

METADATOS : estado de la cuestión

Tesis de la Licenciatura en Bibliotecología y Documentación

María Carolina Rojas
Directora: Lic. Silvia Sleimen



FACULTAD DE HUMANIDADES
UNIVERSIDAD NACIONAL DE MAR DEL PLATA

2010

PÁGINA DE ACEPTACIÓN

METADATOS: Estado de la cuestión

Tesis presentada

por

BIBL. DOC. M. CAROLINA ROJAS

Aprobada en contenido y estilo por:

Lic. Silvia R. Sleimen, Directora

César O. Archuby, miembro del jurado

Gustavo Liberatore, miembro del jurado

Licenciatura en Bibliotecología y Documentación

Dpto. de Documentación / F.H. / UNMDP

Servicio de Información Documenta
Dra. Liliana B. De Boschi
Facultad de Humanidades
U.N.M.D.P.

Agradecimientos

A Eva Méndez Rodríguez, por su colaboración especializada, desinteresada y continua.

A Víctor Herrero, por sugerirme el tema de investigación.

A mis compañeros de trabajo Silvia Sleimen, Mela Bosch, Marcela Coringrato y Andrés Vuotto por sus continuos aportes.

A Rosalía Baltar, que revisó, muy generosamente, la escritura y organización de este documento.

A todos los que a lo largo de estos años no dejaron de apoyarme y me acompañaron en este proceso y me alentaron aplicar los conocimientos adquiridos, en especial a Noemí Conforti, César Archuby, Gustavo Liberatore, María Coira, Luis Porta y tantos otros de los que me estaré olvidando.

A mis amigos, a mi familia y a Claudio que nunca dejan de apoyarme y acompañarme en lo que necesito.

RESUMEN:

Los metadatos implican una nueva forma de estructurar los objetos de información digitales, como alternativa tecnológica para mejorar u optimizar su recuperación en contextos de red, fundamentalmente en *Internet* y en la *World Wide Web*, que han generado en los últimos años un debate profesional y un desarrollo vertiginoso, lo que nos ha llevado a elegirlos como objeto de estudio principal de esta investigación.

Se realiza una revisión bibliográfica acerca de los metadatos, sus tipos, funciones, esquemas y nivel de utilización, para mejorar la recuperación de recursos de información digital. Con ese fin, y para ejemplificar su aplicabilidad, se presenta un prototipo de uso de estas estructuras para la descripción y búsqueda de información en el sitio web del Departamento de Documentación de la Facultad de Humanidades de la Universidad Nacional de Mar del Plata (Argentina).

DESCRIPTORES:

Internet (World Wide Web) / Metadatos / Metainformación / Información digital / Documentación digital / Procesamiento de la información / Recuperación de información / Estructura de contenidos Web / Lenguajes de marcado / Sitios Web / LICAD / Motores de búsqueda / Buscadores locales

RESPONSABLES DEL PROYECTO

Grupo ejecutor: Bibl. Doc. M. Carolina Rojas

Directora: Lic. Silvia R. Sleimen

Entidad responsable: Dpto. de Documentación / F.H. / UNMDP

Tabla de Contenido

1. Introducción	3
2. Marco teórico, planteo, objetivo y método de la investigación.....	5
2.1 Ubicación del Proyecto	5
2.2. Probable impacto de la investigación	6
2.3. Objetivos de la investigación	7
Objetivo general:	7
Objetivos específicos:.....	7
2.4. Formulación del problema	8
2.5. Antecedentes, estado actual de la cuestión y marco teórico.....	9
2.6 Diseño metodológico	15
2.7 Administración del proyecto.....	18
Cronograma de actividades:.....	18
Recursos disponibles	19
Duración	19
Aporte e impacto de los resultados	19
3. Los metadatos: definición, tipos y funciones	20
Contexto para la gestión de metadatos (Greenberg)	25
4. Esquemas de metadatos	33
4.1. La elección de un esquema de metadatos	35
4.1.1. Consideraciones generales	35
4.1.2. Propuestas y modelos de metadatos	36
4.1.3. Software para creación, implementación y gestión de metadatos	38
4.1.4. Etapas del proceso de implementación de metadatos.....	40
4.1.5. La elección de un formato de metainformación para el sitio Web de la LICAD	43
5. El Modelo DC (ISO Standard 15836-2003 - ANSI/NISO Z39.85-2007 - IETF RFC 5013-2007).....	44
6. Aplicación de metadatos al sitio Web de la LICAD.....	61
7. Conclusiones y reflexiones finales.....	66
8. ANEXOS	68
A. Código en JavaScript del buscador	69
B. Arrays de búsqueda para el Módulo de Procesamiento de la información.....	81
9. Glosario.....	86
10. Bibliografía.....	91

Índice de ilustraciones

Ilustración 1 - Contexto para la gestión de metadata. Trad. por la autora, tomado del artículo de Jane Greenberg	25
Ilustración 2 - DataWareHouse – Asociación de metadatos embebidos.....	29
Ilustración 3- DataWareHouse – Asociación de metadatos along-with.....	29
Ilustración 4 - DataWareHouse – Asociación de metadatos service bureau.....	30
Ilustración 5 - DataWareHouse – Asociación de metadatos wrapped.....	30
Ilustración 6 - Relación de estándares de metadatos.....	37
Ilustración 7 - Modelo de etiqueta cualificada de Metadatos según formato DCMI	46
Ilustración 8 - Cabecera HTML en DC.....	47
Ilustración 9 - Cabecera en RDF un archivo HTML con metadatos DC.....	60
Ilustración 10 - Metabot Professional. Generación de un proyecto de edición de etiquetas – selección de esquema (en este caso “Basic Tags” y “Dublin Core”).....	62
Ilustración 11 - Metabot Professional. Generación de un proyecto de edición de etiquetas – selección de directorio.....	62
Ilustración 12 - Metabot Professional. Generación de un proyecto de edición de etiquetas - conjunto de documentos sin editar	63
Ilustración 13 - Metabot Professional. Generación de un proyecto de edición de etiquetas - conjunto de documentos editados	63
Ilustración 14 - Metabot Professional. “HEAD” de uno de los documentos sin procesar	63
Ilustración 15 - Metabot Professional. “HEAD” de uno de los documentos procesados	63
Ilustración 16 - Código buscador en JavaScript	64
Ilustración 17 - Arrays de búsqueda, con los metadatos extraídos de los archivos componentes del sistema	64
Ilustración 18 - Visualización del buscador en un navegador.....	65
Ilustración 19 - Visualización de los resultados de búsqueda	65

Índice de tablas

Tabla 1 - Tipos de metadatos	24
Tabla 2- Concepto de Generación Automática de Metadatos.....	27

1. Introducción

El desarrollo de las tecnologías implica en forma permanente y constante nuevos desafíos a la hora de relacionarse con las diferentes disciplinas. En el caso de las ciencias de la información, con el advenimiento de Internet, se plantean cuestiones en torno a la organización, gestión y recuperación efectiva de la información electrónica.

Entendemos que el caos de información que se ha suscitado con la fácil aplicabilidad de los lenguajes de edición de páginas web, sobre todo con la capacidad de edición *WYSIWYG* (*What You See Is What You Get*, lo que ves es lo que obtienes) -aplicación que permite ver el documento en su presentación final al momento de su procesamiento/creación; posibilita que cualquier usuario de la Red produzca y publique contenidos, sin que estos necesariamente cumplan con normas bien definidas-, impide una eficaz localización de recursos.

Para subsanar dicho problema se recurrió a la utilización formal de herramientas de descripción de propiedades previstas por quienes desarrollaron los formatos de archivos informáticos.

Nos ocupamos, desde el punto de vista documental, de la descripción de recursos electrónicos, mediante la utilización de esquemas de metadatos¹. Se

1 Metadatos: información acerca de la información; son datos sobre las características de un recurso de información, describen sus relaciones, y permiten su descubrimiento (recuperación) y uso efectivo. (Burnett, 1997)

tiene presente en todo momento que Internet funciona como una gran *base de datos*, ya que se sustenta en una serie de protocolos que permiten compartir datos almacenados en diferentes ordenadores y con diferentes plataformas sin que esto sea un inconveniente en el sistema de comunicación; esto obliga a que la información disponible tenga estructuras de datos definidas, en su nivel más simple de representación, y en especial si prestamos atención al código fuente de las páginas web.

Las funciones de descripción, tanto física (forma) cuanto lógica (representación de contenidos) que tienen y cumplen los metadatos facilitan la tarea de los indizadores automáticos (programas que representan el contenido conceptual de los documentos con vistas a su recuperación), además de tener un universo controlado de documentos electrónicos.

La exhaustividad que la aplicación de esta tecnología confiere a los sitios web permite el desarrollo de la *web semántica*, y a su vez, nos acerca a la aproximación de una *sociedad de la información por y para todos*.

A través de un proceso de investigación exploraremos diversas fuentes de información, impresas y electrónicas, para definir qué son los metadatos, sus características, funciones, aplicación y software. Y para completar este estudio se presenta un caso práctico de aplicación de metadatos en un sitio web educativo, el de la Licenciatura en Bibliotecología y Documentación A Distancia (LICAD) del Departamento de Documentación de la Facultad de Humanidades de la Universidad Nacional de Mar del Plata.

2. Marco teórico, planteo, objetivo y método de la investigación

En este capítulo introductorio, fundamental en un trabajo de esta naturaleza, se pretende, por una parte justificar la elección del tema, el marco teórico en el que se ubica, los antecedentes y presupuestos de los que partimos para abordarlo y los límites que establecemos en la investigación; y por otra, se definen los objetivos (generales y específicos) que se intenta alcanzar, a partir del análisis de variables concretas.

2.1 Ubicación del Proyecto

- a) Área de investigación: Procesamiento de la información.
- b) Materia de la investigación: Automatización, descripción y recuperación de la información
 - Temas generales: Información y Documentación digital; Internet (World Wide Web); Descripción de la información (catalogación, clasificación, indización); Recuperación de la información; Estructuración de la información Web; Marcado de documentos (HTML, XML, SGML).
 - Tema específico: Metadatos.
- c) Cobertura del proyecto: Revisión bibliográfica del concepto de metadatos tanto en fuentes impresas cuanto electrónicas, su construcción, desarrollo, aplicaciones y prospectiva de utilización en el contexto de la información electrónica en Internet, desde 1998 a 2005. Además, se presenta un ejemplo de aplicación que permita la utilización de un esquema básico de

metainformación en el sitio web de LICAD (Licenciatura en Bibliotecología y Documentación A Distancia, Departamento de Documentación, Facultad de Humanidades, Universidad Nacional de Mar del Plata), de tal forma que facilite una recuperación de información más precisa y pertinente.

- d) Campo de interés: Dirigida a los profesionales de la información y en especial a los que se dedican a la descripción y recuperación de información en Internet, así como a los encargados de la creación de sitios y recursos web. Adscribimos con este trabajo a la nueva tendencia de la organización de la información digital basada en la información semi-estructurada y en el mercado de documentos.
- e) Tipo de investigación: Estudio exploratorio-descriptivo, diseño transeccional descriptivo.

2.2. Probable impacto de la investigación

Esta investigación responde a un requisito académico de la Licenciatura en Documentación del Dpto. de Documentación / F.H. / UNMDP, y pretende acercar a los bibliotecarios y gestores de información una nueva tendencia del trabajo profesional en el ámbito de la información digital, abordando en ella un tema de especial interés en un contexto como es la metainformación.

Para ejemplificar este trabajo, el dominio informativo elegido es el Web de LICAD porque, por un lado, es el ámbito de información electrónica en cuyo desarrollo la autora está implicada por pertenecer al equipo técnico del Departamento de Documentación de la Universidad Nacional de Mar del Plata, y además, porque resulta un dominio informativo concreto y controlable para los

objetivos de esta investigación, en la que se trata de esclarecer qué son los metadatos y difundir su utilidad.

2.3. Objetivos de la investigación

Objetivo general:

Examinar de manera detallada el estado de la cuestión de los metadatos, a partir de la revisión de las principales fuentes en la materia, así como ejemplificar, a través de una experiencia específica de aplicación, su utilidad para la recuperación de información.

Objetivos específicos:

- ✓ Determinar las ventajas de describir y recuperar la información electrónica en la World Wide Web, mediante la aplicación de metainformación.
- ✓ Describir los metadatos, sus características, funciones, y modos de aplicación.
- ✓ Caracterizar los programas disponibles para su implementación en el ámbito de la documentación.
- ✓ Describir la presentación de un caso de aplicación de metadatos en un sitio web educativo.

2.4. Formulación del problema

Ante el aumento exponencial de recursos Web, es preciso tomar conciencia —no sólo desde una visión informática, sino también bibliotecológica—, de las nuevas tendencias, tecnologías y aplicaciones alternativas a la imposibilidad de procesamiento convencional de la información electrónica.

Este problema generalizado en el ámbito de la recuperación de información de Internet, se puede concretar más en el marco de las bibliotecas digitales y de cualquier servicio de información electrónica, y en nuestro caso, circunscribimos la cuestión a definir qué son los metadatos y cuál es el estado de la cuestión, realizando una revisión bibliográfica y elaborando un trabajo monográfico acerca de las diferentes facetas necesarias para su definición teórica. Agregaremos, a modo de ejemplo, una propuesta de aplicación para la mejora del sistema de descripción-recuperación de contenidos en el sitio Web de LICAD, en el que garantizar el acceso coherente a la información es una obligación por parte de nosotros y un derecho de los alumnos.

Problema: ¿Qué son los metadatos, sus características, funciones, modos de aplicación y de qué software disponemos para su implementación, en el ámbito de la bibliotecología y documentación?

2.5. Antecedentes, estado actual de la cuestión y marco teórico

La red conocida como Internet tiene más de 30 años de existencia, pero fue a principios de los años 90 cuando se establecieron las bases de lo que hoy conocemos como *World Wide Web*, y el momento a partir del cual se desarrolla exponencialmente como medio informativo. Se fundamentó en dos pilares: un protocolo de transferencia de hipertexto (HTTP, *HiperText Transfer Protocol*), y un sencillo lenguaje de marcado (*HTML, HyperText Markup Language*) interpretable por intuitivas aplicaciones de software o navegadores (*browsers*), que han hecho de la Web un entorno documental universal. El crecimiento que ha experimentado la Web en la última década constituye la base del problema que pretenden solucionar los metadatos y por ende, el fundamento del problema de esta investigación.

Cuando Tim Berners-Lee estableció el *HyperText Markup Language* (HTML) tenía como principal objetivo garantizar el funcionamiento de la red desde un punto de vista informático y con este lenguaje de marcado logró, además de un soporte *multi-plataforma*, un modo intuitivo de hacer accesible y manejable la información. El modelo hipertextual simple era más que suficiente para organizar la información interna del CERN (*Centre Europeen de la Recherche Nucleaire-European Council for Nuclear Research*), lugar en el que surge originalmente la

WWW, pero cuando éste fue llevado a escala mundial o global comenzaron a detectarse rápidamente sus limitaciones².

En un primer momento se consideró posible la recuperación de la información a través de *directorios* que, contruidos manualmente, organizaran la información en función de su naturaleza temática. Los casos paradigmáticos de este tipo de servicios de búsqueda de información son, por diversos motivos³:

-Yahoo! <<http://www.yahoo.com>> en el que puede encontrarse una estructura jerárquicamente ordenada de 2.500.000 enlaces⁴ aproximadamente.

-OPD (*Open Directory Project*) <<http://dmoz.org>>, que registra en torno a 4.000.000 de recursos enlazados⁵.

Ciertamente, estas cifras pueden parecer bastante grandes; sin embargo, sólo corresponden a una ínfima parte de la totalidad de la Red, por lo que este tipo de soluciones, si bien útiles, son al mismo tiempo insuficientes. Una forma alternativa de subsanar la limitación de la lentitud de la selección y procesamiento

2 Berners-Lee, Tim. Tejiendo la red: el inventor del World Wide Web nos descubre su origen. Madrid: Siglo XXI, 2000.

3 En el caso de Yahoo! el motivo fundamental para considerarlo como paradigma de este tipo de servicios es fundamentalmente su antigüedad (1994), que le otorga la capacidad de establecer un modelo. En el caso de OPD, su cobertura y la forma cooperativa y abierta de su realización, desde que Tim Bray propusiera el modelo a finales de 1998.

4 Según los últimos datos recogidos en *Search Engine Watch* (diciembre de 2004) <http://www.searchenginewatch.com/reports/> (2.600.000 OPD / 1.800.000 Yahoo!)

ODP and Yahoo Size Projection Charts (consultado diciembre de 2004 / creado 1diciembre 1999 / actualizado 10 enero 2004) <http://www.geniac.net/odp/> (4.000.000 OPD / 2.500.000 Yahoo!)

5 *Ibid.*

manual, en que se fundan los índices de búsqueda se encuentran en el procesamiento (indización) automático.

En esta línea se desarrollaron los motores o robots de búsqueda (*SearchEngine*), aplicaciones que construyen sus bases de datos a través de sistemas totalmente automatizados, conocidas con diversas denominaciones (crawlers, robots, agentes, rastreadores, engines, worms, ants, etc.), cuya principal finalidad consiste en recorrer la Web y extraer de ella la información. Esta tarea la realizan a través de cuatro elementos básicos:

-El robot, formado por un localizador y recolector. El localizador se encarga de revisar las direcciones de servidor almacenadas en su base de datos en función de recuperar y almacenar nuevos enlaces en la base de datos del recolector, cuya función será evaluar esas direcciones en términos de su preexistencia y en tal caso de posibles modificaciones, por medio del análisis del tamaño y fecha del recurso. Si el recurso es novedoso o ha sufrido cambios se los envía al analizador-indexador.

-El sistema de indexación automática o analizador-indexador es el programa encargado de extraer palabras del contenido textual de acuerdo con un conjunto de filtros. Los filtros son los que deciden si se debe indexar toda la página o sólo las primeras palabras, si indexará todas las etiquetas de un documento HTML o sólo el texto contenido en las etiquetas META, TITLE, Hn. Estos datos se

almacenan en un fichero inverso que será gestionado por una base de datos sobre la que buscarán los usuarios.

-El motor de recuperación: es importante considerar este ítem desde dos puntos de vista, el que tiene que ver con el hardware, mayormente en el aspecto de velocidad y capacidad de procesamiento, y el que se relaciona con el software, en cuanto a las capacidades de recuperación de información, vinculado con la estructura de las bases de datos, su forma de almacenamiento, lenguajes de recuperación admitidos (utilización de palabras simples, referencias compuestas, operadores booleanos, operadores relacionales, operadores de adyacencia, búsqueda cualificada o por campos, posibilidad de truncamiento, distinción entre mayúsculas y minúsculas, fichero de palabras vacías, combinación de búsquedas, sistemas de retroalimentación de relevancias, búsquedas por peso).

-La interfaz de usuario: es el subsistema que permite que un usuario pueda realizar una consulta y visualizar sus resultados. En general, los motores de búsqueda presentan un formulario simple, con un cuadro de texto en el que el usuario puede insertar su fórmula de búsqueda, con algunas características distintivas de delimitación de idioma, tipo de documento, fecha de publicación, según el perfil del buscador, y un botón que posibilita la ejecución de la consulta. Presentan, así, los resultados de acuerdo con un algoritmo de relevancia, aspecto que tiene especial importancia puesto que el

orden en que son presentados es determinante de la satisfacción de la demanda del usuario.⁶

Abundan ejemplos en esta línea: Altavista, Northen Light, Google, Excite, InfoSeek, HotBot, AllTheWeb, Lycos, y muchos más.

Aún así, las bases de datos que se conforman no abarcan la totalidad de la red -no es lo mismo usar uno u otro buscador- y jerarquizan los resultados de acuerdo con criterios que muchas veces responden más a los lugares de aparición de los términos buscados (como en títulos o encabezados del documento), o, incluso, a fines comerciales, que a criterios de relevancia de información.⁷

Ante esta situación, en diferentes ámbitos académicos se comenzó a estudiar la posibilidad de desarrollar y aplicar algún tipo de norma que permitiera la descripción, y posterior recuperación, de los contenidos informativos. En esta línea encontramos dos grandes temas: la estructuración del documento completo (XML) y, por otro lado, su descripción catalográfica, en un sentido amplio. La primera, se encuentra dominada por el estándar de descripción XML mientras que en la segunda hallamos los metadatos. Si bien en torno al XML se están realizando muchos estudios, nos centraremos en la segunda de las corrientes y profundizaremos en el análisis de los METADATOS.

6 Hernández Pérez, Antonio. La búsqueda y recuperación de información en internet. EN: La sociedad de la información: política, tecnología e industria de los contenidos. Mercedes Caridad Sebastián (Coord.). Madrid: Fundación Ramón Areces, 1999. p. 213-238.

7 Siri, Laura. Internet: búsquedas y buscadores. Buenos Aires: Norma, 2000.

Aunque los términos METADATA y METADATOS no significan exactamente lo mismo para todos los colectivos profesionales (informáticos, bibliotecarios, gestores de BD, etc.) en el entorno de las bibliotecas digitales se usan cada vez más para referirse a la disponibilidad de datos sobre los recursos de información. Los términos hacen referencia a cualquier elemento utilizado para identificar, describir y localizar los recursos electrónicos en un entorno de red. Los metadatos son, en otras palabras, datos acerca de datos. En el mundo bibliotecario este concepto es bastante familiar ya que los registros bibliográficos son esencialmente metadatos; proporcionan información descriptiva y analítica sobre un objeto de información. Los catalogadores los han empleado como método descriptivo desde hace décadas, especialmente como registros MARC⁸.

Las áreas que definen el mundo de la investigación en metadatos abarcan su definición teórica, la generación de estándares internacionales, la creación de metadatos para nuevos recursos electrónicos, el desarrollo de nuevos tipos de metadatos, la evaluación de su efectividad, la elaboración de lenguajes y herramientas de aplicación y los métodos de gestión e *interoperabilidad* entre diferentes formatos.

No se pretende aquí abordar toda la complejidad que alberga esta problemática sino presentar en esta investigación el resultado de la revisión bibliográfica, tanto en fuentes impresas cuanto electrónicas, acerca del concepto

8 Gilliland-Swetland, Anne J. Setting the Stage. EN: Introduction to Metadata : Pathways to Digital Information. Murtha Baca. Los Ángeles: Getty Research Institute, mayo 2000. [Versión HTML] <http://www.getty.edu/research/institute/standards/intrometadata/pdf/swetland.pdf>

de metadatos, su construcción, desarrollo, aplicaciones y prospectiva de utilización en el contexto de la información electrónica en Internet. Un ejemplo que permita la utilización de un esquema básico de metainformación es su aplicación al sitio web de LICAD, ya que facilita una recuperación de información más precisa y coherente. Se trata pues, de recortar como objeto de estudio —a través de una metodología teórico-práctica— la discusión científica que, desde hace varios años gira en torno a la recuperación de información digital, instando a la utilización de metadatos así como a la investigación en la materia.

2.6 Diseño metodológico

Tipo de estudio: exploratorio-descriptivo

Diseño de investigación: transeccional

Instrumentos de recolección: fichaje de documentos recuperados en bibliotecas y bases de datos on-line.

Universo de estudio: artículos de publicaciones periódicas, libros y documentos electrónicos, publicados entre los años 1998 y 2005.

Variable: Metadatos

Definición conceptual:

-Metadatos: información acerca de la información; son datos sobre las características de un recurso de información, describen sus relaciones, y permiten su descubrimiento (recuperación) y uso efectivo. (Burnett, 1997)

Definición operacional:

-Estructura: sintaxis, semántica.

-Características: extensibilidad, repetibilidad, opcionalidad, modificabilidad.

-Tipos y funciones: administrativos, descriptivos, de conservación, técnicos, legales, de uso o acceso; y para cada uno de ellos: procedencia (internos - externos), métodos de creación (automáticos generados por ordenador - manuales creados por individuos), carácter (creados por individuos que no son especialistas temáticos ni en información - creados por expertos), estatus (estáticos - dinámicos-de larga duración - de corta duración), estructura (estructurados - no estructurados), semántica (controlados - no controlados), nivel (de colección - individuales).

-Modos de aplicación:

- Embebido/Encajado (*Embedded*): aparecen contenidos en el objeto de información digital (EJ: TEI / sintaxis << meta >> HTML: DC)
- Junto con/A lo largo del (*Along-with*): son externos al recurso pero se proporcionarán con él mediante un mecanismo de transferencia de la respuesta a una recuperación de información.
- Oficina de servicios (*Service bureau*) [departamento de servicios: cuerpo que proporciona información computada al público y puede procesar datos]: se almacenan separados del recurso de información, por ejemplo, en bases de datos y permiten su acceso. (EJ. MARC/UNIMARC).
- Envuelto (*Wrapped*): recurso de información encapsulado en los metadatos que lo describen. EAD / URI / DOI⁹.
- Métodos de implementación: manual, automático
- Software de implementación: denominación; institución proveedora; dirección de descarga; versión; modo de distribución (freeware, shareware, de pago); generador de metaetiquetas-generador de metadatos; tipo (plantillas, software cliente, SIG-DLO [Sistema Integrado de Gestión de Objetos Documentales]); interfaz (amigabilidad, sencillez); disponibilidad de ayuda (manejo de la aplicación, descripción de esquemas utilizados, etiquetas que

9 Lupovici, Catherine. Identification des ressources sur Internet et métadonnées: diversité des standards. EN Documentaliste – Sciences de l'information, 1999, vol. 36, N°6, p. 221-325.

soporta); forma de visualización del DLO que se está describiendo (navegador propio, navegador integrado, sin visualización); permiso de visualización del código fuente a través de un editor de textos; flexibilidad (permiso de repetir etiquetas, permiso de agregación de etiquetas ad hoc, permiso para generar nuevos formatos); formatos de codificación de metadatos (los distintos esquemas); integración al sistema de almacenamiento, búsqueda y recuperación de metainformación; control de autoridades (autores y materias); validación de los elementos obligatorios del modelo elegido; inclusión de aspectos de sintaxis (elementos, subelementos, calificadores).

2.7 Administración del proyecto

Cronograma de actividades:

1. Recopilación bibliográfica: publicaciones periódicas, monografías, manuales, páginas web. (3 meses)
2. Definición marco teórico. (3 meses)
3. Elaboración de un ejemplo que permita la aplicación de un esquema básico de metadatos al sitio web de la Licenciatura en Documentación a Distancia. Departamento de Documentación. Facultad de Humanidades. UNMDP. (2 meses)
4. Análisis de los resultados. (2 meses)
5. Elaboración de conclusiones y propuestas. (2 meses)

Recursos disponibles

PC con conexión a Internet, publicaciones periódicas especializadas, manuales de ciencias de la documentación e informática, software específico.

Duración

1 año

Aporte e impacto de los resultados

Como investigación básica: se realizará una revisión exhaustiva de los aspectos teóricos concernientes a los METADATOS dentro del ámbito de las ciencias de la información.

3. Los metadatos: definición, tipos y funciones

"Información acerca de la información" es la definición por excelencia que encontramos en todo documento acerca de metadatos, pero consideramos que con esto solo no basta para indicar el alcance del término.

Entendemos, y ampliando la anterior definición, que los metadatos están constituidos por aquellos elementos que permiten, de forma estructurada, describir las características de un objeto de información dado, para darle significado, ubicarlo dentro de un contexto y facilitar su organización, legibles por computadora.

A partir de este concepto podemos advertir una amplia correspondencia en el mundo de la información digital con las tareas de catalogación y clasificación, y su consiguiente normalización, desempeñadas tradicionalmente por los profesionales bibliotecarios en la organización de los documentos impresos, y es en este punto en el que se hace imprescindible la vinculación con los grupos que se ocupan del desarrollo de estándares internacionales de descripción de objetos digitales.

La estructura de los archivos digitales posibilitan que a cada objeto digital se le pueda asignar elementos de descripción según su formato: si se trata de un fichero generado a partir de una aplicación propietaria, el mismo software provee una herramienta de edición de las propiedades del documento: *Autor, Título,*

Palabras clave, Resumen, Fecha...; en el caso de los documentos generados a partir de lenguajes de marcado, se determinan etiquetas que permiten una descripción general del objeto digital (en el HEAD de un documento HTML mediante etiquetas META) o descripción de cada una de sus partes constitutivas.

La asignación de metadatos a un objeto o recurso de información digital se realiza a partir de la declaración de una propiedad, su valor y la referencia a un perfil normativo en el que se define cada propiedad y sus valores admitidos, y la sintaxis de una sentencia.

- Recursos: Todos los objetos descritos por expresiones RDF se denominan recursos. Un recurso puede ser una página Web completa, tal como el documento HTML <http://www.w3.org/Overview.html>, por ejemplo. Un recurso puede ser una parte de una página Web (un elemento HTML o XML específico dentro del documento fuente). Puede ser también una colección completa de páginas; es el caso de un sitio Web completo. También un objeto que no sea directamente accesible vía Web, p. ej. un libro impreso. Los recursos se designan siempre por URIs más identificadores de anclas opcionales (ver [URI]). Cualquier cosa puede tener un URI; la extensibilidad de URIs