

MODELOS DE INTEROPERABILIDAD EN BIBLIOTECAS DIGITALES Y REPOSITORIOS DOCUMENTALES: CASO BIBLIOTECA DIGITAL COLOMBIANA

Ponente: Laureano Felipe Gómez Dueñas
Idioma: Español
Recursos Técnicos: Presentación en PowerPoint - Video Beam, Internet.
Datos del Autor: Laureano Felipe Gómez Dueñas
Docente- Investigador - Programa Sistemas de Inf. y Documentación - Universidad de La Salle
Magister en Sistemas de Información Digital, Universidad de Salamanca
Especialista en Redes de Información Documental, Pontificia Universidad Javeriana
Ingeniero de Sistemas, Universidad Nacional de Colombia
MSN/Messenger: laureanofg@hotmail.com
Correo Electrónico : felipe.gomez3@gmail.com
Google/Talk: felipe.gomez3@gmail.com
Skype: laureanofg

Resumen

A partir del inventario, análisis y evaluación de los diferentes Repositorios Documentales Digitales y las Bibliotecas Digitales existentes en Colombia, realizados al interior del proyecto de investigación denominado Biblioteca Digital Colombiana (BDCOL), auspiciado por Colciencias, el Ministerio de Educación de Colombia y trece Universidades participantes, Se realizó un documento que contiene un modelo de interoperabilidad entre sistemas de información documental, que incluye la descripción de los estándares y normas, en los que se basará BDCOL para realizar el intercambio y recolección de metadatos y objetos digitales de los diferentes Repositorios institucionales y Bibliotecas Digitales de Colombia, en especial aquellos cuyas instituciones responsables pertenecen a la Red Nacional de Alta Velocidad y Tecnologías Avanzadas (RENATA).

Introducción

Actualmente el término Interoperabilidad es comúnmente usado en la jerga de los profesionales de información para indicar que se están compartiendo recursos, productos y servicios de información especializada, presente en sus sistemas de información, con otras personas y sistemas de información ubicados en cualquier parte del planeta. De esta forma se está evidenciando la construcción de una gran red global de conocimiento académico y científico que permitirá acercar el conocimiento humano a todas las personas y reducir la brecha digital ocasionada por la dificultad de acceder oportunamente a información actualizada y de calidad.

Para especificar el entorno de acción de este trabajo, se ha hecho una aproximación basada en los repositorios de documentos digitales (REDIS), las bibliotecas digitales (BIDIS) y en general, los sistemas de información documental (SIDOCS) existentes en Colombia, los cuales fueron analizados respecto a su organización, colecciones ,servicios, productos, protocolos, etc., buscando establecer las mejores prácticas que permitirían integrar las diversas colecciones de documentos existentes, y a partir de esta generar un sistema de metabúsqueda que proporcione

servicios avanzados de información, y posibilite la visibilización, posicionamiento y uso global de la producción académica-científica Colombiana.

Utilizando una aproximación basada en el contexto de las bibliotecas digitales y los Sistemas de Información Documental en el contexto de BDCOL, se define la interoperabilidad como:

“La capacidad de un sistema de información para comunicarse y compartir datos, información, documentos y objetos digitales de forma efectiva (con una mínima o nula pérdida de su valor y funcionalidad), con uno o varios sistemas de información (siendo generalmente estos sistemas completamente heterogéneos, distribuidos y geográficamente distantes), mediante una interconexión libre, automática y transparente, sin dejar de utilizar en ningún momento la interfaz del sistema propio”¹.

Es importante considerar que aunque se puede lograr la interoperabilidad utilizando modelos propietarios creados por los desarrolladores de los propios sistemas de información, en el contexto en que se está enfocando ese trabajo; la interoperabilidad para que sea trabajada universal indistintamente por los sistemas de formación debe estar basada en estándares abiertos, de esta forma “una solución informática tendrá futuro porque se podrá comunicar con todo su entorno, o, por el contrario, será meramente pasajera y al final tan cara como cualquier moda”²

De esta manera se puede observar que la interoperabilidad se manifiesta por:

- La capacidad de los sistemas para trabajar entre sí en tiempo real o programado.
- La capacidad del Software para trabajar en diferentes sistemas.
- La capacidad de los datos para ser intercambiados entre diferentes sistemas (portabilidad)

En relación con la existencia de los distintos sistemas de información documental, repositorios y bibliotecas digitales que comprenden actualmente **BDCOL** (Aproximadamente 13 sistemas de información), se busca analizar los distintos niveles de interoperabilidad desde cuatro aspectos fundamentales, según el modelo propuesto por Krsulovic³:

- **Interoperabilidad Sintáctica:** En general se refiere a la capacidad de los sistemas de información para leer datos procedentes de otros similares y obtener una representación que pueda ser compatible. Esto se logra mediante la utilización de formatos/modelos estandarizados de codificación y estructuración de documentos y metadatos. Esto incluye el uso de lenguajes y metalenguajes estructurados (XML, ASN1), junto con modelos de metadatos estandarizados (Dublin Core, Marc21, Etd-ms, MODS, etc..) cuyos elementos representan una sintaxis común entre los diferentes sistemas.
- **Interoperabilidad Semántica:** Se puede entender como la capacidad de los sistemas de información (Bibliotecas Digitales y Repositorios Institucionales), para intercambiar información basándose en un **significado común** de los términos y expresiones contenidos en los metadatos y documentos, con el fin de asegurar la consistencia,

¹ Gómez Dueñas Laureano Felipe. Interoperabilidad en los Sistemas de Información Documental (SID): la información debe fluir. Revista Códice. Vol 3 No. 1. 2007.

² Consejo Superior de Informática de La Comisión Europea. Marco Europeo de Interoperabilidad Versión 1.0., <http://europa.eu.int/idabc/>

³ Krsulovic, E. Blog de la Web Semántica. <<http://www.dcc.uchile.cl/~ekrsulov/prj/ws-blog/>>

representación y recuperación de los contenidos. Esto involucra el uso de esquemas semánticos que incluyen vocabularios controlados (tesauros, listados de encabezamiento de materias, anillos de sinónimos, taxonomías, ontologías, etc.).

- **Interoperabilidad Estructural:** Corresponde con los modelos lógicos comunes y la capacidad de los sistemas de información para comunicarse e interactuar en ambientes heterogéneos (en grupos abiertos/cerrados, con control centralizado/distribuido). Esto incluye la definición y utilización de protocolos especializados como Z39.50, OAI-PMH, SRU, RSS, etc.
- **Interoperabilidad de Infraestructura:** utilización de un medio físico/lógico que sirva como un canal que permita realizar los procesos de intercambio de datos, información, documentos, metadatos y objetos digitales. independiente de la plataforma, soportados por protocolos de intercambio y acceso comunes a redes de datos tales como ISO-OSI y TCP/IP.

Es conveniente comprender que en ámbito de las distintas unidades de información se busca prioritariamente satisfacer las necesidades de información del usuario, permitiéndole buscar, recuperar y acceder efectiva y oportunamente a los distintos documentos relevantes que existen en determinada colección documental. Aquí subyacen tres ideas principales relacionadas con la interoperabilidad de los SIDOCS, REDIS y BIDIS:

- La importancia de poseer una colección grande y compleja, la cual cubra prácticamente cualquier necesidad de información y cuyos contenidos sean seleccionados y evaluados siguiendo criterios académicos y científicos, lo cual permitirá redundar en la calidad de los mismos y la pertinencia a la hora de su consulta y posterior uso.
- El enfoque y la importancia de la efectividad en el proceso de recuperación de información, el cual le permita al usuario acceder rápidamente a documentos relevantes de acuerdo con su necesidad de información.
- El uso de meta-esquemas de información estructurada (**Metadatos**) que permitan integrar distintas colecciones heterogéneas y distribuidas más fácilmente mediante el establecimiento de un marco común, de tal forma que permitan mejorar la recuperación de los documentos.

ESTÁNDARES DE METADATOS

Desde la historia remota, una función principal del profesional en información ha sido registrar y describir el conocimiento creado por el hombre, buscando la difusión y preservación de la memoria humana. Sin embargo en la actualidad debido a la existencia de grandes cantidades de recursos (documentos) digitales, esta función ha sido asociada completamente al concepto de los metadatos que le imprime unas cualidades especiales más allá del hecho de representar abstractamente un recurso. Una base para comenzar a definir los metadatos esta en sus raíces etimológicas las cuales indican:

| |
|---------------------------------|
| $\mu\epsilon\tau\alpha$ + datum |
|---------------------------------|

| | |
|-----------|--|
| Μετα | “junto a”, “después de”, “entre” o “con” |
| Datum | Dato |
| Metadatos | “Datos junto a después de entre con los datos” |

Sin embargo, esta aproximación es bastante pobre respecto a la verdadera potencia que actualmente ofrecen los metadatos, por ahora se puede agregar a esta breve introducción que los metadatos contextualizan y dan significado explícito suficiente para que un computador (aplicación) pueda gestionar datos e información de manera automática. El creador de la Web, Tim Berners Lee⁴ describe el concepto de "*metadata*", como "una máquina comprensible de información acerca de objetos Web", donde estos objetos Web, pueden ser cualquier tipo de documentos que se pueda visualizar en una navegador (Páginas Web, Audio, Video, Animaciones, Imágenes, etc..). También cabe añadir que uno de los propósitos básicos de los metadatos esta dado para facilitar y mejorar la recuperación de información para acceder directamente al documento de contenido completo. Según García Martínez, define los metadatos como: "datos asociados a documentos que ayudan a los usuarios potenciales a tener un conocimiento anticipado de la existencia característica de los mismos; todo ello teniendo en cuenta que soportan una gran variedad de operaciones y que un usuario puede ser una persona o un programa"⁵; Mientras que Eva Méndez los define como: "los metadatos son estructuras de información legibles por máquinas, cuya finalidad es hacer útiles a los datos, de distintas formas, según las necesidades concretas de cada servicio de información digital y según la aplicación que se les otorgue"⁶. Así mismo Rosa San Segundo Manuel, los define como: "Los metadatos o datos representacionales son definidos como el dato sobre los datos, 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"⁷.

Analizando el contexto de BDCOL, se entienden los metadatos como: ***“Información Estructurada ó Semi-estructurada que describe, explica, localiza, administra y amplía la información acerca de un recurso ó fuente de información, de forma que tenga algún sentido para las maquinas (Sistemas de información) y los humanos, y se pueda utilizar en los procesos de interoperabilidad; por concerniente, los metadatos son la suma total de lo que se puede decir acerca de cualquier objeto de información en cualquier nivel de agregación”.***

Así mismo se optó en BDCOL, por utilizar el estándar Dublin Core (ISO 15836:2003) Simple (DC simplificado) y con Calificadores (DC calificado), agregándole algunas modificaciones, como modelo base de interoperabilidad entre los distintos SIDOCS, REDIS y BIDIS.

⁴ Berners-Lee, Tim.; Connolly, D. Y Swick, R (1999) *Web Architecture: Describing and Exchanging Data*. <http://www.w3.org/199/04a/WebDat>.

⁵ García Martínez, Ana María. Definición y estilo de los objetos de información digitales y metadatos para la descripción. Boletín de la asociación Andaluza de bibliotecarios, N. 63, Junio 2001, p.23-47

⁶ Méndez Rodríguez, Eva M^a. La descripción de documentos electrónicos a través de metadatos: una visión para la Archivística desde la nueva e-Administración. Revista d'Arxius, 2003, pp. 47-82. < http://e-archivo.uc3m.es:8080/dspace/bitstream/10016/878/1/EMendez_Arxius.pdf>

⁷ San Segundo Manuel, Rosa. Organización del Conocimiento en Internet: Metadatos Bibliotecarios Dublin Core. VI Jornadas Españolas de Documentación, FESABID 98. <<http://www.uag.mx/eci/infosource/Articulos/Profinfo/organizacionconocimiento.pdf>>

| CONTENIDO | PROPIEDAD INTELECTUAL | INSTANCIACIÓN |
|---------------------------|---------------------------|----------------------------|
| Título (title) | Autor (creator) | Fecha (date) |
| Materia (subject) | Editor (publisher) | Tipo (type) |
| Descripción (description) | Colaborador (contributor) | Formato (format) |
| Fuente (source) | Derechos (rights) | Identificador (identifier) |
| Idioma (language) | | |
| Relación (relation) | | |
| Cobertura (coverage) | | |

Tabla 1. Elementos Dublin Core Simple (ISO 15836:2003)

| DCTERMS |
|--|
| La utilización del modelo Dublin Core Calificado (DCTERMS) lo conforman 55 Elementos http://dublincore.org/documents/dcmi-terms/ |
| abstract, accessRights, accrualMethod, accrualPeriodicity, accrualPolicy, alternative, audience, available, bibliographicCitation, conformsTo, contributor, coverage, created, creator, date, dateAccepted, dateCopyrighted, dateSubmitted, description, educationLevel, extent, format, hasFormat, hasPart, hasVersion, identifier, instructionalMethod, isFormatOf, isPartOf, isReferencedBy, isReplacedBy, isRequiredBy, issued, isVersionOf, language, license, mediator, medium, modified, provenance, publisher, references, relation, replaces, requires, rights, rightsHolder, source, spatial, subject, tableOfContents, temporal, title, type, valid |

Tabla 2. Elementos Dublin Core Calificado (DCTERMS)

Adicionalmente para las instituciones que utilicen el estándar **MARC21**, se recomienda utilizar la tabla de equivalencias Marc21/Dublin Core, que ha sido desarrollada por la Biblioteca del Congreso de los Estados Unidos de Norteamérica <http://www.loc.gov/marc/marc2dc.html>

LA INICIATIVA DE ARCHIVOS ABIERTOS Y EL PROTOCOLO OAI-PMH

La Iniciativa de Archivos Abiertos (OAI), busca proporcionar un sencillo modelos de interoperabilidad que incluye una arquitectura y especificaciones técnicas necesarias para que cualquier persona ó institución que haya creado contenidos académicos y científicos pueda hacerlo visible y accesible a través del uso de los distintos SIDOCS, REDIS y BIDIS

La iniciativa propone el uso del protocolo para la recolección de metadatos (OAI-PMH), como un modelo de interoperabilidad estructural que hace posible la comunicación entre distintos sistemas de información, que facilitan la visibilidad y accesibilidad de los contenidos. Es importante recalcar que, aunque el protocolo OAI-PMH permite básicamente el intercambio de metadatos, existen muchas iniciativas que permitirían a partir del uso de este protocolo transmitir también los textos completos de los documentos que se referencian (DIDL).

Reme Melero y José Barrueco señalan tres características fundamentales del protocolo⁸:

1. **Simplicidad:** Se concibió bajo la premisa de la sencillez. Conscientes de los problemas de implementación que habían tenido otras iniciativas anteriores como Z39.50 o Dients. Los creadores buscaron una fórmula simple que estuviera al alcance de cualquier potencial implementador.
2. **Normalización:** Basado en estándares ampliamente utilizados en Internet como son el protocolo http (HiperText Transport Protocol) para la transmisión de datos y órdenes, XML (Extended Markup Language) para la codificación de los datos y Dublin Core como elemento básico de metadatos.
3. **Recolección:** Frente a otros sistemas de agregación de contenidos como la búsqueda distribuida (Z39.50) o los sistemas de sindicación de contenidos vía RSS. OAI-PMH ha optado por la recolección de metadatos. En este caso, existe una entidad que pone a disposición de los interesados información bibliográfica sobre los documentos que almacena (Proveedor de Datos / PD) y los agregadores de contenidos que recogen periódica y sistemáticamente todos o parte de los metadatos expuestos para, localmente, implementar servicios de valor añadido (Proveedores de Servicio / PS).

La comunicación entre PD y PS se realiza utilizando el protocolo **http** (a través de los métodos GET y POST) para emitir preguntas y obtener repuestas a partir de una lista de argumentos con la forma de pares del tipo: "clave=valor". El PS puede pedir al PD que le envíe metadatos según determinados criterios como por ejemplo la fecha de creación de los registros. En respuesta, el PD envía un conjunto de registros codificados en formato XML.

Existen solamente seis peticiones (Verbos) que un PS puede realizar a un PD:

1. **Identify:** utilizado para recuperar información sobre el PD: nombre, versión del protocolo que utiliza, dirección del administrador, etc. **<http://pd.org/OAI-script?verb=Identify>** .
2. **ListMetadataFormats:** devuelve la lista de formatos bibliográficos que utiliza el servidor. **<http://PD.org/OAI-script?verb=ListMetadataFormats>**. Aunque el protocolo especifica que como mínimo los metadatos se expresen en formato Dublín Core no calificado especialmente diseñado por la iniciativa (oai_dc). No restringe el uso de formatos adicionales de metadatos (codificados en XML), que

⁸ Melero Melero, Remedios; Barrueco Cruz, José Manuel. Acceso Abierto y Repositorios de Documentos. Unidades de Autoformación SEDIC.
http://www.sedic.es/autoformacion/acceso_abierto/presentacion.htm>

permitan una descripción más rica de los documentos como por ejemplo MARCXML, LOM, ETD-MS, etc..

3. **ListIdentifiers:** recupera los encabezamientos de los registros en lugar de los registros completos. Tiene un argumento obligatorio (metadataprefix) para especificar el formato de metadatos en el que se quiere obtener la respuesta. Permite, además, argumentos como el rango de fechas entre los que queremos recuperar los datos. **http://PD.org/OAI-script?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2007-01-01&until=2007-08-01**
4. **ListRecords:** igual que el anterior pero recuperando los registros completos en n modelo de metadatos especificado por el verbo (metadataprefix). **http://PD.org/OAI-script?verb=ListRecords&metadataPrefix=oai_dc&from=2007-01-01&until=2007-08-01**
5. **GetRecord:** utilizado para recuperar un registro concreto. Necesita dos argumentos: identificador del registro solicitado (identifier) y especificación del formato de metadatos (metadataprefix) en que se desea obtener. **http://PD.org/OAI-script?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:PD:1/1479**
6. **ListSets:** recupera un conjunto de registros específicos, estos conjuntos son creados opcionalmente por el servidor para facilitar una recuperación selectiva de metadatos. Se trata de una clasificación de los contenidos según diferentes criterios como materias, lenguaje, tipología documental, etc. **<http://PD.org/OAI-script?verb=ListSets>**

DIRECTRICES DE INTEROPERABILIDAD EN BDCOL

Con el fin de buscar un modelo de integración general en el marco de BDCOL, se requieren establecer unas directrices que permitan normalizar las diferentes interpretaciones de las normas y estándares en las implementaciones de SIDOCS, REDIS y BIDIS que realiza cada institución perteneciente a BDCOL, estas directrices contienen características obligatorias y otras recomendadas que serán especificadas a lo largo del documento. Estas directrices permitirán entre otras:

- Construir servicios de calidad (por ejemplo, la búsqueda)
- Estandarizar la calidad de los metadatos de los repositorios locales
- Resolver la semántica y cuestiones de clasificación
- Asegurar la interoperabilidad (sintáctica, semántica, estructuras y de infraestructura)
- Mejorar el acceso al texto completo

Para una mejor comprensión de las directrices, estas fueron agrupadas según los distintos aspectos fundamentales de interoperabilidad mencionados anteriormente:

- Interoperabilidad Sintáctica
 - Esquemas de codificación de Caracteres (UTF-8)
 - Estándares de Metadatos
 - Identificación de documentos de acceso abierto (accessRights)
 - Formatos de Elementos por Defecto
 - Fecha
 - Identificador
 - Idioma
 - Coberturas Geográficas/Espaciales
- Interoperabilidad Semántica
 - Tipologías Documentales
 - Vocabularios Controlados
- Interoperabilidad Estructural
 - implementación del protocolo **OAI-PMH**
- Interoperabilidad de Infraestructura
 - Protocolo HTTP y Direcciones IP de Internet y RENATA
 - ID única y acceso directo a los Documentos Completos

Interoperabilidad Sintáctica

- **Esquemas de codificación de Caracteres:**

La codificación de caracteres es el método que permite codificar cualquier documento mediante la representación de un conjunto de símbolos de un alfabeto que pertenece a un sistema de representación, como un número o una secuencia de pulsos eléctricos en un sistema electrónico, aplicando normas o reglas de codificación.

En el caso de los metadatos y documentos que contienen texto existen múltiples esquemas de codificación y almacenamiento de los mismos en archivos que, muchas veces

debido a esta disparidad de esquemas, resulta incompresible su lectura e interpretación adecuada por parte de los diversos programas de computó.

“Como solución a estos problemas, desde 1991 se ha acordado internacionalmente utilizar la norma **Unicode**, que es una gran tabla, que en la actualidad asigna un código a cada uno de los más de cincuenta mil símbolos, los cuales abarcan todos los alfabetos europeos, ideogramas chinos, japoneses, coreanos, muchas otras formas de escritura, y más de un millar de símbolos especiales”⁹.

Para el caso de los Metadatos y Documentos de BDCOL, buscando solucionar problemas de incompatibilidad en los metadatos y documentos intercambiados y recuperados, se ha observado que no es conveniente utilizar sistemas de codificación incompatibles entre sí (ASCII, BCD, latin-1, ISO 8859-1, euc*, cp* y cientos más), por tanto se especifica que **TODAS las instituciones participantes en BDCOL** deben codificar sus metadatos y documentos en **UNICODE** en cualquiera de estas dos versiones:

- UTF-8
- UTF-16

- **Estándares de Metadatos:**

Para la especificación de un modelo de metadatos común para BDCOL, se determinaron en primera instancia las **colecciones** documentales básicas que permitirían determinar unos elementos constitutivos, estas tipologías son:

Tesis: Incluye todos los trabajos o producciones relacionadas con proyectos de grado o fin de carrera, proyectos de especialización, tesis de maestrías y doctorados.

Artículos: Se refiere a la producción académica, de investigación y divulgación de Universidades, Grupos de Investigación, Investigadores, etc. Esta Colección incluye artículos de revistas, informes de investigación, e-papers/e-prints.

Seriadas: Esta colección incluye publicaciones vistas como un todo, que presentan una periodicidad de publicación (como revistas, periódicos, guías, boletines, etc..) que agrupan la producción académica e intelectual de investigación y que permite la divulgación de resultados de Universidades, Grupos de Investigación, Investigadores, etc..

Audiovisuales: Se refiere a todo material digital cuyo contenido integre animaciones, imágenes, video y sonido en forma prioritaria sobre el componente texto.

Patrimonial: Este conjunto de metadatos incluye aquellos documentos históricos de archivos de instituciones públicas o privadas cuyo contenido sea de interés potencial para investigadores o ciudadanos colombianos. Por ejemplo, actas de

⁹ Colaboradores de Wikipedia. Codificación de caracteres [en línea]. Wikipedia, La enciclopedia libre, 2009 [fecha de consulta: 23 de mayo del 2009]. Disponible en <http://es.wikipedia.org/w/index.php?title=Codificaci%C3%B3n_de_caracteres&oldid=26585903>.

constitución, correspondencia de personajes influyentes en la historia de Colombia, Informes anuales institucionales, entre otros.

General: Se refiere a una colección digital general, en la cual está enmarcado cualquier tipo de material no seriado/periódico, que no pertenece a ninguna colección definida anteriormente para BDCOL.

A partir de estas colecciones, se evaluaron los posibles elementos específicos que deberían manejar cada una de estas, tomando como base los siguientes modelos:

- El estándar Dublin Core simple (ISO 15836:2003)
- La norma Dublin Core con calificadores (DCTERMS)
- La norma para el manejo de Tesis ETD-MS
- La norma de descripción archivística ISAD(G)

Después de analizar los elementos de cada uno de estos modelos de metadatos y compararlos con las necesidades propias de BDCOL, se generó el siguiente mapa de elementos/colecciones, el cual describe el perfil de aplicación de metadatos propio para BDCOL:

| | | | | | |
|--|--|--|--|--|--|
| Relación (dc:relation) | Relación (dc:relation) | | Relación (dc:relation) | Relación (dc:relation) | Relación (dc:relation) |
| Derechos de acceso : (dcterms.accessRights) Open Acces | Derechos de acceso : (dcterms.accessRights) Open Acces | Derechos de acceso : (dcterms.accessRights) Open Acces | Derechos de acceso : (dcterms.accessRights) Open Acces | Derechos de acceso : (dcterms.accessRights) Open Acces | Derechos de acceso : (dcterms.accessRights) Open Acces |
| Título Alternativo (dcterms.alternative) | Título Alternativo (dcterms.alternative) | Título Alternativo (dcterms.alternative) | Título Alternativo (dcterms.alternative) | Título Alternativo (dcterms.alternative) | Título Alternativo (dcterms.alternative) |
| Colaborador (dc:contributor) | Colaborador (dc:contributor) | Asesor/ director de la tesis (dc:contributor) | Colaborador (dc:contributor) | Colaborador (dc:contributor) | Colaborador (dc:contributor) |
| | | Resumen (dcterms:abstract) | | | |
| | | | | Notas del Archivista (ISADG:notes) | |
| | | | | Localización (dc:description.location) | |
| | | | | | Dueño de los derechos (dcterms.rightsholder) |
| | | Editor (thesis.degree.grantor) | | | |
| | | Título obtenido (thesis.degree.name) | | | |
| | | Nivel de estudio: (thesis.degree.level) | | | |
| | | Departamento / Programa (thesis.degree.discipline) | | | |

Tabla 3. Elementos de Metadatos BDCOL

- **Identificación de documentos de acceso abierto**

Para que el sistema metabuscador/recolector de BDCOL pueda identificar solo aquellos objetos digitales que sean de Acceso Abierto (Open Access), se ha encontrado indispensable implementar un elemento en el modelo de metadatos que identifique esta característica. Por consiguiente se ha decidido utilizar el elemento del esquema DCTERMS denominado "accessRights" (**DCTERMS.accessRights**), que a su vez corresponde con un calificador específico del elemento Dublin Core "rights" (**DC.rights.accessRights**), este elemento permite obtener información sobre quién puede acceder al recurso o una indicación sobre su seguridad; adicionalmente puede incluir información respecto al acceso o restricciones basadas en la privacidad, seguridad u otras reglas.

Para el caso de BDCOL, en este elemento se debe especificar quién puede acceder al recurso o una indicación de su estado de acceso según las siguientes recomendaciones:

- Se utilizarán 3 niveles validos:
 - **Restricted:** Para el caso de documentos restringidos en su totalidad
 - **Limited Access:** Para el caso de documentos parcialmente restringidos, y se permite el acceso a partes específicas del documento.
 - **Open Access:** Si se permite acceso irrestricto a todo el objeto digital de información

Si no existe este elemento "dc.rights.accessRights" ó "dcterms.accessRights", se asumirá que el documento corresponde con el nivel más alto "**Open Access**" (Acceso Abierto) y será tratado como tal.

Ejemplo HTML:

- `<meta name="dc.rights.accessRights" content=" Open Access" />`
- `<meta name="dcterms.accessRights" content=" Open Access" />`

Ejemplo XML:

- `<dcterms:accessRights> Open Access </dcterms:accessRights>`

- **Formatos de Elementos por Defecto**

Fechas: Este elemento tiene su equivalente en Dublin Core simple denominado "date", por lo cual se puede utilizar indistintamente el prefijo DC ó DCTERMS, sin embargo, en términos de normalización, es aconsejable utilizar básicamente el prefijo DC. Por defecto para **BDCOL** (a menos que se indique lo contrario con calificadores DCTERMS), la fecha se asociará con la creación o la disponibilidad del recurso

Para **BDCOL**, se recomienda codificar el valor de la fecha según la especificación **W3CDTF**, la cual consiste en un conjunto de reglas para la codificación para fechas y tiempos desarrolladas por el W3C- basado en ISO 8601 [W3CDTF <http://www.w3.org/TR/NOTE-datetime>]

Para registrar fechas (Años, Meses y Días), se utilizará la notación normalizada según la **ISO 8601** de la forma AAAA-MM-DD.

Para registrar tiempo (Horas, Minutos y Segundos), se utilizará la notación especificada en la norma **ISO 8601** que incluye las horas de la forma HH:MM:SS.

Para registrar fechas completas que incluyan años, meses, días, horas, minutos y segundos, se utilizará la notación especificada en la norma **ISO 8601** que incluye la adición de los signos T y Z de la forma: AAAA-MM-DDTHH:MM:SSZ.

En el caso que se esté especificando un periodo tiempo se puede utilizar el esquema de codificación DCMI Period (start=2000-01-26; end=2000-02-20;)

Ejemplo HTML:

```
<meta name="DC.date" scheme="DCTERMS.W3CDTF" content="2008-07-11" />
<meta name="DC.date" scheme="ISO8601" content="2008-07-11T10:15:20Z" />
```

Ejemplo XML:

```
<dc:date xsi:type="dcterms:Period">start=2000-01-26; end=2000-02-20;</dc:date>
<dc:date>1977-07-11</dc:date>
```

Identificadores: Este elemento tiene su equivalente en Dublin Core simple denominado "identifier", por lo cual se puede utilizar indistintamente el prefijo DC ó DCTERMS, sin embargo, en términos de normalización, es aconsejable utilizar básicamente el prefijo DC. La práctica más recomendable es identificar el recurso por medio de una cadena de caracteres o por un número conforme a un sistema formal de identificación.

Algunos sistemas identificación formal de recursos son:

- El Uniform Resource Identifier (URI) (que incluye el Localizador Uniforme de Recursos (URL) y el Localizador Uniforme de Nombres (URN)).
- URL Persistente (PURL)
- OpenURL
- El Digital Object Identifier (DOI)
- Handle ID
- El International Standard Book Number (ISBN)
- El International Standard Serial Number (ISSN)

Este elemento puede ser repetible, sin embargo por defecto para **BDCOL se debe tener en cuenta que:** la primera ocurrencia de este elemento corresponde con la URL única del objeto digital y debe apuntar al objeto en sí o a la especificación de los metadatos del mismo.

En el contexto de BDCOL se recomienda:

- Utilizar como identificador un **enlace persistente** tipo **PURL, HANDLE, DOI**
- Colocar la mayor cantidad de identificadores posibles, entre ellos la dirección URL del recurso mismo ó del registro de metadatos asociado.
- Así mismo cuando se utiliza el protocolo OAI-PMH para diseminar estos recursos, se debe asignar un identificador URN único a cada registro de metadatos.

Ejemplo HTML:

```
<meta name="DC.identifier" content="http://docudomo.org/pipe/" />
<link rel="DC.identifier" scheme="DCTERMS.URI" href="http://www.lasalle.edu.co" />
<meta name="DC.Identifier" scheme="DOI" content="10.1029/2003JP002251">
<meta name="DC.Identifier" content="RCSPC-FISHER-Y3045C-053">
<meta name="DC.identifier" content="http://www.bdcoll.org:8080/index.php">
<meta name="DC.identifier" scheme="HANDLE" content="9781873671009">
```

Ejemplo XML:

```
<dc:identifier xsi:type="dcterms:URI"> http://www.ukoln.ac.uk/</dc:identifier>
<dc:identifier xsi:type="dcterms:ISBN">ISBN</dc:identifier>
<dc:identifier xsi:type="dcterms:ISSN">ISSN</dc:identifier>
<dc:identifier xsi:type="dcterms:URI">urn:oai:www.lasalle.edu.co:kobv:11-1234567</dc:identifier>
```

Idioma: Este elemento tiene su equivalente en Dublin Core simple denominado "language", por lo cual se puede utilizar indistintamente el prefijo DC ó DCTERMS, sin embargo, en términos de normalización, es aconsejable utilizar básicamente el prefijo DC.

Aunque para el control de los contenidos de este elemento se puede utilizar las siguientes normas y estándares de contenido:

- El estándar ISO 639-2 e ISO 639-3: Códigos para la representación de nombres de lenguajes.
- La norma RFC 1766 : Etiquetas de Internet para la identificación del lenguaje, especifica un código de dos caracteres tomado de ISO 639, seguidos opcionalmente por un código de país de dos caracteres tomado de ISO 3166.
- RFC3066: Etiquetas para la identificación de lenguajes, especifica una sub-etiqueta que tiene un código de dos letras tomados de la ISO 639-1 o un código de tres letras tomado de la ISO 639-2, seguido opcionalmente de un código de país de dos letras tomado de la ISO 3166

La mejor práctica recomendada por BDCOL es utilizar un vocabulario controlado, como **RFC 4646** (<http://www.ietf.org/rfc/rfc4646.txt>), la cual reemplaza a los RFC anteriores 3066 y 1766. Esta norma se utiliza en conjunción con la norma ISO 639-2 (http://www.loc.gov/standards/iso639-2/php/code_list.php), y define etiquetas de dos y tres letras para identificar el idioma principal, con subetiquetas opcionales.

El uso del **RFC 4646** incluye las siguientes funcionalidades y características¹⁰:

1. Las subetiquetas válidas pueden hallarse todas en un solo lugar, el nuevo registro de la IANA (<http://www.iana.org/assignments/language-subtag-registry>).
2. Las posiciones y longitudes de las subetiquetas están fijas, lo que facilita el procesamiento de las etiquetas de idioma.
3. Hay más flexibilidad en relación con los posibles componentes de una etiqueta de idioma.

Ejemplo HTML:

```
<meta name="DC.Language" scheme=" DCTERMS.RFC4646" content = "es-CO">
<meta name="DC.language" scheme="DCTERMS.ISO639-2" content ="spa" />
<meta name="DC.Language" scheme=" DCTERMS.RFC3066" content="fr-FR" />
<meta name="DC.language" scheme="DCTERMS.RFC4646" content="en-US"/>
```

Ejemplo XML:

```
<dc:language xsi:type="dcterms:ISO639-3"> jpn</dc:language>
<dc:language xsi:type="dcterms:RFC4646">en-US</dc:language>
<dc:language xsi:type="dcterms:RFC4646">es-CO</dc:language>
```

Coberturas Geográficas/Espaciales: Este elemento tiene su equivalente en Dublin Core simple denominado “coverage” es su definición de alcance geográfico/espacial, por lo cual se puede utilizar indistintamente el prefijo DC ó DCTERMS, sin embargo en términos de normalización, es aconsejable utilizar básicamente el prefijo DC. Normalmente este tipo de cobertura incluirá la localización espacial (un nombre de un lugar o unas coordenadas geográficas) o la jurisdicción (por ejemplo una denominación de una entidad administrativa).

Para el Caso de **BDCOL**, se recomienda la utilización de las siguientes normas y estándares de contenido:

Nombres de Países: **ISO3166 (DCTERMS:ISO3166 / <http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html>):** Códigos para la representación de los nombres de países

Espacios Geográficos:

- **DCMI Box (DCTERMS:Box / <http://dublincore.org/documents/dcmi-box/>):** La Caja identifica una región del espacio que usa sus límites geográficos
- **DCMI Point (DCTERMS:Point / <http://dublincore.org/documents/dcmi-point/>):** El Punto identifica un punto en el espacio que usa sus coordenadas geográficas

¹⁰ Ishida, Richard. Etiquetas de idioma en HTML y XML. Consorcio W3C. 2006 (Traducción 2008). <http://www.w3.org/International/articles/language-tags/Overview.es.php>

Ejemplo HTML:

```
<meta name="DC.coverage" scheme="DCTERMS.ISO3166" content="IT"/>
```

Ejemplo XML:

```
<dc:coverage xsi:type="dcterms:ISO3166">CO</dc:coverage>  
<dc:coverage xsi:type:"dcterms:point">36 58 27 N, 122 01 47 W</dc:coverage>  
<dc:coverage xsi:type="dcterms:Box">northlimit=-13.71; southlimit=-20.25;  
westlimit=166.52; eastlimit=169.89</dc:coverage>
```

Interoperabilidad Semántica

- **Tipologías Documentales (Nivel Bibliográfico)**

Con el fin de normalizar las diversas tipologías documentales que se utilizarán en el elemento **TIPO (dc.type)** del modelo de metadatos de Dublin Core para que estos sean altamente interoperables, se sugiere la utilización de un vocabulario controlado que proveer cierta estandarización en los contenidos.

El elemento **dc.type** describe la naturaleza, genero, la categoría del contenido del recurso ó su tipología documental. Este elemento incluye la normalización de términos que permita normalizar categorías generales, funciones, géneros ó niveles de agregación del contenido

Inicialmente se **Recomienda** utilizar para **BDCOL** el vocabulario controlado de tipos de recursos de información, ofrecidos por la iniciativa Dublin Core (DCMI Vocabulary / DCMIType / <http://dublincore.org/documents/dcmi-type-vocabulary>), el cual se presenta en la tabla siguiente:

| Nombre del término | Definición |
|-----------------------------|--|
| Colección (Collection) | Una colección es un recursos de información que abarca varios recursos, en el sus partes pueden ser descritas y navegadas separadamente. Ejemplo: memorias de congresos, una compilación de recursos, etc. |
| Conjunto de Datos (Dataset) | Corresponde a un conjunto de datos codificados con una estructura documental predefinida, debido a esta, puede ser utilizado directamente en una base de datos. Ejemplo: un reporte contable, una lista de clase ó un listado de usuarios. |
| Evento (Event) | Un evento es un suceso no persistente basado en el tiempo. El tipo de información que poseerá el metadato corresponderá con información que permita presentar los detalles principales del evento asociado (ubicación, duración, responsables, temática, etc.). Ejemplo: clase |

| | |
|--|--|
| | virtual, exhibición, videoconferencia, lanzamiento de un sitio Web, talleres, etc. |
| Imagen (Image) | Corresponde con la representación de un recurso visual físico o digital diferente al texto. Se puede equiparar a la representación visual de los contenidos multimedia no interactivos. Ejemplo: imágenes y fotografías, pinturas, impresiones, dibujos, planos, mapas, animaciones y videos. |
| Recursos Interactivo (InteractiveResource) | Un recurso interactivo requiere interacción con el usuario para ser utilizado y comprendido. Dependiendo de esta interacción se muestran unos contenidos u otros. Por ejemplo: Chats, mundos virtuales, objetos de aprendizaje, formularios Web, multimedia interactiva, etc. |
| Imagen en Movimiento (MovingImage) | Corresponde a recursos que, al ser visualizados dan la impresión de movimiento, generalmente son un conjunto de imágenes presentadas sucesivamente. Ejemplo: animaciones, películas, programas de televisión, videos, etc. |
| Objeto Físico (PhysicalObject) | Corresponde a recursos físicos, tangibles y que se manipulan en el mundo real. Las representaciones digitales de estos recursos se convierten en imágenes, animaciones, textos, etc. Ejemplo: una casa, un árbol, etc. |
| Servicio (Service) | Corresponde con un sistema que provee uno o más servicios de información. Ejemplos son: un servicio bancario, un servicio de previsión climática, un sistema de préstamo interbibliotecario, un servidor Z39.50, OAI, a un servidor de autenticación. |
| Software (Software) | Corresponde con un programa de computador que realiza unas tareas generales o específicas. Puede estar compilado (.exe) ó aparecer en código fuente (.ast, .php, .cpp, etc.). |
| Sonido (Sound) | Un sonido es un recurso asociado a la sensación en el órgano del oído, producida por el movimiento ondulatorio, debido a la energía que es transmitida por cambios de presión en el aire y otros medios, cuyo contenido es reproducido como audio. Ejemplo: Archivos del tipo MP3, Midi, Wav, etc. |
| Imagen Estática (StillImage) | Corresponde con una representación visual estática. Ejemplo: pinturas, dibujos, diseños gráficos, planos y mapas. |
| Texto (Text) | Un texto es un recurso compuesto por una composición de signos codificado en un sistema de escritura (como un alfabeto) que forma una unidad de sentido. Ejemplo: |

| | |
|--|--|
| | Libros, Cartas, Poemas, Revistas, etc. |
|--|--|

Tabla 4. Vocabulario DCMITYPE

Sin embargo, los componentes de este vocabulario son muy genéricos respecto a las colecciones que maneja BDCOL, por esto, se desarrolló un nuevo vocabulario controlado, denominado **vocabulario de los niveles bibliográficos de la Biblioteca Digital Colombiana (BDCOL type Vocabulary [BDCOLType] / con espacio de nombres BDCOL:<http://www.bdcoll.org/documents/bdcoll-type-vocabulary>, cuyo esquema XSD se encuentra en <http://www.bdcoll.org/documents/bdcoll-type-vocabulary/BDCOLType.xsd>), el cual se recomienda utilizar en todos los documentos objetos digitales que se integren bajo esta iniciativa, que puede ser utilizado de las siguientes maneras:**

- o Reemplazando el vocabulario DCMIType, cuando se utiliza en los metadatos el elemento **dc.type**.
- o Manejándolo como un elemento repetible del campo **dc.type**, indicando preferiblemente mediante el uso de un **Schema** ó un **xsi:type**, que se está utilizando el vocabulario **BDCOLType**.

```

<?xml version="1.0" encoding="UTF-8"?>
- <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns="http://www.bdcoll.org/documents/bdcoll-type-vocabulary/"
  targetNamespace="http://www.bdcoll.org/documents/bdcoll-type-vocabulary/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" elementFormDefault="qualified" attributeFormDefault="unqualified"
  version="1.00.000">
- <xs:annotation>
  <xs:documentation xml:lang="es_CO">Vocabulario Controlado de Tipologías Documentales de BDCOL XML Schema para el
    espacio de nombres: "http://www.bdcoll.org/documents/bdcoll-type-vocabulary/" namespace Creado en: 2009-06-18
    Creado Por: Laureano Felipe Gómez Dueñas(laureanogomez@unisalle.edu.co) Versión: 1.0</xs:documentation>
</xs:annotation>
- <xs:simpleType name="BDCOLType">
- <xs:union>
  - <xs:simpleType>
    - <xs:restriction base="xs:Name">
      <xs:enumeration value="Article"/>
      <xs:enumeration value="Multimedia Resource"/>
      <xs:enumeration value="Book"/>
      <xs:enumeration value="BookPart"/>
      <xs:enumeration value="ConferenceObject"/>
      <xs:enumeration value="Data Set"/>
      <xs:enumeration value="Historical document archive"/>
      <xs:enumeration value="Image"/>
      <xs:enumeration value="Institutional document"/>
      <xs:enumeration value="Journal"/>
      <xs:enumeration value="Learning Object"/>
      <xs:enumeration value="Map"/>
      <xs:enumeration value="Newspaper"/>
      <xs:enumeration value="Patent"/>
      <xs:enumeration value="PeerReviewed"/>
      <xs:enumeration value="Presentation"/>
      <xs:enumeration value="Project"/>
      <xs:enumeration value="Portal - Web Site"/>
      <xs:enumeration value="Simulation"/>
      <xs:enumeration value="Software"/>
      <xs:enumeration value="Sound"/>
      <xs:enumeration value="Technical Report"/>
      <xs:enumeration value="Thesis"/>
      <xs:enumeration value="WorkingPaper"/>
    </xs:restriction>
  </xs:simpleType>
</xs:union>
</xs:simpleType>
</xs:schema>

```

Figura 1. Esquema XSD BDCOLType

Por defecto, la ausencia del indicador de vocabulario controlado utilizado en cada registro de metadatos, indica el uso del vocabulario **BDCOLType**. **Las tipologías documentales que componen el Vocabulario BDCOLType son:**

| Nombre del término | Definición | Ejemplos |
|---------------------|---|---|
| Article | Corresponde con documento compuesto por una composición de signos codificado en un sistema de escritura (como un alfabeto) que forma una unidad de sentido y que pertenece a una publicación periódica. | Artículo de una revista, Noticia de un magazín, boletín ó periódico. |
| Multimedia Resource | Corresponde con un recurso interactivo que incluye, animación, video y sonido, que requiere interacción con el usuario para ser utilizado y comprendido. Dependiendo de esta interacción se muestran unos contenidos u otros. | Chats, mundos virtuales, objetos de aprendizaje, formularios Web, multimedia, etc. |
| Book | Corresponde con documento compuesto por una composición de signos codificado en un sistema de escritura (como un alfabeto) que forma una unidad de sentido y que se publica como un todo en uno o limitado número de volúmenes. | Libros |
| BookPart | Corresponden con una parte de un libro ó un volumen monográfico. | Capítulo de un libro; separata |
| Data Set | Corresponde a un conjunto de datos codificados con una estructura documental predefinida, debido a esta, puede ser utilizado directamente en una base de datos. | Todo tipo de listados y reportes consolidados como un todo, Un reporte contable, una lista de clase ó un listado de usuarios. |
| Historical document | Incluye aquellos documentos | Incluye actas de |

| | | |
|------------------------|---|--|
| archive | históricos de archivos de instituciones públicas o con funciones públicas, cuyo contenido sea de interés potencial para investigadores o ciudadanos colombianos. | constitución, correspondencia de personajes influyentes en la historia de Colombia, informes anuales institucionales, entre otros. |
| Image | Corresponde con la representación de un recurso visual físico o digital diferente al texto. Se puede equiparar a la representación visual de los contenidos multimedia no interactivos. | Incluye imágenes y fotografías, pinturas, impresiones, dibujos y planos. |
| Institutional document | Corresponde con documentos administrativos, de gestión o que reflejen el desarrollo de la identidad institucional de una organización. | PEUL, Librillos de identidad institucional, planes académicos, syllabus. |
| Journal | Incluye la producción académica e intelectual de investigación que permite la divulgación de resultados de Universidades, Grupos de Investigación, Investigadores, etc., como también materiales que presentan una periodicidad en su publicación, tales como revistas, eprints, etc. | Revistas, Folletos, Boletines |
| Learning Object | Recurso digital que puede ser reutilizado en diferentes contextos educativos. | Material Docente, Trabajos de Clase (Alumnos), Presentaciones PPT |
| Legal | Incluye todo tipo de documentos legales, normas, leyes, códigos, minutas, sentencias, etc.. | Normas/Leyes |
| Map | Corresponde con una representación gráfica y métrica | Planos, Esquemas, Afiches |

| | | |
|-------------------|---|--|
| | de una porción de territorio sobre una superficie bidimensional, generalmente plana, pero que puede ser también esférica como ocurre en los globos terráqueos. | |
| Newspaper | Una publicación periódica de prensa escrita es una publicación escrita impresa editada con periodicidad que incluye generalmente noticias y novedades. | Depende de su periodicidad, puede ser diaria (en cuyo caso suele llamarse diario), semanal (semanario), mensual o anual (anuario) |
| Patent | Corresponde con un tipo de documento que es una concesión legal emitida por un gobierno que permite al inventor excluir a otras personas de fabricar, utilizar o vender un invento, declarado como propio, durante el plazo de vigencia de la misma | Patentes |
| Presentation | Corresponde con documentos del tipo presentación/exposición, generalmente desarrollados en un evento académico / científico como un congreso, seminario, etc.. | Este tipo de documentos manejan los formatos PPT, ODT y Flash (SWF). |
| Project | Corresponde con un tipo de documento que contiene un conjunto de secuencia de tareas programadas y planificadas con un fin específico. | Proyectos, Planes estratégicos |
| Portal - Web Site | Corresponde recursos bibliográficos que se completa o modifica por medio de actualizaciones que no permanecen separadas, sino que se integran en un todo. | Bases de datos, páginas web, archivos de datos abiertos, Noticias, blogs, novedades, eventos, expresiones personales e Institucionales en Internet |

| | | |
|------------------|--|--|
| Simulation | Corresponde con un tipo de aplicación altamente interactiva que permite al alumno diseñar o representar un escenario determinado, generalmente mediante el uso de realidad virtual ó sistemas especializados. | Archivos VRML, etc.. |
| Software | Corresponde con un programa de computador que realiza unas tareas generales o específicas. Puede estar compilado (.exe) ó aparecer en código fuente (.ast, .php, .cpp, etc.). | |
| Sound | Un sonido es un recurso asociado a la sensación en el órgano del oído, producida por el movimiento ondulatorio, debido a la energía que es transmitida por cambios de presión en el aire y otros medios, cuyo contenido es reproducido como audio. | Archivos del tipo wav, mp3, Midi, etc.. |
| Technical Report | Corresponde con un informe técnico (también: informe científico) el cual es un documento que describe el proceso, el progreso, o los resultados de la investigación científica o técnica | |
| Thesis | Incluye todos los trabajos o producciones relacionadas con proyectos de grado o fin de carrera, proyectos de especialización, tesis de maestrías y doctorados. | Tesis, Informes de Practicas, Informe de Investigación |
| Working Paper | Un documento de trabajo corresponde con un registro de un proyecto de investigación o actividades de conclusiones | |

| | | |
|--|----------------|--|
| | provisionales. | |
|--|----------------|--|

Tabla 5. Vocabulario BDCOLTYPE

Ejemplo de uso DCMITYPE y BDCOLTYPE en HTML

```
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="collection" />
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
<meta name="DCTERMS.type" scheme="BDCOLType" content="Analítica de Seriado" />
```

Ejemplo de uso DCMITYPE y BDCOLTYPE en XML:

```
<dc:type>Software</dc:type>
<dc:type xsi:type="dcterms:DCMIType">Sound</dc:type>
<dcterms:type xsi:type="dcterms:DCMIType">Collection</dcterms:type>
<dc:type xsi:type="BDCOL:BDCOLType">Tesis</dc:type>
<dcterms:type xsi:type="BDCOL:BDCOLType">Objeto de Aprendizaje</dcterms:type>
```

- **Vocabularios Controlados**

Los esquemas de Codificación de Vocabulario indican que el contenido asociado a un elemento de metadatos, corresponde con los valores provistos por un vocabulario controlado. Para el caso de **BDCOL**, se recomienda utilizar los siguientes vocabularios:

Elemento dc.subject:

- **DDC (Esquema DCTERMS / DCTERMS:DDC):** Corresponde con el Sistema de Clasificación Universal DEWEY (<http://www.oclc.org/dewey/>)
- **LCC (Esquema DCTERMS / DCTERMS:LLC):** Son los esquemas de clasificación del conocimiento desarrollados por la Biblioteca del Congreso de los Estados Unidos de Norteamérica (<http://lcweb.loc.gov/catdir/cpsol/lcco/lcco.html>).
- **LCSH (Esquema DCTERMS / DCTERMS:LCSH):** Son el conjunto de términos (vocabulario controlado) desarrollado por la Biblioteca del Congreso de los Estados Unidos de Norteamérica (<http://authorities.loc.gov/>).
- **MESH (Esquema DCTERMS / DCTERMS:MESH):** Abreviatura usual, en idioma inglés, de "Medical Subject Headings", una lista de encabezamientos más populares en el área de la Medicina (<http://www.nlm.nih.gov/mesh/meshhome.html>).
- **NLM (Esquema DCTERMS / DCTERMS:NLM):** El conjunto de recursos conceptuales (sistema de clasificación) especificados por la Biblioteca Nacional de Medicina (<http://wwwcf.nlm.nih.gov/class/>).
- **UDC (Esquema DCTERMS / DCTERMS:UCD):** Corresponde con el sistema de clasificación decimal universal. (<http://www.udcc.org/>)
- **LEMB (Esquema BDCOL / BDCOL:LEMB):** Listado de Encabezamiento de Materias para Bibliotecas (<http://www.lembdigital.com/>)
- **TAC (Esquema BDCOL / BDCOL:TAC):** Tesoro Ambiental para Colombia (<http://www.minambiente.gov.co/tesauro/naveg.htm>)
- **TDC (Esquema BDCOL / BDCOL:TDC):** El tesoro sobre biodiversidad de Colombia del El Sistema de Información sobre Biodiversidad de Colombia -SIB-

(<http://www.siac.net.co/sib/tesauros2/WebModuleTesauros/index.jsp>)

- **TEE (Esquema BDCOL / BDCOL:TEE):** Tesoro Europeo de la Educación (<http://www.redined.mec.es/consultas.php>)
- **UNESCOTHESES (Esquema BDCOL / BDCOL:UNESCOTHESES):** El Tesoro de la UNESCO es una lista controlada y estructurada de términos para el análisis temático y la búsqueda de documentos y publicaciones en los campos de la educación, cultura, ciencias naturales, ciencias sociales y humanas, comunicación e información (<http://databases.unesco.org/thessp/index.html>)
- **DECS (Esquema BDCOL / BDCOL:DECS) :** Descriptores en Ciencias de la Salud, corresponde al tesoro desarrollado por Bireme/OPS que está basado en MESH (<http://www.bireme.org>).

Elemento dc.format:

- **IMT (Esquema DCTERMS / DCTERMS:IMT):** Corresponde con el listado de extensiones de archivo desarrollado por la IANA (Internet Assigned Numbers Authority) y corresponden con los tipos de archivos que especifica los diversos tipos de contenidos, subtipos de contenidos, el conjunto de caracteres utilizado, tipos de acceso, y la conversión de los valores de MIME (Multipurpose Internet Mail Extensions - Extensiones de Correo Internet Multipropósito).

Elemento dc.coverage

- **TGN (Esquema DCTERMS / DCTERMS:TGN):** Corresponde con el Tesoro de los nombres Geográficos (http://www.getty.edu/research/conducting_research/vocabularies/tgn/index.html)

Ejemplos HTML:

```
<meta name="DC.subject" scheme="BDCOL:LEMB" content="guerra del golfo" xml:lang="es" />
<meta name="DC.subject" scheme="DCTERMS:LCSH" content="Persian Gulf War, 1991" xml:lang="en" />
<meta name="DC.Subject" scheme="DCTERMS:MESH" content="Myocardial Infarction">
<meta name="DC.Subject" scheme="DCTERMS:LCSH" content="Vietnamese Conflict">
<meta name="DC.Subject.Classification" scheme="DCTERMS:DDC" content="158.25">
<meta name="DCTERMS.format" scheme="DCTERMS:IMT" content="text/html " />
```

Ejemplo XML:

```
<dc:subject xsi:type="bdcol:LEMB">Historia de Colombia</dc:subject>
<dc:subject xsi:type="dcterms:MESH" />Smoking</dc:subject>
<dcterms:subject xsi:type="bdcol:TEE" />Educación Primaria</dcterms:subject>
<dc:coverage xsi:type="dcterms:TGN">Colombia</dcterms:coverage>
<dcterms:format xsi:type="dcterms:IMT">audio/x-wav</dcterms:format>
```

Interoperabilidad Estructural

- **Implementación del protocolo OAI-PMH**

La iniciativa BDCOL especifica el uso del protocolo **OAI-PMH** como el protocolo básico y mínimo para la integración y recolección de metadatos. De esta manera es **Obligatorio** que los diferentes SIDOCS, REDIS y BIDIS que deseen pertenecer a **BDCOL** sean conformes con el protocolo OAI-PMH (v. 2.0, junio 2002) emitido por la Iniciativa de Archivos Abiertos (Open Archives Initiative <http://www.openarchives.org>).

De acuerdo con este modelo, cada uno de los SIDOCS, REDIS y BIDIS que recolecte **BDCOL** serán entendidos como “proveedores de datos”, BDCOL será entendido como el “proveedor de Servicios” de los contenidos académico y científicos de Colombia y al mismo tiempo BDCOL será “proveedor de datos” de colecciones unificadas, para ser recolectado por Bibliotecas Digitales, buscadores y sistemas de recuperación de información internacionales.

Así mismo aunque se privilegia la recolección de metadatos utilizando el esquema de metadatos `oai_dc`, **se recomienda** implementar el esquema global de metadatos de **BDCOL** denominado `oai_bdcoll` (Espacio de Nombres namespace: <http://www.bdcoll.org/documents/metadata/>, Esquema XSD http://www.bdcoll.org/documents/metadata/oai_bdcoll.xsd), el cual integra todos los elementos de metadatos indicados para BDCOL.

Adicionalmente **BDCOL** proveerá la infraestructura lógica que permitirá diferenciar las diversas colecciones recolectadas y su posterior diseminación nacional e Internacional tanto individual como colectivamente. Para esto se **requiere** que las colecciones suministradas por cada uno de los participantes de **BDCOL** contengan:

- Cada registro de Metadatos recolectado deberá contener dos Elemento de identificación unívocos y con características especiales:

- Un Identificador Global del registro para ser identificado en los procesos de recolección según las indicaciones del protocolo OAI-PMH, utilizando la especificación **URN**, y que deber ser visualizado cuando se ejecuta el verbo “ListIdentifiers”. Por Ejemplo:

```
<identifier>oai:tegra.lasalle.edu.co:10185/4</identifier>
```

- Un identificador tipo URL que será puesto en la primera ocurrencia del elemento “**dc.identifier**”, el cual contendrá un enlace para visualizar el registro directamente desde su fuente original (desde el proveedor de datos). Se debe tener en cuenta que esta URL debe ser **PERMANENTE** y debe ser válida en **TODO** momento. Por ejemplo:

```
<dc.identifier>http://hdl.handle.net/10185/5</dc.identifier>
```

- Cuando un objeto digital sea borrado/eliminado del SIDOC, REDIS ó BIDI, debe registrarse en la recolección según los lineamientos del protocolo OAI-PMH, en el cual se debe colocar el atributo “**deleted**” en el elemento “**header**” (cabecera) del registro desplegado.

```
<record>
<header status="deleted" >
<identifier>oai:tegra.lasalle.edu.co:10185/4</identifier>
<datestamp>2007-08-25T15:19:53Z</datestamp>
<setSpec>hdl_10185_3</setSpec>
</header>
</record>
```

Interoperabilidad de Infraestructura

- **Protocolo HTTP y Direcciones IP de Internet y RENATA:**

La iniciativa **BDCOL** requiere el uso del protocolo base de Internet **TCP/IP**, además de utilizar el protocolo para la transferencia de HiperTexto denominado **HTTP**, para el acceso a los objetos digitales que integran los documentos, metadatos, así como el uso del protocolo **OAI-PMH**.

Para esto se recomienda utilizar el protocolo HTTP con su puerto por defecto (80), con el fin de evitar problemas de recolección y visibilidad ocasionadas por algún tipo de problema de seguridad de los Muros de Fuego (FireWall) ó los filtros de Página web (Proxy). Así mismo se recomienda configurar la codificación por defecto de los documentos HTTP en los servidores web con el atributo "utf-8"

Así mismo, para optimizar los tiempos de intercambio de información entre SIDOCS, REDIS y BIDIS y facilitar la recolección por parte del metabuscador de BDCOL, **se recomienda** el uso de direcciones IP que sean de uso transparente en la **red RENATA/CLARA/Internet2/GEANT2**.

- **ID única y acceso directo a los Documentos Completos**

Con el fin de validar el acceso libre e irrestricto a los documentos de los distintos SIDOCS, REDIS y BIDIS, a través del sistema integrador / recolector de **BDCOL**, **se recomienda** que todos los **documentos y metadatos** tengan un acceso directo (URL lógica que se pueda acceder mediante el protocolo HTTP), tanto para su consulta como su posterior descarga. Adicionalmente, este acceso directo servirá para evaluar futuros esquemas de replicación y preservación de objetos digitales y otros servicios de valor añadido.

Se recomienda que se utilicen mecanismos de persistencia y preservación a largo plazo en el uso de las URL y los identificadores de acceso utilizados por cada uno de los REDIS y BIDIS, como: PURL, HANDLE, DOI. Adicionalmente se recomienda la utilización de servidores de enlaces OpenURL que permitan rastrear y localizar los distintos recursos en caso que se llegue a presentar algún problema con las URL.

- **Software**

La iniciativa **BDCOL recomienda** para la implementación de SIDOCS, REDIS y BIDIS, utilizar software de acceso abierto (Software Libre), el cual permita implementar todas las normas y especificaciones de interoperabilidad recomendado por **BDCOL**, los programas de Software Recomendados son:

| Software | Tipos de Documentos |
|--|---|
| DSPACE (http://www.dspace.org/) | Principalmente maneja documentos monográficos, tesis, reportes, patentes, etc.. |
| EPRINTS (http://www.eprints.org) | Principalmente maneja documentos monográficos, tesis, reportes, patentes, etc.. |
| OJS /Open Journal System (http://pkp.sfu.ca/?q=ojs) | Este software está especialmente diseñado para manejar publicaciones seriadas y todo el proceso editorial asociado |
| OCS / Open Conference System (http://pkp.sfu.ca/?q=ocs) | Principalmente maneja presentaciones y documentos pertenecientes a eventos generalmente periódicos que incluyen congresos, conferencias, talleres, etc.. |
| Greenstone (http://www.greenstone.org) | Principalmente maneja documentos monográficos, tesis, reportes, patentes, etc.. Además es muy utilizado para colecciones de documentos multimediales, incluidos animaciones, imágenes y sonidos |
| CWIS (http://scout.wisc.edu/Projects/CWIS/) | Es principalmente utilizado para manejar colecciones digitales referenciales que incluyen enlaces a sitios web, portales, grandes documentos multimedia que son administrados por un sistema especializado. |

Tabla 6. Software Recomendado para REDIS y BIDIS

Bibliografía

Berners-Lee, Tim.; Connolly, D. Y Swick, R (1999) Web Architecture: Describing and Exchanging Data. <<http://www.w3.org/199/04a/WebDat>>.

Colaboradores de Wikipedia. Codificación de caracteres [en línea]. Wikipedia, La enciclopedia libre, 2009 [fecha de consulta: 23 de mayo del 2009]. Disponible en <http://es.wikipedia.org/w/index.php?title=Codificaci%C3%B3n_de_caracteres&oldid=26585903>

Consejo Superior de Informática de La Comisión Europea. Marco Europeo de Interoperabilidad Versión 1.0, <<http://europa.eu.int/idabc> >

García Martínez, Ana María. Definición y estilo de los objetos de información digitales y metadatos para la descripción. Boletín de la asociación Andaluza de bibliotecarios, N. 63, Junio 2001, p.23-47

Gómez Dueñas Laureano Felipe. Interoperabilidad en los Sistemas de Información Documental (SID): la información debe fluir. Revista Códice. Vol 3 No. 1. 2007.

Krsulovic, E. Blog de la Web Semántica. <<http://www.dcc.uchile.cl/~ekrsulov/prj/ws-blog/> >

Ishida, Richard. Etiquetas de idioma en HTML y XML. Consorcio W3C. 2006 (Traducción 2008). <<http://www.w3.org/International/articles/language-tags/Overview.es.php>>

Melero Melero, Remedios; Barrueco Cruz, José Manuel. Acceso Abierto y Repositorios de Documentos. Unidades de Autoformación SEDIC. <http://www.sedic.es/autoformacion/acceso_abierto/presentacion.htm>

Méndez Rodríguez, Eva M^a. La descripción de documentos electrónicos a través de metadatos: una visión para la Archivística desde la nueva e-Administración. Revista d'Arxius, 2003, pp. 47-82. <http://e-archivo.uc3m.es:8080/dspace/bitstream/10016/878/1/EMendez_Arxius.pdf>

San Segundo Manuel, Rosa. Organización del Conocimiento en Internet: Metadatos Bibliotecarios Dublin Core. VI Jornadas Españolas de Documentación, FESABID 98. <<http://www.uag.mx/eci/infosource/Articulos/Profinfo/organizacionconocimiento.pdf> >