MHP y que cuentan con un mecanismo de generación de metadatos en Dublin Core).

Evidentemente, esto puede ser un procedimiento combinado con el que podríamos denominar manual, de adquisición de obras por compra o por depósito legal de recursos tradicionales y electrónicos de fondos locales.

Conclusiones

Entre los servicios que la biblioteca pública tiene definidos está el desarrollo de la colección local y la atención al público local. En este sentido, la ontología se inscribiría en un marco doble, por un lado, tendríamos el aspecto de instrumento para cumplir uno de los fines que se fija entre los objetivos estratégicos de la biblioteca pública, el desarrollo de la colección local y, por otro lado, dar satisfacción a uno de los programas más activos en las bibliotecas de todo el mundo como es el «bibliotecario responde». Cuando el bibliotecario hubiera de responder a preguntas relacionadas con la colección local acudiría lógicamente a la ontología, o bien la ontología sería consultada directamente por los usuarios como una herramienta especializada de búsqueda semántica.

Por lo tanto, el beneficio sería doble, por un lado, contaríamos con recursos automatizados que responderían cualquier pregunta sobre la colección local de la biblioteca y sobre el patrimonio cultural local, y por otro lado, una línea de actuación que permitiera ir creando aquellos instrumentos que la ontología demanda y que son imprescindibles para que la trabazón conceptual que se precisa.

Incluso, dando un paso más allá, podemos ver una aplicación que sería muy bien recibida por los poderes públicos pues daría a la biblioteca un verdadero significado de actuación social, que ya tiene en muchos casos, pero no de forma tan explícita, en la que, según la terminología de los economistas se denomina el *back-office* del modelo de negocio, que permitiría que la biblioteca tuviera, a través de su ontología, perfectamente ordenados una serie de recursos informativos necesarios para emprender proyectos, especialmente, de desarrollo sostenible para el cual hace falta un conjunto grande de informaciones sobre el patrimonio y sobre el territorio, que se encuentran en la biblioteca, aunque no ordenadas o dispuestas de modo tal que puedan ser utilizables para proyectos de este tipo. En ese sentido, la biblioteca cumpliría una nueva función social como uno de los referentes del desarrollo y se vería claramente enriquecida su función. También disminuirían las intermediaciones a quien va a la biblioteca para recoger información para proyectos e iniciativas de desarrollo, pues las podría obtener directamente. Sin duda alguna, podría jugar sus cartas de una manera fuerte para que se considerase que la elaboración de ontologías locales centradas en un territorio concreto, en una ciudad determinada, en una circunscripción precisa, fueran especialmente valoradas por la sociedad civil o por la administración municipal, autonómica, o incluso central, que verían a su vez en la biblioteca un conjunto de recursos informativos ordenados y que proporcionan un caudal de información interrelacionada a varios niveles que permite llevar a cabo planificaciones y proyectos de todo tipo.

NOTAS

1 En prensa.

2 <http://www.larramendi.es>.

3 <http://www.digibis.com>.

4 <http://www.larramendi.es/BVMPELAYO.htm>.

5 Algunos ejemplos de sitios *web* semánticos basados en ontologías pueden encontrarse en Gene Ontology Consortium www.geneontology.org; United States National Library of Medicine, Unified Medical Language System (UMLS) <umlsks.nlm.nih.gov>; y Mnemosyne Image Atlas <www.mnemosyne.org>.

6 *Six Challenges for the Semantic Web* / V. Richard Benjamins, Jesús Contreras. April 2002. www.isoco.com/isococom/whitepapers/files/SemanticWeb-whitepaper-137.pdf.

7 <http://www.loc.gov/marc/marcxml.html#marcdtd>.

8 <http://www.loc.gov/standards/marcxml/>.

9 <http://www.loc.gov/ead>.

10 www.amico.org.

11 <http://xml.coverpages.org/CIMI-SupportingDoc15.pdf>.

12 www.ifla.org.

13 <http://www.ifla.org/V1/7/icabs.htm>.

14 <http://www.ica.org>.

15 <http://icom.museum>.

16 www.oclc.org/research/projects/viaf/default.htm.

17 www.library.yale.edu/eac.

18 www.w3.org/2001/sw.

Recientemente ha surgido el proyecto MarcOnt para la creación de un conjunto de herramientas que permitan la traducción entre MARC 21 y Dublin Core y ontologías destinado a la mejora de los procesos de búsqueda y recuperación de información en el entorno de la Web Semántica. MarcOnt está directamente relacio-

- nado con Elvis Digital Library, e-Library with Semantics <http://elvis-dl.sourceforge.net>.
- ¹⁹ www.w3.org/2001/sw.
- ²⁰ *The Semantic Web* / Tim Berners-Lee, James Hendler, Ora Las-sila // Scientific American. May 2001.
- ²¹ www.w3.org/2001/sw/Europe.
- ²² *Mapping Databases to Ontologies: How to build fully automa-ted query and reporting systems* / by Marvin Elder. <http://www.dmreview.com/whitepaper/WID517.pdf>.
- ²³ *Relational databases on the Semantic Web* / Tim Berners-Lee. - www.w3.org/DesignIssues/RDB-RDF.html.
- ²⁴ Federación Internacional de Asociaciones de Bibliotecarios y Biblio-tecas. *Requisitos Funcionales de los Registros Bibliográficos: Informe final* / Grupo de Estudio de la IFLA sobre los Requisitos Funcionales de los Registros Bibliográficos [Aprobado por el Comité Permanente de la Sección de Catalogación de la IFLA]; Traducción de Xavier Agenjo y María Luisa Martínez-Conde [Madrid?]; Fede-ración Internacional de Bibliotecarios y Asociaciones de Biblio-tecas: Dirección General del Libro, Archivos y Bibliotecas del Minis-terio de Cultura, [2004] 239 p. I.S.B.N.: 84-8181-213-7.
- ²⁵ www.mcu.es/pruebaprensa.
- ²⁶ Véase *El monje digital* / Xavier Agenjo Bullón. En prensa.
- ²⁷ The Protégé Ontology Editor and Knowledge Acquisition System. <http://protege.stanford.edu/>.
- ²⁸ KAON - The Karlsruhe Ontology and Semantic Web Tool Suite. http://www.aifb.uni-karlsruhe.de/Forschungsgruppen/Projekte/viewProjekt?id_db=26.
- ²⁹ Véase *Editing ontologies: a survey of editing tools* / by Michael Denny. <http://www.xml.com/pub/a/2002/11/06/ontologies.html>.
- ³⁰ http://cidoc.ics.forth.gr/official_release_cidoc.html.
- ³¹ En este sentido sería muy conveniente contar con un buen el elaborado por la Institución Fernando el Católico de Aragón con más de 2000 títulos publicados y no necesariamente impresos en Aragón, con lo cual no es seguro que aparezcan en la base de datos de Depósito Legal de esta Comunidad.
- ³² *Influencia del ICABS en el futuro digital de las Bibliotecas* / Xavier Agenjo y Francisca Hernández. Comunicación presentada al Foro Biblioteca y Sociedad: experiencias de innovación y mejora. Murcia, 14-16 de octubre de 2004.
- ³³ <http://www.openarchives.org/>.
- ³⁴ <http://www.mla.gov.uk/news/snippets.asp?print=print&month=8&year=2004>.