

Organización del conocimiento en Internet: Metadatos bibliotecarios Dublin Core

Organización del Conocimiento en Internet: Metadatos bibliotecarios Dublin Core

Rosa San Segundo Manuel

*Profesora titular de Universidad en el área de Biblioteconomía y Documentación
Universidad Carlos III de Madrid*

1. Concepto de metadatos

Los metadatos o datos representacionales son definidos como el dato sobre los datos¹, es un conjunto de elementos que poseen una semántica comúnmente aceptada, o sea tratan de representar la información electrónica tan dispersa y representan a la descripción bibliográfica de recursos electrónicos. Estos datos abarcan ámbitos tanto individuales como colectivos, también documentos, recursos de Internet e incluso objetos reales. Nace de la necesidad de recuperar la información electrónica tan dispersa. Los metadatos tratan, principalmente, de describir el contenido y la localización del objeto de la información en Internet.

Una de las características más importantes de los metadatos va a ser su capacidad de relación o de establecer enlaces. De esta forma se han hecho imprescindibles en la recuperación global de la información en Internet, puesto que se trata de indizar y clasificar incommensurables cantidades de información de diversos tipos. Se tratará de integrar de forma heterogénea fuentes de información muy diversas, así como integrar diferentes formatos de bases de datos. O sea, se emplean metadatos para organizar el contenido de la información en Internet. De esta forma en la definición de metadatos podemos incluir y de forma más específica, se refiere a la información accesible por Internet. Por lo que los metadatos tienen el objetivo primordial de que los documentos introducidos en la Red incluyan todos los datos necesarios para su posterior búsqueda, localización y recuperación. Ya que introducir o publicar dentro de Internet es una tarea sencilla, sin embargo la localización, control y uso de la información es una tarea más compleja. Por tanto, será una tarea primordial establecer las normas y elementos que ha de contener cualquier descripción y catalogación de recursos en Internet.

Un catálogo de biblioteca o un repertorio bibliográfico son tipos de metadatos². Estos tipos de metadatos emplean, fundamentalmente, reglas de catalogación y formatos para transmitir la información como los formatos MARC. Por lo que nuestra primera idea de metadatos van a ser los catálogos bibliotecarios y bibliográficos. O sea, cada ficha es un metadato de un libro o bien de un autor y los metadatos proporcionan una información básica sobre las obras de un autor y lo relaciona con otras obras del mismo autor u otras obras de similar contenido. Lo que hasta ahora venía denominándose descripciones bibliográficas o registros bibliográficos hoy día van a ser denominados metadatos o sea que tienen como objetivo la descripción de los recursos de Internet.

Existen varios gestores de metadatos que tratan de unificar el mapa representado para cada documento, los elementos de los datos y la conversión de varias sintaxis en una sola. El movimiento de metadatos en Internet trata de integrar distintos formatos de metadatos de las bases de datos para ser integrados conjuntamente aunando el legado de los catálogos automatizados de las bibliotecas y una estructura de catálogos electrónicos o también denominados catálogos hipertextuales, donde su idiosincrasia radica, no solo en las formas tradicionales de acceso, sino en la propia estructura del hipertexto con enlaces tanto para la clasificación sistemática como para la alfabética e incluso para toda la descripción bibliográfica. Ya que se trata de una estructura articulada con distintos tipos de enlaces. De esta forma los servicios y fuentes de los catálogos electrónicos van a estar accesibles también a través de los denominados buscadores y de las propias fuentes de las páginas Web. Puesto que la nueva estructura de las bases de datos es accesible a través de las páginas Web, supone que las usuales bases de datos transformen su propia estructura.

Así la arquitectura navegable y jerárquica reporta a los diferentes metadatos de distintos formatos para que converjan en uno, y además posibilita establecer una estructura de enlaces que los haga accesibles. Exis-

ten varios modelos de metadatos, pero en la aplicación bibliográfica y bibliotecaria se ha extendido e implantado, de forma mas mayoritaria, el formato denominado "Dublin Core" o Círculo de Dublin, creado por las iniciativas de las asociaciones de bibliotecarios americanos, y en concreto por OCLC (On Line Computer Library C). Se trata de un formato bastante standard para las fuentes de Internet, originariamente bibliográficas y bibliotecarias. Es un formato de metadatos, basado en la asociación de superenlaces, y estableciendo mapas semánticos similares a los elementos y estructuras con metadatos standares. O sea, se trata de un sistema de conversión de metadatos que abarca y contiene metainformación, esta conversión necesita todavía de la intervención humana e identifica y enlaza las páginas Web. En definitiva, es un formato muy simple que incluso puede ser aplicado por catalogadores no muy expertos.

Los metadatos Dublin Core tratan de ubicar, en el entramado de Internet, los datos necesarios para describir, identificar, procesar, encontrar y recuperar un documento introducido en la Red. Si este conjunto de elementos Dublin Core se lograra aceptar internacionalmente supondría que todos los robots que indizan documentos en Internet encontrarían, en la cabecera de los mismos, todos los datos necesarios para su indización y además estos datos serían uniformes. La eficacia de estos robots como Altavista, Yahoo y otros mejoraría notablemente. Si el Dublin Core lograra a ser operativo los metadatos de la cabecera de los documentos facilitarían su indización automática y mejorarían la efectividad de los motores de búsqueda³

2. Evolución del concepto de metadatos

Con la pretensión de crear identificadores y localizadores que organicen y sistematicen la heterogénea información que alberga Internet han surgido varios intentos de normalizar esta información tan dispersa. El primero ha sido la TEI (Text Encoding Initiative) o texto codificado, se trata de un proyecto internacional con el objeto de establecer las normas para intercambio de documentación electrónica válida para fines investigadores en el marco de las instituciones académicas. Se trata de normas para codificar o marcar texto y posteriormente procesarlo automáticamente, además incluye una cabecera que comprende información sobre el documento electrónico o metadatos. La cabecera TEI o cabecera de texto codificado aporta datos útiles para poder describir un documento, y recuperarlo con el auxilio de los motores de búsqueda, supone un paralelismo, aunque no tan ortodoxo, de las normas de descripción bibliográfica. Los datos de las cabeceras TEI, en muchos casos, son aprovechados por las bibliotecas que adquieren los documentos, o pueden acceder a ellos. De esta forma se ha tratado de hacer conversión de cabeceras TEI al formato MARC, esto ha presentado serias dificultades y han sido los bibliotecarios quienes han demandado una mayor normalización de las cabeceras TEI, y en concreto de los bibliotecarios norteamericanos que son quienes han empleado el sistema. Sin embargo estos documentos electrónicos no tienen como único usuario y fin las bibliotecas por lo que su descripción aunque deba normalizarse no debe complicarse. Además, en el ámbito bibliotecario internacional existe un consenso en lo que hace referencia a los elementos de la descripción bibliográfica pero no en los puntos de acceso. La cabecera TEI comprende los elementos siguiente⁴:

1. Descripción bibliográfica del documento codificado, muy similar a la descripción bibliográfica normalizada para bibliotecas o sea las normas ISBD, aunque la introducción de la información, para la cabecera TEI, no requiere de normas tan precisas y rigurosas, por lo que esta operación puede llevarse a efecto por personas menos expertas que los bibliotecarios profesionales..

2. Descripción de la codificación, o sea la forma como se ha codificado el documento, también la relación entre el documento electrónico y la fuente, y finalmente los niveles de codificación y otros aspectos.

3. Descripción del perfil o información no bibliográfica como lengua, fechas y otros.

4. Descripción de la revisión o modificaciones del documento electrónico.

Estos elementos de la cabecera TEI tienen la gran carencia de no establecer, de forma precisa, las informaciones mas relevantes, esto es los puntos de acceso.

3. Origen y proyección del Círculo de Dublin

El formato Dublin Core nace para tratar de proporcionar metadatos para los materiales accesibles en red. El primer intento de normalizar el procesamiento técnico de la información de Internet fue pronto reconocido, pues existía ya gran sensibilidad por las búsquedas y recuperaciones de los recursos de Internet. Se entendieron los metadatos como un recurso que podían facilitar un consenso en la edición de documentos en la red. El formato Dublin Core se ha gestado bajo el patrocinio de OCLC a partir de 1995, que ha pretendido avanzar en su nueva concepción de los catálogos a través de las páginas Web. El primer objetivo, en este sentido, que implantó OCLC para los catálogos fue la indización encadenada, donde de los números de clasificación de un Sistema de Clasificación tradicional, (CDU, DDC, LCC, y otros) se derivan encabezamientos de materia y viceversa. Este procedimiento en cadena fue ideado por Ranganathan y ha sido llevado a la práctica no solo por la Bibliografía Nacional Inglesa y otras muchas bibliografías nacionales, sino que OCLC lo ha incorporado a su catálogo hipertextual basándose en la Clasificación Decimal de Dewey electrónica que tiene asociada la lista de encabezamientos de materias de la Biblioteca del Congreso de Washington.

OCLC ha pretendido avanzar en los catálogos hipertextuales y ha patrocinado el formato Círculo de Dublin. Este tiene su origen en un núcleo o círculo intelectual de Dublin (primer encuentro de 1995), en el estado Ohio en EEUU, sede de OCLC de ahí toma su nombre, pues basa su denominación en el Dublin de Joyce, o en el París de Benjamín, o en la Praga de Kafka, o sea Dublin como lugar de encuentro y foco intelectual. Como lo fuera al inicio del siglo XX el Círculo de Viena en el ámbito de la metodología científica. El Círculo de Dublin pronto va a considerar los documentos como objetos e integra todos los posibles tipos de materiales (tercer encuentro de 1996) y su tratamiento técnico, ya sean bases de datos, imágenes digitales, bancos de imágenes, textos electrónicos, videos y películas en formato digital y no digital, objetos multimedia, grabaciones sonoras en formato digital y no digital. De esta forma se pretenden integrar los distintos catálogos en un dominio de acceso cruzado y múltiple a través de los enlaces hipertextuales.

El formato Dublin Core consta de 15 elementos que son empleados como descriptores para, los documentos-como objetos. Además se ha incluido otras aportaciones como la estructura Warwick (segundo encuentro de 1996) que es un modelo sintáctico para transportar paquetes de metadatos basados en sets de descriptores como en el Dublin Core. Y los calificadores de "tipo" y "esquema" (cuarto encuentro de 1997) dan a los usuarios del Dublin Core cierta habilidad de añadir calificativos a sus elementos. Se ha pretendido, también, la creación de servicios de registro para esquemas de metadatos, con el objeto de llegar a una mayor normalización o a un cuerpo de normas internacionales (quinto encuentro de 1997).

La historia del Círculo de Dublin es corta pero intensa, en poco período de tiempo ha tenido jugar varias reuniones bajo el patrocinio de OCLC y ha conseguido implantar y extender el formato Círculo de Dublin.

1. El grupo de trabajo del Círculo de Dublin tuvo su primer seminario o encuentro en Dublin, Ohio, en marzo de 1995, donde se trató de establecer recursos en red para facilitar la búsqueda y recuperación de la información, creando el marco para la descripción de las fuentes electrónicas. Los 52 participantes seleccionados de esta primera reunión fueron muy heterogéneos había fundamentalmente bibliotecarios investigadores de Biblioteconomía y Documentación y otros, y supuso, que tuvieran intereses y puntos de vista diferentes, pero todos ellos convergieron en la necesidad de crear unos elementos en metadatos estándares. A partir de aquí nació el Dublin Core o Círculo de Dublin que consta de una cabecera de datos para describir los recursos de Internet. Este fue el logro más eficaz pues elaboraron una pequeña página de instrucciones destinada a los autores y editores de información en la red para que pudieran introducir su información sin excesivos conocimientos de Documentación. Se creó una lista de elementos donde se definía el recurso.

Estos elementos iniciales de los documentos en red que serán aplicables a cualquier documento y será una norma a seguir y va a denominarse Dublin Core. Se trata de un núcleo de elementos básicos no un número mínimo de elementos. Se estableció la denominación de Círculo de Dublin aplicado para la descripción de este núcleo de elementos descriptivos. En principio, se establecieron los 13 elementos siguientes: Título, Autor, Materia, Editor, Otros colaboradores, Fecha de publicación, Forma, Formato, Identificador o número de identificación, Relación, Fuente de la que deriva el recurso, Lengua, Localización temporal, espacial y características.

Estos elementos Dublin Core deben comprender la información siguiente:

- Título: Nombre o título asignado al documentos o a la fuente
- Autor: Creador intelectual del contenido del documento o de la fuente
- Materia: palabras claves o frases que resumen el contenido de un documento o de la fuente, el empleo de vocabularios controlados y esquemas de clasificación formales es aconsejable.
- Resumen: Se trata de un resumen o descripción textual del documento en caso de documentos-como objetos y se hará descripción del contenido para fuentes visuales.
- Editor: Editor o entidad responsable de la publicación del documento puede ser editor comercial o sin fines de lucro.
- Otros autores: Otros colaboradores u otras menciones de responsabilidad, que aportan contenido intelectual al documento.
- Fecha de publicación: Fecha de creación de la fuente, este dato debe hacerse según la norma ISO 8601, según la cual la estructura para consigna la fecha es AAAA-MM-DD, o sea 14 de abril de 1937 será 1947-04-14.
- Forma: La forma puede ser referente a la forma material como reportaje tenido, diccionario, también a la forma literaria como ensayo, poema, novela o a la forma en paganas Web como home página o primera página.
- Formato: Sirve para identificar el programa informático y el equipo material.
- Identificador: Número único, individualizado e internacional que identifica la fuente puede ser una URL, URN o bien un ISBN.
- Fuente actual: Versiones, adaptaciones, revisiones de documentos originales, indicaremos solo el documento que estamos procesando solo si fuera necesario y cambiara totalmente la forma respecto del documento original lo consignaremos, en este caso podría hacerse mención de los metadatos del documento original, ya sea el autor, el numero internacional, fecha u otros datos del documento original.
- Lengua: Lengua en la que esta el contenido intelectual de la obra, Ejemplos: en, de, es,fi, fr, ja, th, zh.

.Relación originaria: Ediciones, Traducciones, Vaciado y Traslación:

.Sí se reseñan distintas ediciones, 1ª ed, o 2ª ed, se consignará Es Edición De.

.Si se trata de una traducción se consignará Es Traducción De.

.Si se trata de un procesamiento analítico o Vaciado de los capítulos o las imágenes del libro o documento se indicara Es Parte de.

•Si se trata de una traslaciones de un documento en texto a imagen se consignará Es Traslación De

- Tipo de material y sus fechas: Localización temporal y espacial del contenido intelectual de la fuente, tipos de materiales especiales y las duraciones temporales)

- Derechos: Copyright u otros.

Estos elementos del formato Dublin Core pueden clasificarse según tres aspectos⁶:

1. Contenido:

Título, Tema, Resumen, Fuente actual, Lengua, Relación originaria y Localización especial y temporal.

2. Propiedad intelectual:

Autor, Editor, Otros colaboradores y Derechos.

3. Instancia:

Fecha de edición, Forma, Formato, Número internacional.

El uso del empleo del formato Dublin Core supone un formato de intercambio para metadatos descriptivos, contiene su propia descripción de sus datos en red y supone una interoperabilidad semántica entre dominios cruzados. Además su empleo no requiere especialistas. Las aplicaciones futuras llevaran a diferentes tipos de metadatos.

2. El segundo encuentro tuvo lugar en la Universidad de Warwick, Inglaterra, en abril de 1996, con 50 participantes procedentes de 11 países, donde se abordó el problema de la implementación que requería una sintaxis formal y una estrategia organizativa. Fue organizado por la Oficina de los trabajos en Red de Biblioteconomía y Documentación UKOLN y la oficina de información de OCLC se estableció un lenguaje standard con una sintaxis en hipertexto, o sea, se trató de establecer criterios comunes para el HTML (Hypertext Markup Language). El objetivo prioritario era la aceptación de unas reglas para las fuentes de información en red. Se estableció una sintaxis concreta tipo Definición de Documento Tipo en GGML. La sintaxis del Círculo de Dublin estableció como lenguaje de enlaces generalizados y normalizados el SGML (Standar Generalized language), esto es posible en lenguaje HTML (Hypertext Markúp language), o lenguaje de enlace's en hipertexto. Esta sintaxis fue ideada para las etiquetas en htm, para los Metadatos de los documentos Web. Se estableció, también, una arquitectura de contenido que incorporaba diferentes paquetes de metadatos y que podían mantenerse de forma independiente, esto se plasmó en el Warwick Framework. Esta estructura significó una forma de complementar los modelos de metadatos y posibilitar el intercambio de paquetes de metadatos, además supone una forma que permite satisfacer las necesidades de distintas comunidades con diferentes niveles de complejidad. La Estructura Warwick presenta el nivel mas alto, pero puede que esto plantee la necesidad de un registro de tipos de metadatos como el registro de tipos de Internet media o IANA también conocido como MIME tipos. La estructura Warwick aparece como una integración efectiva de metadatos en la infraestructura de información global. La realización de una arquitectura tal requiere de un gran esfuerzo por parte de muchos frente a la diversidad. El grupo de trabajo Warwick representa la necesidad y demanda de una mejor fuente de descripción en Internet.

En este segundo encuentro, se estableció una guía para los autores y editores que elaboran la descripción de los recursos y también para quienes administran estos recursos, en definitiva, se elaborará una guía para un nivel mas sencillo y otra para un nivel mas complejo. Ee van a elaborar paquetes de metadatos que van a posibilitar que estos se incluyan en el propio documento descrito o bien de forma separada con una referencia URL De esta forma el registro Dublin Core puede ser un paquete, y el registro MARC otro, igualmente otros paquetes pueden incluir otro tipo de información como metadatos de campo, datos administrativos, derechos y otros. Así mismo, se podría establecer un conjunto de metadatos múltiple que sería un registro tipo y que crearía nuevos tipos de metadatos con el fin de que la sintaxis fuera bien transferida entre estos grupos de metadatos.

3. El tercer encuentro del Círculo de Dublin se reunió de nuevo en Dublin, Ohio, en septiembre de 1996, con 70 participantes, al que se adhieren numerosas bibliotecas nacionales como la de Australia o la Biblioteca del Congreso de Washington, donde se van a incluir en el proyecto no solo bases de datos, sino también documentos o imágenes digitales. En los dos encuentros anteriores se trabajó, fundamentalmente, con documentos textuales, por contra, en este tercer encuentro se va a trabajar con documentos visuales como las fotografías, diapositivas, ficheros de imágenes. Se estableció que los diferentes tipos de materiales especiales debieran ser objeto de un tratamiento técnico distinto. Y se trató de elaborar un modelo de descripción de recursos visuales en Internet. Además se estableció la conversión del formato MARC al formato Núcleo de Dublin con los 13 siguientes elementos: Título (campo 245 del formato MARC); Autor (100 del MARC); Tema (653 del MARC, aunque abarca tanto clasificación sistemática como alfabética con listas de encabezamientos de materias, tesauros y otros); Editor (antiguo campo 260 del formato MARC relativo a los datos de publicación); Otros autores (720 del MARC); Fecha de publicación (260 \$c del MARC); Formato (texto en HTML, ASCII u otros); Identificador (URL, URN, ISBN o bien ISSN); Fuente (versión en papel, electrónica u otras); Lengua (Normas ISO, Formato MARC, Z 39.53 u otros); Vaciado (Vaciado de los capítulos o las imágenes del libro o documento); Tipo de material y sus fechas (tipos de materiales especiales y las duraciones temporales); Hay tres últimos elementos que⁴ se añadirán en el encuentro de Cambera en 1997: Resumen (520 del MARC relativo al área de notas); Forma (tipo de página electrónica o forma del documento) y Derechos (Copyright).

Esto significa que a partir del tercer encuentro a pesar que el Dublin Core sea independiente de cualquier sintaxis o sea de cualquier modificación, sin embargo, si se ha posibilitado que sea convertible fácilmente al formato MARC, de esta forma la bibliotecas pueden emplear estos recursos para sus fines ya que incluyen en sus catálogos recursos de Internet.

4. El cuarto encuentro del Círculo de Dublin se celebró en Camberra, Australia, en marzo de 1997, donde se estableció el acceso temático, ya sea mediante clasificación sistemática o alfabética, mientras los buscadores en Internet tienen un acceso sólo por palabras clave. Aquí surgieron dos posturas, de una parte los minimalistas partidarios de la simplicidad en la creación de metadatos y por otra parte los estructuralistas que aseveran graves dificultades en el intercambio de información debido a la simplicidad de los metadatos. pues la flexibilidad de significados formales o de los elementos ya extendidos o cualificados puede llevar a que sean usados por diferentes comunidades de forma distinta. Aquí se añadieron tres elementos adicionales como los Resúmenes, Forma y derechos.

5. El quinto encuentro se celebró en Helsinki, Finlandia, en octubre de 1997 al que ya asiste representación de sesenta países y se trató de consensuar los metadatos en páginas Web.

4. Descripción electrónica de los elementos Dublin Core

*

La descripción de los recursos electrónicos ha ido tras la trayectoria de las normas de descripción y su aplicación automatizada. Ha habido anteriores intentos de descripción de los recursos en Internet con el objeto de mejorar su localización.

Destacan los identificadores uniformes de recursos o URI (Uniform Resource Identifiers), estos identificadores tratan de acceder a la información de internet a través de motores de búsqueda de directorios.

El identificador o localizador de recursos URL (Uniform Resource locator) ha posibilitado reseñar de forma normalizada las direcciones de Internet, este identificador no da más información que la mera localización del recurso, pero cuando este cambia de localización queda totalmente inservible. Sin embargo, cuando un recurso cambia de dirección puede encontrarse una solución ideada por los usuarios y desarrollada por OCLC el denominado localizador persistente de recursos PURL (Persistent Uniform Resource) Locator. Ambos localizadores el URN y el URL son útiles para mejorar las búsquedas y el acceso a los directorios y a los propios motores de búsqueda.

Otro identificador de Internet es el nombre del recurso URN (Uniform Resource Name) que distingue por su nombre propio todos los recursos de Internet, su funcionamiento es muy similar al ISBN de los libros.

La sintaxis o sistema de codificación de un documento es totalmente independiente de los elementos Dublin Core. O sea, un sistema de codificación cualquiera, o el SGML, u otros. Presenta puntos de discusión mayores que lo relativo a la semántica en la que parece haber un mayor consenso, ya que pueden mostrarse varias sintaxis y varias formas de estructurar los datos. Desde que las páginas Web son el mayor medio en Internet el desarrollo de metadatos en las páginas Web es la estrategia primaria. Y en un momento posterior, va a ser el desarrollo de metadatos en HTML. Por tanto, los desarrollos de metadatos se van a producir o expresar en GSML DTD, esto va a quedar expresado mediante el empleo de enlaces.

La sintaxis de HTML en Dublin Core se representa de la forma siguiente:

```
<META NAME="DC.Título"CONTENT="
<META NAME="DC.Autor"CONTENT="
<META NAME="DC.Materia"CONTENT="
<META NAME="DC.Resumen"CONTENT="
<META NAME="DC.Editor"CONTENT="
<META NAME="DC.Otros autores"CONTENT="
<META NAME="DC.Fecha"CONTENT="
<META NAME="DC.Forma"CONTENT="
<META NAME="DC.Formato"CONTENT="
<META NAME="DC.Identificador"CONTENT="
<META NAME="DC.Fuente"CONTENT="
<META NAME="DC.Lengua"CONTENT="
<META NAME="DC.Referenciao Vaciado"CONTENT="
<META NAME="DC.Materiales especiales y Fechas"CONTENT="
<META NAME="DC.Derechos"CONTENT="
```

Las diferentes comunidades de usuarios y las distintas áreas de aplicación requieren diferentes elementos y niveles de complejidad, por tanto serán necesarios modelos de metadatos complementarios que requieren una arquitectura que acomode la variedad de los modelos de metadatos. Esta arquitectura debe modular los distintos tipos de metadatos. Los problemas que se generan por estos motivos plantean la problemática de extensividad del formato Dublin Core, como elementos adicionales que son necesarios añadir, en este sentido nace la necesidad de enlazar la descripción Dublin Core con otros tipos de descripciones, esto se planteó en el segundo encuentro y tuvo su solución en el tercero. El consenso de estos trabajos debe converger en la necesidad de una arquitectura que podría acomodar la diversidad de modelos y niveles de descripción que caracteriza el mundo heterogéneo de las fuentes electrónicas.

Todos los campos son repetibles y pueden tener asociados cuantos enlaces se considera necesarios. Tomando en consideración esta capacidad para describir estas fuentes con mayor precisión han sido ideados los calificadores "esquemas" y "tipos" relativos a cada elemento. Los calificadores⁸ "esquema" y "tipo" identifican el sistema empleado en la descripción de cada elemento. En la descripción Dublin Core se emplean para expresar información pertinente relativa a cada elemento. Por ejemplo el elemento autor puede reseñarse cuantas veces sea necesario para especificar los distintos calificadores "tipos" que se expresan en este elemento como los datos relativos a los autores: nombre del autor, su correo electrónico, su dirección, su teléfono, su fax, su lugar de trabajo y otros muchos datos. En definitiva tenemos que los "Esquemas" hacen referencia a las normas generales y los "Tipos" a los datos o elementos de esas normas generales. Por lo que los "Esquemas" pueden expresarse en línea y los "tipos", generalmente, no pueden expresarse así. En definitiva, los elementos Dublin Core son opcionales y repetibles, el esquema, además permite emplear calificadores opcionales para cada elemento que posibilitan indicar la normativa empleada en caso de haber usado normas de descripción bibliográfica usuales. Las posibles ambigüedades que se pueden presentar con los elementos Dublin Core se pueden sortear mediante el HTML enlaces para cada elemento del Dublin Core haciendo uso de una referencia realizada mediante un enlace y para el empleo de la estructura de cada enlace esta prevista una definición de la sintaxis en línea.

La sintaxis⁹ de los elementos Dublin Core se expresa de la forma siguiente:

Dublin CORE TÍTULO

Dublin CORE CALIFICADORES:

Esquema: ISBD; AACR2, TEI

Tipo: Título, Subtítulo, Título paralelo y otros

```
<META NAME="DC.Title"CONTENT="(Scheme=ISBD) La ilustre fregona>
<META NAME="DC.Title"CONTENT="(Scheme=AACR2)
<META NAME="DC.Title"CONTENT="(Scheme=TEI)
<META NAME="DC.Title"CONTENT="(Type=Subtítulo)= Matemáticas: 2ºcurso>
<META NAME="DC.Title"CONTENT="(Scheme=Título paralelo)
```

DC AUTOR

DC CALIFICADORES:

Esquema: ISBD; AACR2, TEI, IBERMARC, Autoridades Biblioteca Nacional

Tipo: Autor, Autor colectivo, Autor corporativo, Dirección, Teléfono, fax y otros.

```
<META NAME="DC.Autor"CONTENT="(scheme=ISBD) UNAMUNO, Miguel de>
<META NAME="DC.Autor"CONTENT="(scheme=AACR2)
<META NAME="DC.Autor"CONTENT="(scheme=IBERMARC) 100 I $a Tomas de Aquino, $c santo.
<META NAME="DC.Autor"CONTENT="(scheme=TEI)
<META NAME="DC.Autor"CONTENT="(Type=name) Pedro Fernández
<META NAME="DC.Autor"CONTENT="(Type=email) P.Fernández@casa2.es
<META NAME="DC.Autor"CONTENT="(Type=postal) C/ Bravo Murillo 34, 3
<META NAME="DC.Autor"CONTENT="(Type=phone) 91 624 93 28
<META NAME="DC.Autor"CONTENT="(Type=fax) 91 624 93 25
<META NAME="DC.Autor"CONTENT="(Type=affiliation)Universidad CarlosIII
<META NAME="DC.Autor"CONTENT="(Type=homepage) http://www.uc3m.es
```

DC MATERIA

DC CALIFICADORES:

Esquema: CDU, LCSH, CDD21, BBK, CC y otros

<META NAME="DC.Subject"CONTENT="(Scheme=LCSH)

<META NAME="DC.Subject"CONTENT="(Scheme=Encabezamientos de materias)

Bibliotecas-Adquisición

Biblioteconomía-Diccionarios

<META NAME="DC.Subject"CONTENT="(Scheme=CDU) 946.0"1936/1939">

<META NAME="DC.Subject"CONTENT="(Scheme=CDD)

<META NAME="DC.Subject"CONTENT="(Scheme=CDD21)

<META NAME="DC.Subject"CONTENT="(Scheme=BBK)

<META NAME="DC.Subject"CONTENT="(Scheme=CC)

DC RESÚMEN

DC CALIFICADORES:

Esquema: URL y otros

<META NAME="DC.Resumen "CONTENT=" Se trata de un resumen textual del contenido de la fuente y tratado el documento como un objeto"

<META NAME="DC.Resumen "CONTENT="(Scheme=URL)=

http://purl.org/metadata/dublin_core_elements#description

DC EDITOR

DC CALIFICADORES:

Esquema: ISBD; AACR2, TEI, IBERMARC

Tipo: Nombre de editor comercial, nombre del editor no comercial o entidad, dirección, teléfono, correo electrónico y otros.

<META NAME="DC.Publisher "CONTENT="(Scheme=ISBD)=

<META NAME="DC.Publisher "CONTENT="(Scheme=AACR2)

<META NAME="DC.Publisher "CONTENT="(Scheme=TEI)

<META NAME="DC.Publisher "CONTENT="(Type=Editor comercial)=Anaya>

<METANAME=" DC.Publisher "CONTENT="(Type=Entidad)=Universidad CarlosIII>

<META NAME="DC.Publisher "CONTENT="(Type=Fax)

DC OTROS AUTORES

DC CALIFICADORES:

Esquema: ISBD; AACR2, TEI, IBERMARC

Tipo: Nombre personal (apellidos, nombre), Autor corporativo, Dirección, Teléfono, Correo electrónico y otros.

<META NAME="DC.Otros autores"CONTENT="(Scheme=ISBD)=LEON, Felipe>

<META NAME="DC.Otros autores "CONTENT="(Scheme=AACR2)

<META NAME="DC.Otros autores "CONTENT="(Type=Fax)

DC FECHA

DC CALIFICADORES:

Esquema: ISO 8601 (AAAA-MM-DD), otros

Tipo:

<META NAME="DC.Fecha "CONTENT="(Scheme=ISO 8601) 1998-12-01>

<META NAME="DC.Fecha "CONTENT="(Scheme=ISBD)1994>

<META NAME="DC.Fecha "CONTENT="(Scheme=AACR2)

DC FORMA

DC CALIFICADORES:

Esquema: AACR2, ISBD (tipo de material y al área de descripción física)

Tipo:

<META NAME="DC.Forma "CONTENT="(Scheme=ISBD)=Archivo de ordenador>

<META NAME="DC.Forma "CONTENT="(Scheme=ISBD)=1 v. :il :29 cm

<META NAME="DC.Forma "CONTENT="(Scheme=AACR2)=

DC FORMATO

DC CALIFICADORES:

Esquema: IMT(Internet Media Tipos), como html, sgml, word perfect, y otros

Tipo:

<META NAME="DC.Formato "CONTENT="(Scheme=IMT) =text/html

<META NAME="DC.Formato "CONTENT="(Scheme=IMT) =Word perfect 6.1

<META NAME="DC.Formato "CONTENT="(Scheme=IMT) =Windows 3.1

DC IDENTIFICADOR

DC CALIFICADORES:

Esquema: URL,URN, ISBN, ISSN, FPL, SICI¹⁰

Tipo:

<META NAME="DC.Identificador "CONTENT="(Scheme=URL)=http://www.Uc3m.Es

<META NAME="DC.Identificador "CONTENT="(Scheme=ISBN) =0-8230-3255-8

<META NAME="DC.Identificadir "CONTENT="(Scheme=ISSN) =0943-7444

DC FUENTE

DC CALIFICADORES:

Esquema: Texto libre, URL,URN, ISBN, ISSN

Tipo:

<META NAME="DC.Fuente "CONTENT="(Scheme=URL)=http://www.Uc3m.Es

<META NAME="DC.Fuente "CONTENT="(Scheme=ISBN)= 0-9230-4355-8

<META NAME="DC.Fuente "CONTENT="(Scheme=ISSN)=0935-6643

DC LENGUA

DC CALIFICADORES:

Esquema: ISO.639-1¹¹ (en. fr, ger, sv), Z39.53

Tipo:

<META NAME="DC.Lengua "CONTENT="(Scheme=ISO 639-1)=en

<META NAME="DC.lengua "CONTENT="(Scheme=ISO 639-1)=sv

<META NAME="DC.lengua "CONTENT="(Scheme=ISO 639-1)=fr

DC REFERENCIA, VACIADO O CONTENIDO

DC CALIFICADORES:

Esquema: URL, URN, ISBN

Tipo: DC

<META NAME="DC.Relation "CONTENT="(Scheme=URL)=

<META NAME="DC.Relation "CONTENT="(Scheme=URN)=

<META NAME="DC.Relation "CONTENT="(Scheme=ISBN)=

DC MATERIALES ESPECIALES Y FECHAS

DC CALIFICADORES:

Esquema: A determinar por el grupo de trabajo coverage

Tipo: Localización espacial y temporal

```
<META NAME="DC.Coverage.Lugar" CONTENT="España">
<META NAME="DC.Coverage.Tiempo" CONTENT="Dinastía Ming">
<META NAME="DC.Coverage" CONTENT="(Scheme=IMT)=Windows 3.1
```

DC DERECHOS

DC CALIFICADORES:

Esquema: URL, URN

Tipo:

```
<META NAME="DC.Rights" CONTENT="Dominio público">
<METANAME="DC.Rights" CONTENT="(Scheme=URL)=http://www.someplace.se/copyrights.htm
```

Puede ocurrir que aparezcan los dos calificadores Esquema y Tipo y se presentaran de la forma siguiente:

```
<META NAME="DC.Autor" CONTENT="(TYPE=name)Pedro Fernández">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#author">
```

```
<META NAME="DC.Autor" CONTENT="(TYPE=email)P.Fernández@casa2.es">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#author">
```

```
<META NAME="DC.Autor" CONTENT="(TYPE=postal)C/ Bravo Murillo 34, 3">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#author">
```

```
<META NAME="DC.Autor" CONTENT="(TYPE=phone)91 624 93 28">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#author">
```

```
<META NAME="DC.Autor" CONTENT="(TYPE=fax)91 624 93 25">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#author">
```

```
<META NAME="DC.Autor" CONTENT="(TYPE=affiliation)Universidad CarlosIII">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#author">
```

```
<META NAME="DC.Autor" CONTENT="(TYPE=homepage)http://www.uc3m.es">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#author">
```

Cuando todos los elementos del Dublin Core no permiten una descripción total de la fuente o del documento el formato puede extenderse mediante un "paquete"

```
<META NAME="package" CONTENT="(TYPE=begin) Dublin Core es un...>
<META NAME="package" CONTENT="(TYPE=spatial) Los descriptores....>
<META NAME="package" CONTENT="(TYPE=end) la posibilidad de crear ...>
```

Si se trata de un esquema se procederá de forma similar

```
<LINK REL=SCHEMA.ad HREF="http://www.ncl.ac.uk/~napm1/ads/ahds_descriptor-elements#precision">
```

5. Proyectos que avalan el consenso del Círculo de Dublin

Son muchos los organismos e instituciones que tratan de definir y delimitar los formatos adecuados para poder insertar o publicar un documento en Internet, estos formatos se han de emplear en el momento en que se introduce la información para su localización posterior. Así organismos bibliotecarios nacionales, inter-

nacionales y centros de normalización tratan de definir estos formatos que van a reportar una gran acceso a la información bibliotecaria, bibliográficas y otras.

Proyecto de metadatos nórdico 'o NORDINFO, nace de la necesidad de intercambio bibliotecario y bibliográfico, el plan preliminar ha sido desarrollado por Juha Hakala de la biblioteca universitaria de Helsinki, este plan quedó avalado en el tercer encuentro del Círculo de Dublin aunque su gestación aconteció en el segundo encuentro.

Proyecto interoperativo de fuentes de nombres uniformes o proyecto TURNIP se inició en el centro tecnológico de sistemas distribuidos (DSTC) de Australia y ha producido una URN que empleó el formato de Dublin Core.

OCLC ha iniciado diversos proyectos para el desarrollo de metadatos en formato Círculo de Dublin¹².

También se emplea como acceso a la colección de mapas de OCLC en formato MARC y el proyecto Alejandría de Bases de datos de la Universidad de California en Santa Bárbara

El proyecto Espectro DE OCLC explora el empleo de interfaces basándose en una descripción electrónica.

El proyecto Escorpión de OCLC investiga la asignación automática de encabezamientos basada en el sistema de Clasificación Decimal Dewey y trata de usar, describir y generar automáticamente metadatos de clasificación basados en los elementos del Dublin Core. La base del proyecto Escorpión es que la Clasificación Decimal de Dewey puede aplicarse para asignar materias alfabéticas automáticamente. Se trata de construir un instrumento de reconocimiento automático de materia basado en las tablas de la Clasificación Decimal de Dewey. Esta clasificación parte de que es empleada en gran número de bibliotecas¹³ como en 135 países, está traducida en 30 lenguas y en EEUU se emplea en el 95% de las bibliotecas públicas y escolares además de un elevado número de bibliotecas universitarias y especializadas.

La Clasificación de Dewey se basa en un esquema jerárquico en el que los dígitos representan subdivisiones de las ciencias o disciplinas. El mantenimiento de la Dewey Electrónica se plasma a través de OCLC en ESS (Editorial Support System). El ESS se emplea para la conversión automática de bases de datos que están clasificadas y se les asigna encabezamientos de materia de forma automática. En los esquemas de la Clasificación de Dewey hay aproximadamente unos 30.000 términos alfabéticos, algunos de ellos extremadamente ambiguos, para poder ejercitar el asignar encabezamientos de materia de forma automática será necesario eliminar la ambigüedad de los términos alfabéticos. Escorpión es una herramienta que aplica la CDD a los documentos y los transforma en clasificación para los usuarios¹⁴. El añadir a este proyecto los encabezamientos de materias de la Biblioteca Del Congreso supone que estamos introduciendo otra jerarquía o bien otro vocabulario controlado organizado con otra jerarquía. Uno de los problemas con los que ese encuentra este proyecto radica en la estructura de enlaces que tiene hoy la CDD21.

El proyecto Kilroy trata de encontrar una herramienta útil para otros proyectos de organización del conocimiento.

El proyecto Documento Nacional y Servicio de Información el NDIS¹⁵ lo desarrolla la biblioteca Nacional de Australia y la Biblioteca Nacional de Nueva Zelanda, y ofrecen un servicio de búsqueda en las bases de datos australianas. Este proyecto trataba de integrar un servicio de información cruzada en cualquiera de los formatos existentes, este proyecto integra la aplicación de los elementos del Dublin Core, tanto para información bibliográfica como para información no bibliográfica, esto supone un almacenamiento, acceso y búsquedas en bases de datos cruzadas.

Otro proyecto ha sido el que ha pretendido la conexión entre el formato Dublin Core y el formato MARC que ha puesto de manifiesto a través de los trabajos elaborados en la Biblioteca del congreso de Washington que no ha sido fácil esta interconexión de forma total¹⁶.

El proyecto de la Biblioteca Digital de Alejandría (ADL) consistió en la traslación de su información en el formato Dublin Core.

El proyecto WordSmith a sido impulsado por el científico Jean Godby quien ha tratado de aplicar un set de herramientas para procesar el lenguaje natural a las colecciones de documentos Web, donde trata de encontrar un fórmula para indexación de documentos usando frases seleccionadas por el análisis estadísticos de los datos.

El Proyecto Organización del Conocimiento ha sido planteado por Diane Vizine-Goetz, trata de trabajar en la aplicación de los Sistemas de Clasificación para organizar el conocimiento y como fuente para descubrir el ámbito de Internet. O sea, la base de este proyecto es la clasificación Decimal de Dewey y la posibi-

lidad de aplicarla a la ordenación de libros y al mundo electrónico de Internet. Este grupo de trabajo trata de encontrar e implantar una vía automatizada de la estructura jerárquica de la Clasificación de Dewey, o de otras fuentes como la Lista de encabezamientos de la Biblioteca del Congreso (LCSH).

6. Aplicación práctica del formato Dublin Core

La aplicación práctica del formato Dublin Core se ha proyectado en varias grandes bibliotecas como la biblioteca del Congreso de Washigton, la Biblioteca Nacional de Australia y la Biblioteca Nacional de Nueva Zelanda, que han tratado de integrar los elementos del Dublin Core en sus catálogos, aunque, por el momento, el proyecto no se ha visto consumado. Pues, para que conversión tenga fiabilidad y mejore las capacidades de los catálogos y de los motores de búsqueda será fundamental que se articulen normas internacionales que reculen estos procedimientos, pues si los elementos no quedan totalmente controlados nunca estarán completos como los elementos que se incluyen en el formato MARC.

Otra aplicación práctica de interés es la realizada por la unión de catálogos en línea de OCLC o el denominado World Cat, que emplea como interface de acceso a los elementos establecidos del Dublin Core.

La descripción formal para metadatos Dublin Core con calificadores en RDF (Resource Description Framework) y la forma de extenderlos ha sido una de los puntos mas importantes de trabajo¹¹. El trabajo de DRF se realiza bajo los auspicios de World Wide Web Consortium W3C, la organización que trata la estandarización en la RED, hay un grupo de trabajo que esta trabajando en la representación del Dublin Core en RDF¹⁸. Esta estructura de metadatos pretende efectuar una interoperabilidad de intercambio entre máquinas la estructura DRF puede emplearse para varias áreas incluida la descripción de fuentes

Una implantación importante del Dublin Core radica en el generador automático de metadatos Dublin core, o sea, se plantea como modelo el DC-dot o generador Dublin Core, o sea creador automático Dublin Core del UKOLN19 desarrollado por British Library Research and Innovation Centre con colaboración de la Universidad de Baht. Ya existen otros varios creadores o generadores de metadatos.

7. Futuro del formato Dublin Core

La relevancia del formato Dublin Core radica en tres aspectos a considerar: El primero son los elementos Dublin Core, el segundo la sintaxis desarrollada en HTML y finalmente la gestión como página Web o lugar determinado en Internet. La gestión de los posibles, usos del formato Dublin Core puede centrarse en tres aspectos: como descripción compartida, como instrumento de autoridades, y como instrumento de gestión de lugares en la Red, ya que existen multitud de problemas en la edición de páginas web pues las hay estáticas, dinámicas y externas, estas dos ultimas son objeto de muchas modificaciones y en muchos casos se han editado mediante metadatos.

En definitiva, el futuro mas inmediato para el inicio del próximo milenio apunta a la creación, acceso y métodos de organizar los datos de distintos materiales de las bibliotecas clasificados en línea. Las postrimeras del siglo XIX supusieron un gran cambio de paradigma, en lo que a las clasificación documental se refiere. De igual forma, en los albores del siglo XXI, la organización de la información presenta una novedosa arquitectura hipertextual y no lineal que va a comportar un nuevo cambio de paradigma para el próximo milenio en la organización del conocimiento.

Las reuniones del Círculo de Dublin han sido de gran provecho pues son numerosas los organismos que han adoptado la forma del Dublin Core para el diseño de sus páginas en Internet, para organizar las propias páginas y para facilitar la tarea a los buscadores.

Además, muchas páginas tienen enlaces con materias (palabras claves) y resúmenes como metaelementos pues son empleados por grandes motores de búsqueda como Alta vista, aunque todavía ninguno de los grandes motores de búsqueda realizan sus búsquedas mediante meta-preguntas. Pues todavía no existe ningún índice de metadatos de especial significancia. Sin embargo, las meta-preguntas Dublin Core Autor y Dublin Core Materia comprenden la misma información que otros documentos electrónicos procesados mediante otros sistemas. En el diseño de páginas Web se hace uso de HTML y de la metaidentificación o

metaseñalización, o sea, las palabras claves funcionan como metaseñalización pues sirven para indizar una página Web. Y se ha pretendido elaborar un índice de metadatos pan-temáticos que recogen la metainformación reseñada en tema. Incluso hay propuestas para crear un registro de metadatos, pues para describir las fuentes de Internet con metadatos y para interpretar los metadatos es necesario conocer la forma electrónica de la fuente. Se trata de conocer la semántica del esquema u ontología. Los registros de metadatos contienen información sobre la semántica, estructura y sintaxis, de los elementos metadatos²⁰. Se pretende conseguir el gran índice de metadatos que sea accesible por materias a través de los grandes motores de búsqueda.

Además, uno de los aspectos más importantes de los metadatos Dublin Core es que el elemento materia²¹ no comprende únicamente palabras clave, sino que incluye la descripción del contenido completo del documento o de la fuente, o sea comprende las clases de un sistema de clasificación, descriptores tomados de un Tesauro, encabezamientos de materia y otros.

Los textos electrónicos procesados mediante metadatos Dublin Core podrían funcionar, al igual que la cabecera TEI, como funcionan los CIP en los libros, pues son los propios libros los que incluyen los datos que identifican y describen al libro. Los CIP son los centros catalográficos nacionales correspondientes y encargados de asignar estos datos. Sin embargo tanto en los sistemas de metadatos Dublin Core como en la cabecera TEI, ambos están ideados para que sea el propio autor o editor quien elabore la cabecera. De esta forma en varios de los elementos Dublin Core pueden aplicarse las normas catalográficas y de clasificación, indización y todo el proceso técnico usual en las bibliotecas, aunque parece difícil que la cabecera Dublin Core, así elaborada, incluya los datos tan elaborados y normalizados.

Se puede decir que el formato del Dublin Core es un formato genuinamente anglosajón pues son EEUU, Canadá, Australia e Inglaterra los países que han promocionado y posibilitado el desarrollo e implantación del formato. La exportación del mismo a todos los países de su área de influencia es rápida y total, pues la pretendida globalización actual abarca a los ámbitos político, económico, y financiero, y como infraestructura e instrumento de estos, tendremos, por tanto, la globalización de la organización y recuperación de la información.

Bibliografía

- MILLER, Paul. Metadata for the masses. URL: <http://www.ariadne.ac.uk/issue5/metadata-masses/>
- MILLER, Paul; GREENSTEIN, Daniel [Ed]. A practical implementation of the Dublin Core: Discovering Online Resources Across the Humanities. URL: <http://ahds.ac.uk/public/metadata/discovery>
- Generación automática de HTML y metadatos: <http://www.ukoln.ac.uk/metadata/dcdot>
- Sistema de metadatos Dublin Core. URL: <http://orc.rsch.oclc.org:9016/dems/index.html>
- POWEL, A. Dublin Core management URL: <http://www.ariadne.ac.uk/issueO/dublin/>
- SAN SEGUNDO MANUEL, Rosa. Sistemas de Organización del conocimiento. Madrid :BOE, Universidad Carlos III, 1996
- THOMSON. R.; SHAFER, K; VIZINE-GOETZ, D. Evaluating Dewey Concepts as a Knowledge Base for Automatic Subject Assignment. URL: http://ors.rsch.oclc.org:6109/eval_dc.html
- WEIBEL, S.; KUNZE, L; LAGOZE, C. Dublin Core Metadata for Simple Resource Discovery. <ftp://ftp.ietf.org/internet-drafts-kunze-dc-02.txt>
- WEIBEL, S.; GLDBY, J.; MILLER, E. OCLC/NCSA Metadata Workshop Report. URL: http://www.oclc.org:5047/oclc/rese...s/metadata/dublin_core_report.html
- XU, Amanda. Metadata Conversion and the Library OPAC. URL: <http://web.mit.edu/waynej/vvwww/xu.htm>