



**Centro de Investigación y de Estudios
Avanzados del Instituto Politécnico
Nacional**

**Departamento de Ingeniería Eléctrica
Sección de Computación**

Meta-X: Sistema para creación de Metadatos.

Tesis que presenta

Rubén Isaí Rivera Rodríguez

para obtener el Grado de

Maestro en Ciencias

en la Especialidad de

Computación

Director de la Tesis

Sergio Víctor Chapa Vergara

México D.F.

Agosto 2005

Dedicatoria

Dedico este trabajo a mi madre Rebeca Rodríguez Ortíz
y a mi madrina Elia Rodríguez Ortíz.
Las amo, gracias por todo.

Agradecimientos

Primero agradezco a Dios, porque sin Él no estaría aquí, gracias por la paz y fortaleza que me hace sentir con su presencia, y que me ha ayudado a lograr mis objetivos.

Agradezco a mi madre Rebeca Rodríguez Ortíz y a mi madrina Elia Rodríguez Ortíz por todo su amor. Por sus consejos, por su guía, porque han sido un ejemplo para mí de dedicación, de fortaleza y de respeto. Por la confianza, la fé que siempre me han tenido, porque sin ustedes no podría haber llegado a donde estoy. Y porque son las personas que más amo y admiro en el mundo. Gracias .

A mi hermana Saraí y a mi sobrino Irad, porque han sido una motivación para seguir adelante.

A mi tía Miriam, y a sus hijos Alejandro, Gerardo y Sofía, por su cariño y por su fé en mí, por siempre hacerme reír.

Al Dr. Sergio V. Chapa Vergara (Doc) por todo el apoyo que me ha brindado, por su amistad, sus consejos y por sus oportunos regaños, su confianza en mí y por su infinita paciencia, que hicieron posible que terminara este trabajo.

A la Dra. Ana María Martínez Enríquez y al Dr. Felipe Gómez Castañeda por su tiempo para revisar esta tesis y por sus comentarios e ideas para concretar este trabajo.

A Sofy, porque ha sido una madre para mí en el Cinvestav, por todo su amor, su amistad, su eterna sonrisa que siempre me levantaba el ánimo, por su apoyo en todos los aspectos, tanto académicos como personales, porque siempre tuvo tiempo para escucharme.

A Flor y a Feli, por su buen humor y por su ayuda para completar todos los trámites que nos solicitaban.

A mis amigos de toda la vida, Camilo, Omar (Patiño), Luis (Panzón), Nancy, Luz, Ethel, Javier (Gummy), Bello, Omar Garnica, Juan Canales, por su franqueza y por su confianza, porque siempre estuvieron ahí cuando me sentía agobiado y siempre hallamos juntos la forma de salir de los problemas y en algunas ocasiones también la manera de meternos en ellos. Gracias, porque con ellos he aprendido a no tomarme la vida tan en serio.

A Gaby Rivera porque siempre me ha apoyado y ha tenido las palabras indicadas para no dejarme caer, por todo el cariño que me ha dado.

A los amigos que he conocido en el CINVESTAV, Nancy Soto, Lalo López, Betsa, Ulises, Noé, Nancy, Marisol y Mayté. Porque vivimos muchas experiencias, algunas un poco más difíciles que otras, pero que al fin y al cabo superamos y que nos dieron la oportunidad de conocernos.

A Toño, por su amistad y porque siempre se dió tiempo para compartir sus conocimientos y explicarme lo que en ocasiones se me complicaba entender.

A Lore por ayudarme a terminar esta tesis, por sus comentarios a este trabajo, por sus bromas y sus infinitas muestras de amistad y palabras de ánimo que me ayudaron a hacer las cosas de manera más serena y no sucumbir a la presión.

A Adriana porque su amistad, su paciencia, su tiempo y sus consejos me ayudaron a superar algunos de los momentos más difíciles que he vivido.

Al CONACyT por su apoyo económico sin el cuál no habría sido posible concluir este trabajo.

Al CINVESTAV porque me ha dado la oportunidad de alcanzar una de mis metas académicas. Y porque aquí he conocido personas excelentes, y he vivido experiencias que me son muy apreciadas y que me han hecho crecer en todos los aspectos de mi vida.

Resumen

Actualmente los sistemas de información son utilizados hoy día en una gran diversidad de campos, como son la medicina, la geografía, biología, etc. Un punto importante para cualquier sistema de información es mantener su semántica, esta nos permite representar con mayor precisión el universo del discurso y, para capturar esta semántica se utilizan los metadatos, comúnmente definidos como : información acerca de la información.

Como se mencionó antes, en los sistemas de información se presenta el problema de falta de semántica. Surge la necesidad de tener alguna forma de apoyar el diseño de los sistemas de información desde las primeras etapas del desarrollo, a esto es lo que llamamos metadiseño.

Nuestro sistema, llamado Meta-X, plantea una arquitectura cuya base son un conjunto de bases de datos, a los que denominamos repositorios que, almacenarán requerimientos, metadatos y reglas semánticas. En el presente trabajo, nos enfocamos a la parte de la creación de metadatos, para lo cual desarrollamos una herramienta que nos permite capturar metadatos, apoyándonos en el Estándar de Contenido para Metadatos, y salvarlos en un formato XML.

Nuestra herramienta fue desarrollada sobre la plataforma Mac OS X, utilizando C-Objetivo como lenguaje de programación. La captura de metadatos es realizada a través de interfaces visuales programadas utilizando el framework Cocoa.

Palabras Clave: Bases de datos, metadatos, creación de metadatos, metadiseño, Estándar de Contenido para Metadatos

Abstract

Nowadays the information systems are used in different areas as medicine, geography, biology, etc. A main issue for any information system is keeping its semantic. Semantic allow us to represent more accurately the universe of discourse. In order to capture this semantic we use metadata, which are commonly defined as: information about information.

As mentioned earlier, the information systems exhibits the problem of lack of semantics. Therefore, arise the necessity of having a way to support the design of the information systems from the beginning of the development, this is what we called metadesign.

Our system, called Meta-X, propose an architecture based on a set of databases, which are named repositories. These repositories have the function of storing metadata, requirements and semantic rules. At this work, we focus on metadata creation, developing a tool which allow us to capture metadata based on Content Standard Digital Geographic Metadata (CSDGM). The metadata are stored as XML files.

Meta-X was developed with Objective-C under Mac OS X. The metadata capture is done through visual interfaces with the aid of the Cocoa Framework.

Keywords: Metadata, databases, metadata creation, metadesign, CSDGM.

Índice general

Agradecimientos	III
Resumen	V
Abstract	VII
Introducción	1
1. Ingeniería de Metadatos	5
1.1. ¿Qué son los metadatos?	5
1.2. ¿Qué hacen los metadatos?	6
1.2.1. Descubrimientos de Recursos	6
1.2.2. Organización de Recursos Electrónicos	7
1.2.3. Interoperabilidad	7
1.2.4. Identificación Digital	7
1.2.5. Preservación	8
1.3. Estructuración de Metadatos	8
1.3.1. Esquemas de Metadatos y Conjuntos de Elementos	9
1.3.2. Metadatos para Conjuntos de Datos	12
1.3.3. Extensiones y Perfiles	13
1.4. Creación de Metadatos	13
1.4.1. Herramientas para Creación de Metadatos	14
1.5. Control de Calidad de los Metadatos	16

1.6.	Impacto de los metadatos en nuevas tecnologías de bases de datos	17
1.6.1.	Sistemas de Información Geográfica (SIG)	17
1.6.2.	Centros de Distribución de Datos (Clearinghouses)	18
2.	Estándar de Contenido para Metadatos:	
	Geoespaciales Digitales	19
2.1.	Desarrollo del Estándar	19
2.2.	Objetivo	20
2.3.	Alcance del Estándar	21
2.4.	Organización del Estándar	21
2.5.	Secciones del Estándar	23
2.5.1.	Metadato	23
2.5.2.	Identificación	24
2.5.3.	Organización de los datos espaciales	28
2.5.4.	Referencia Espacial	30
2.5.5.	Referencia de Metadatos	34
2.6.	Secciones de Apoyo	35
2.6.1.	Cita	35
2.6.2.	Periodo	37
2.6.3.	Contacto	37
3.	Sistema Meta-X	41
3.1.	Arquitectura del Sistema	42
3.1.1.	Herramienta de creación de metadatos	42
3.1.2.	Servidor de Bases de Datos	42
3.2.	Diseño e Implementación de Meta-X	45
3.2.1.	¿Qué es Cocoa?	46
3.2.2.	Meta-X, aplicación multidocumentos	46
3.2.3.	Vista en Meta-X	51
3.2.4.	Controladores en Meta-X	53

3.2.5. Modelo en Meta-X	56
4. Caso de Estudio	61
4.1. Modelo conceptual	61
4.1.1. Entidad - Relación	61
4.1.2. Repositorio de Metadatos	63
4.2. Caso de estudio de Meta-X	69
4.2.1. INEGI	72
5. Conclusiones y Trabajo Futuro	81
5.1. Conclusiones	81
5.1.1. Aportaciones	82
5.2. Trabajo Futuro	82

Índice de figuras

2.1. Organización del Estándar para Contenido de Metadatos Geoespaciales Digitales	22
3.1. Arquitectura General del Sistema	43
3.2. Cocoa y sus componentes	46
3.3. Arquitectura basada en documentos	48
3.4. Paradigma MVC	49
3.5. Meta-X y su arquitectura multidocumento	50
3.6. Vista de la interfaz gráfica de Meta-X	52
4.1. Notación “Crow’s Foot” del Modelo ER	62
4.2. Diagrama ER General del Repositorio de Metadatos	65
4.3. Submodelo ER de la sección de identificación	67
4.4. Submodelo ER de la sección de Información de Referencia del Metadato	68
4.5. Submodelo ER de la sección de Información de Cita	70
4.6. Submodelo ER de la sección de Información de Contacto	71
4.7. Vista de la pantalla de captura de información de cita	73
4.8. Vista de la pantalla de captura de la descripción de metadatos	75
4.9. Vista de la pantalla de captura de la descripción de metadatos	76
4.10. Vista de la pantalla de captura de información de Dominio Espacial	77
4.11. Vista de la pantalla de captura de información de contacto	79

Índice de cuadros

1.1. Herramientas para Creación de Metadatos	14
3.1. Variables y métodos de acceso de la clase CiteInfo	57
3.2. Variables y métodos de acceso de CntInfo	57
3.3. Variables y métodos de acceso de TimeInfo	58
3.4. Variables y métodos de acceso de la clase IdInfo	58
3.5. Variables y métodos de acceso de la clase MetaInfo	59

Introducción

Antecedentes

Uno de los problemas que se presentan cuando desarrollamos sistemas de información es la falta de semántica. Esta nos permite representar con mayor precisión el universo del discurso, y se encuentra en los metadatos. A los metadatos los podemos definir como información acerca de la información o datos acerca de los datos. Algunos usos que se le dan a los metadatos son: proporcionar información para catálogos de datos, proveer la información necesaria para interpretar datos transferidos por otra organización.

Descripción de la Herramienta

Nuestra herramienta llamada Meta-X, fue desarrollada en el lenguaje C-Objetivo, para la plataforma Mac OS X. Apoya el desarrollo de sistemas de BDs en un nivel de metadiseño, proponiendo una arquitectura cliente servidor, cuya base es un conjunto de repositorios los cuales están destinados a almacenar metadatos, requerimientos y reglas semánticas. En esta tesis nos enfocamos en la herramienta para creación de metadatos por medio del llenado de plantillas o formas prediseñadas; el contenido de estas plantillas toma como base al Estándar de Contenido para Metadatos Geoespaciales, los archivos creados son almacenados en formato XML.

Alcances

Meta-X facilita la publicación de los metadatos en internet al almacenarlos en formato XML. Además de que al estar basada en un estándar, permite que los metadatos creados, puedan ser entendidos y manejados por una comunidad más grande de usuarios. En lo que respecta a los repositorios, estos nos permitirán interactuar con otras herramientas para trabajo con bases de datos, como son EVEX, una herramienta para modelado de sistemas mediante diagramas Entidad-Vínculo-Extendido, y LIDA que es un lenguaje iconográfico para el desarrollo de aplicaciones, ambos sistemas fueron desarrollados en el CINVESTAV. De forma tal que, en conjunto tenemos un sistema para desarrollo de software que cubre las etapas de especificación del software, diseño e implementación, dentro del proceso de desarrollo de sistemas de información.

Organización de la tesis

En las secciones anteriores hemos descrito el contexto de nuestro trabajo. Los capítulos restantes están organizados de la siguiente manera:

Capítulo 1. Introducimos la definición de metadatos, algunos de los esquemas que existen para trabajar con metadatos en diversos ámbitos y su importancia para nuevas tecnologías de la información como son los sistemas de información geográfica (SIG).

Capítulo 2. Explicamos la arquitectura general del Estándar de Contenido para Metadatos Geospaciales, explicando de forma más detallada las secciones del mismo que son de mayor interés para el presente trabajo.

Capítulo 3. Explicamos la arquitectura de nuestra herramienta, la metodología utilizada para la creación de metadatos.

Capítulo 4. Proponemos un modelo conceptual del repositorio de metadatos siguiendo

la metodología entidad-relación. Después, utilizamos nuestra herramienta para crear los metadatos de una fotografía de una región geográfica; esta información fue proporcionada por el INEGI.

Capítulo 5. Presentamos las conclusiones de nuestro trabajo y proponemos una guía para trabajo futuro.

Capítulo 1

Ingeniería de Metadatos

En este capítulo exponemos, en la primera sección, la definición de metadatos, su importancia y la manera en que se clasifican; en la segunda sección, algunos de los esquemas de metadatos existentes para distintas disciplinas, explicando brevemente sus características; en la parte final, describimos lo que son los Sistemas de Información Geográfica (SIG) y los Centros de Distribución de Metadatos (Clearinghouses).

1.1. ¿Qué son los metadatos?

El término metadato es usado de forma distinta en diferentes comunidades, por ejemplo, en el ambiente de las bases de datos (BD), los metadatos son toda la información almacenada en el diccionario de la BD [5], mientras que otros, utilizan el término únicamente para registros que describen recursos electrónicos. En el entorno bibliotecario, los metadatos son comunmente usados para la descripción de cualquier esquema de recursos, aplicado a cualquier tipo de objeto, digital o no.

Así, encontramos que los metadatos se definen comúnmente, de manera simple como, datos acerca de los datos o, de forma equivalente, información acerca de la información [2]. Otra definición un poco más amplia es:

Los metadatos son información estructurada que describe, explica, localiza, o de alguna manera facilita la obtención, el uso o la administración de un recurso de información [13].

Los metadatos en el contexto de sistemas de información espacial digital se definen como, la información de fondo, la cual describe el contenido, la calidad, la condición, y otras características apropiadas de los datos [8]. Nuestro trabajo se dirige a estos sistemas de información geoespaciales.

A los metadatos podemos clasificarlos en tres clases principales [7]:

- **Metadatos Descriptivos.** Esta clase de metadatos ayuda a los usuarios en la búsqueda de un recurso, a distinguir un recurso de otro y a entender el asunto o contenido del recurso. En otras palabras, describen un recurso para propósitos tales como descubrimiento e identificación. En esta clase se incluyen elementos como son: título, autor, palabras clave y resumen.
- **Metadatos Estructurales.** Indican como agrupar objetos compuestos, como ejemplo tenemos, la relación entre los artículos, el número y el volumen de una publicación serial, o la manera en la cual se ordenan las páginas de un libro formar secciones, capítulos, etc.
- **Metadatos Administrativos.** Esta clase de metadatos provee información que ayuda en la administración de un recurso, por ejemplo: la fecha de creación e identidad del creador, el tipo de archivo del que se trata, quienes pueden acceder a la información, etc.

1.2. ¿Qué hacen los metadatos?

Una razón importante para crear metadatos descriptivos es facilitar el descubrimiento de información relevante. Y además ayudar a organizar los recursos electrónicos, facilitar la interoperabilidad y la integración de recursos, proveer identificación digital y apoyar la preservación.

1.2.1. Descubrimientos de Recursos

Las funciones de los metadatos en el descubrimiento de recursos son:

- Permitir la búsqueda de recursos a partir de criterios de importancia.
- Identificar los recursos.
- Reunir recursos similares y distinguir los que no lo son.
- Proporcionar información de localización.

1.2.2. Organización de Recursos Electrónicos

Los recursos basados en Web van creciendo de manera exponencial, y con ellos se ha incrementado también la utilidad de los portales para organizar los enlaces a los recursos basados en la audiencia o en el tema. Lo anterior se logra construyendo páginas dinámicas que accesan a los metadatos almacenados en una base de datos. Existen varias herramientas para extraer esta información y convertirla a un formato adecuado para aplicaciones Web.

1.2.3. Interoperabilidad

La interoperabilidad es la habilidad de múltiples sistemas con diferente hardware, distintas plataformas de software, estructuras de datos e interfaces, para intercambiar datos con una mínima pérdida de contenido y de funcionalidad.

Usando esquemas de metadatos, protocolos de transferencia compartidos, y *crosswalks* [4] (Una tabla que mapea las relaciones y equivalencias entre dos o más formatos de metadatos), se puede realizar la búsqueda de estos recursos a través de la red de forma transparente.

1.2.4. Identificación Digital

La mayoría de los metadatos incluyen elementos tales como números estándar para identificar de manera única el trabajo u objeto al cual se refieren. La localidad de un objeto digital puede ser dada mediante un nombre de archivo, una URL, o algún identificador

persistente como PURL (Persistent URL Persistente) o DOI (Identificador de Objetos Digitales).

1.2.5. Preservación

Los metadatos son clave para asegurar que los recursos sobrevivirán y continuarán siendo accesibles en el futuro. La preservación requiere elementos especiales para darle seguimiento a un objeto digital (¿de dónde vino? y ¿cómo ha cambiado con el tiempo?), detallar sus características físicas, y documentar su comportamiento para poder emularlo con nuevas tecnologías.

Muchas organizaciones han hecho esfuerzos por definir un esquema de metadatos para la preservación digital, entre ellas, la Biblioteca Nacional de Australia, el Proyecto Británico Cedars, y conjuntamente el Grupo de Trabajo del Centro de Bibliotecas de Cómputo en Línea y el Grupo de Bibliotecas de Investigación (RLG).

1.3. Estructuración de Metadatos

Los esquemas de metadatos son conjuntos de elementos de metadatos diseñados para un propósito específico, tal como describir a un recurso de información específico. La definición o significado de los elementos en sí, es conocida como la semántica del esquema, mientras que, los valores que le asignamos a los elementos de los metadatos son el contenido. Los esquemas de metadatos generalmente especifican los nombres de sus elementos y su semántica. Opcionalmente, pueden llegar a especificar reglas que determinen la manera en que el contenido debe ser formulado (por ejemplo, como identificar el título principal), reglas para su representación (por ejemplo, el uso de mayúsculas) y los valores que son permitidos (por ejemplo, términos que deben ser usados de un vocabulario específico). También puede haber reglas de sintaxis que establecen la manera en que los elementos y su contenido deben ser codificados; un esquema de metadatos sin reglas de sintaxis preestablecidas es llamado independiente de sintaxis. Los metadatos pueden ser codificados en cualquier sintaxis definida; muchos esquemas de metadatos actuales utilizan SGML (Standard Generalized Mark-up Language) o XML (Extensible Mark-up Language).

1.3.1. Esquemas de Metadatos y Conjuntos de Elementos

Muchos esquemas distintos de metadatos están siendo desarrollados en una variedad de ambientes de usuario y disciplinas. Algunos ejemplos son:

Dublin Core

El Conjunto de Elementos de Metadatos Dublin Core surgió de un taller organizado en 1995 por el Centro de Bibliotecas de Cómputo en Línea (OCLC por sus siglas en inglés) y el Centro Nacional de Aplicaciones de Supercómputo. Dado que el taller se realizó en Dublin, Ohio, el conjunto de elementos fue nombrado Dublin Core (Núcleo Dublin).

El objetivo original de Dublin Core fue definir un conjunto de elementos que pudieran ser utilizados para describir sitios Web por sus autores. Son 15 los elementos que actualmente conforman este núcleo: título, creador, asunto, descripción, editor, contribuyente, fecha, tipo, formato, identificador, origen, idioma, relación, cobertura y derechos. Todos ellos son opcionales y pueden repetirse y presentarse en cualquier orden.

La Iniciativa de Codificación de Texto (TEI)

La Iniciativa de Codificación de Texto (TEI por sus siglas en inglés) es un proyecto internacional para desarrollar pautas a seguir para el mercado de textos electrónicos, tales como, novelas, obras de teatro, poesía y principalmente para apoyar la investigación en humanidades.

Se asume que todos los textos codificados con TEI son versiones electrónicas de los textos impresos.

Estándar para la Codificación y Transmisión de Metadatos (METS)

El estándar para la codificación y transmisión de metadatos fue desarrollado para satisfacer la necesidad de una estructura de datos estándar para describir objetos complejos de bibliotecas digitales. METS es un esquema XML para crear instancias de documentos XML que expresan la estructura de objetos de bibliotecas digitales, los metadatos descriptivos y administrativos asociados, y los nombres y localidades de los archivos que

comprenden el objeto digital.

Un documento METS contiene siete secciones: encabezado, metadatos descriptivos, metadatos administrativos, sección de archivos, mapa estructural, enlaces estructurales, comportamiento.

Esquema de Descripción de Objetos de Metadatos (MODS)

El Esquema de Descripción de Objetos de Metadatos (MODS) es un esquema descriptivo de metadatos derivado de MARC 21 (un formato conciso para datos bibliográficos), y su propósito es soportar registros existentes de MARC 21 o poder crear registros de descripción nuevos. Incluye un subconjunto de campos de MARC 21 y usa un lenguaje basado en etiquetas en vez de los numéricos usados en MARC 21. Al igual que METS, MODS se expresa usando como lenguaje de esquema a XML.

Descripción Codificada de Archivos (EAD)

La Descripción Codificada de Archivos (EAD) fue desarrollada como una manera de marcar los datos contenidos en la guía de búsqueda, de tal forma que puedan ser buscados y mostrados en línea.

En archivos y colecciones especiales, la guía de consulta o búsqueda es una herramienta importante para la descripción de los recursos. La guía de búsqueda difiere de los registros de catálogos por ser mucho más larga, más narrativa y explicativa, y está estructurada jerárquicamente. Generalmente la guía de búsqueda describe la colección como un todo, indicando qué tipos de materiales contiene y su importancia. La guía describe las series en las cuales está organizada la colección y termina con una lista del contenido de las cajas y carpetas físicas que comprenden la colección.

EAD es particularmente popular en bibliotecas académicas, sociedades históricas y museos con grandes colecciones especiales. Muchas de estas colecciones poseen materiales únicos que no están disponibles en otro lugar.

Metadatos de Objetos para Aprendizaje (Learning Object Metadata) (LOM)

El estándar para Metadatos de Objetos para Aprendizaje (LOM) fue desarrollado por el Comité de Estándares para la Tecnología de Aprendizaje de la IEEE, con el fin de poder hacer uso y reutilización de tecnología de apoyo al aprendizaje, como el entrenamiento basado en computadora.

LOM define el conjunto mínimo de atributos para administrar, localizar y evaluar los objetos para aprendizaje. Los atributos se agrupan en 8 categorías: general, ciclo de vida, técnica, educacional, derechos, relación, anotaciones y clasificación.

E-Commerce - <indec> y ONIX

ONIX e <indec> son esquemas de metadatos desarrollados para apoyar el comercio electrónico. El framework <indec> (Interoperabilidad de Datos en Sistemas de E-Commerce) fue desarrollado en un esfuerzo internacional conjunto por el Programa Info 2000 de la Comisión Europea. Los colaboradores fueron dueños, editores y miembros de la industria de la grabación, quienes querían desarrollar un framework para un estándar de metadatos que apoyara el comercio en red, con propiedad intelectual.

ONIX es un esquema de metadatos basado en XML desarrollado por editores bajo el auspicio de grupos de la industria de comercio de libros tanto en Estados Unidos como en Europa. ONIX para libros contiene elementos para registrar una amplia gama de información tanto promocional como de evaluación.

Objetos Visuales - CDWA y VRA

Los metadatos usados para describir objetos visuales, como pinturas y esculturas tienen requerimientos especiales. La Fuerza de Tarea de Información de Arte (AITF), desarrolló un framework conceptual para describir y acceder información acerca de objetos e imágenes llamados Categorías para la Descripción de Trabajos de Arte (CDWA). Alrededor de 30 categorías fueron definidas, la mayoría de ellas también tienen subcategorías. Algunos ejemplos de datos relevantes son: orientación, dimensiones, condición, inscripciones, tratamiento para conservación e historia de la exhibición.

El núcleo VRA (VRA Core) se basa en CDWA y lo extiende para definir un solo conjunto de elementos de metadatos para describir la obra, ya sea que se trate de pintura, escultura, fotografía, o de una representación visual de ellas.

Metadatos para Multimedia MPEG

El Grupo de Expertos en Imágenes en Movimiento (MPEG) de ISO/IEC ha desarrollado una suite de estándares para la presentación codificada de audio y video digital. Uno de ellos es el MPEG-7, este describe los elementos, la estructura y las relaciones de los metadatos que son usados para describir objetos audiovisuales, incluyendo imágenes estáticas, gráficos, modelos 3D, audio, video o colecciones multimedia. Este es un estándar multiparte que se enfoca en: herramientas de descripción, que contienen descriptores que definen la sintaxis y la semántica de cada elemento del metadato, un lenguaje de definición de descripción que define la sintaxis de las herramientas de descripción, y herramientas del sistema que apoyan el almacenamiento y transmisión, sincronización de descripciones con el contenido, así como la administración y protección de la propiedad intelectual.

1.3.2. Metadatos para Conjuntos de Datos

Los esquemas de metadatos para conjuntos de datos permiten que datos originales en los campos de la ciencia y de las ciencias sociales sean compartidos de una forma que antes de la internet no habría sido posible. Uno de estos conjuntos de elementos que más desarrollo ha tenido es el Estándar de Contenido para Metadatos Geoespaciales Digitales (CSDGM) del Comité Federal de Datos Geográficos (FGDC), mejor conocido como FGDC-STD-001-1998.

Los conjuntos de datos geoespaciales incluyen datos topográficos y demográficos, SIG (Sistemas de Información Geográfica) y archivos de cartografía asistida por computadora. Son usados en una amplia variedad de áreas, incluyendo estudios de suelo y de uso de tierras, conteo de la biodiversidad, rastreo de la climatología y los cambios globales, e imágenes de satélite. El estándar de contenido FGDC es requerido para usarlo con los recursos creados y fundados por el gobierno de los Estados Unidos de América y está siendo

utilizado también por otros países.

1.3.3. Extensiones y Perfiles

La mayoría de los esquemas desarrollados hasta ahora, son modificados para poder ajustarse a situaciones del mundo real. Estas modificaciones se dividen en dos tipos: extensiones y perfiles.

Una extensión es la adición de elementos a un esquema para soportar la descripción de un tipo de recurso particular o satisfacer las necesidades de un grupo en particular. Las extensiones incrementan el número de elementos.

Los perfiles son subconjuntos de un esquema que son implementados por un grupo en particular. Los perfiles pueden restringir el número de elementos que serán usados, refinar definiciones de elementos para describir tipos de recursos específicos con más precisión y especificar los valores que un elemento puede tomar.

En la práctica, muchas aplicaciones usan ambas, perfiles y extensiones de esquemas de metadatos.

1.4. Creación de Metadatos

Muchos de los metadatos estructurales y administrativos básicos los provee el personal técnico que se encarga de la digitalización o creación de un objeto digital, o son creados a través de un proceso automatizado.

Para los metadatos descriptivos en cambio, es mejor que sea el personal originador del recurso quien provea la información, en especial cuando se trata de conjuntos de datos científicos, donde el originador es quien tiene mayor comprensión del tema y de los usos que se le pueden dar a la información. Sin embargo, a veces se prefiere que sea el personal técnico quien cree los metadatos descriptivos, ya que los autores o creadores de los datos no tienen el tiempo o las habilidades necesarias para hacerlo. Y después los resultados pueden ser revisados por los especialistas para confirmar que sean correctos y consistentes con los esquemas de sintaxis y guías que proporcionaron.

1.4.1. Herramientas para Creación de Metadatos

Existen varias herramientas de software comerciales que ayudan a la creación de metadatos. Estas comúnmente son desarrolladas para soportar un esquema específico de metadatos. Las podemos dividir como sigue:

- Plantillas o Formas. Estas permiten que el usuario ingrese los valores de los metadatos en campos predeterminados.
- Herramientas de Marcado. Se encargan de estructurar los valores y atributos de los metadatos en un lenguaje de esquema específico, por lo general XML.
- Herramientas de Extracción. Estas crean automáticamente los metadatos a partir de un análisis del recurso. Por lo general, dichos recursos se limitan a solo texto.
- Herramientas de Conversión. Traducen de un formato de metadatos a otro.

Algunas herramientas que apoyan la creación de metadatos y que hacen uso del Estándar de Contenido para Metadatos Digitales Geoespaciales (CSDGM) del FGDC se muestran en la tabla 1.4.1:

Cuadro 1.1: Herramientas para Creación de Metadatos

Nombre	Sistema Operativo	\$	Exporta	Requerimientos
ArcCatalog	WinNT	Sí	XML	ArcInfo 8.02
ArcView Metadata Collector	Windows, Unix	No	Texto, HTML	ArcView 3.x
BIC Metadata Tool	Web	No	Texto, SGML	cgi-script
CorpsMet95	Win 95	No	Texto	Standalone
DataLogr	Windows	Sí	Texto	Standalone

Data Dictionary (DataDict)	Windows	No	mptext	ArcInfo / Info AML
Dataset Catalog Database Sys	Windows	No	Otro	Standalone
Document AML	Windows, UNIX	No	mptext	Arc / Info AML
Fgdcmeta AML	Windows, UNIX	No	mptext	Arc / Info AML
GeoData MDB	Windows	No	Otro	Access
Geospatial Metadata Mgt Sys	Windows	No	mptext	Access
M3Cat	Web	Sí	Texto	IIS y Access u Oracle
Metadata 2 (MD2)	Windows	No	Otro	Access
Metadata Entry System	Web	No	mptext, SGML	cgi-script
MetaLite System	Windows	No	mptext, XML, SGML	Standalone
Metadata Management System	Windows	Sí	HTML	Standalone o DB
MetaMaker 2.10	Windows	No	mptext	Standalone

Spatial Meta-data Management System (SMMS)	Windows	Sí	mptext, XML, SGML, HTML	Standalone
Theme Browser	Windows	Sí	HTML	ArcView
MetaStar	Java-Web	Sí	HTML, SGML, XML, text	Standalone
Xt Metadata Editor	UNIX	No	XML, SGML, mptext	Standalone
Metavist 2005	Win 2000 y XP	No	XML	—

1.5. Control de Calidad de los Metadatos

La creación de metadatos, ya sea de forma automática o manual puede causar problemas de calidad. Los elementos obligatorios podrían faltar, o fallas en la sintaxis del esquema podría evitar que fueran procesados de forma correcta, quizá la terminología sea inconsistente causando que sea más difícil encontrar la información relevante.

En [7] se proponen 6 principios que deben cumplir los buenos metadatos:

- Deben ser apropiados para los materiales en la colección, usuarios de la colección, y para el uso que actual y futuro del objeto digital.
- Deben soportar la interoperabilidad.
- Usan estándares para controlar el vocabulario utilizado para reflejar el qué, cómo, cuándo y dónde del contenido.
- Incluyen una oración clara sobre las condiciones y términos de uso para el objeto digital.

- Soportan la administración a largo plazo de los objetos en la colección.
- Son objetos en sí mismos, por lo que deben poseer las características de estos: autoridad, autenticidad, que puedan ser archivados, persistencia e identificación única.

1.6. Impacto de los metadatos en nuevas tecnologías de bases de datos

1.6.1. Sistemas de Información Geográfica (SIG)

Los Sistemas de Información Geográfica (SIG) son una tecnología basada en computadora de propósito general para manejo de datos geográficos en forma digital. Está diseñada para capturar, almacenar, manipular, analizar y mostrar diversos conjuntos de datos espaciales o georeferenciados [16]. Contienen datos geométricos (coordenadas e información topológica) y datos de los atributos (información que describe las propiedades de objetos geométricos, como puntos, líneas y áreas)

Resulta más conveniente ver a un SIG como un proceso más que como una cosa. Estos sistemas describen los objetos del mundo real en términos de [14]:

- Posición. Posiciones geográficas
- Atributos que no están relacionados con la posición, como por ejemplo, el tipo de suelo, elevación o población.
- Relaciones Espaciales. La manera en que los objetos están ligados.
- Información temporal. La evolución de los datos tanto espaciales como no espaciales en el tiempo.

Como ejemplos de la utilidad que tiene la información proporcionada por los SIG tenemos, determinar dónde establecer una estación de bomberos, o en estudios del medio ambiente podrían ayudar a localizar los lugares donde la erosión del suelo es más severa,

estas son consideraciones clave usando esta información. En cada caso, qué es y dónde está debe tenerse en cuenta.

Un SIG contiene los siguientes subsistemas [6]:

- Entrada de datos.
- Almacenamiento, recuperación y representación de datos.
- Administración, transformación y análisis de datos.
- Reporte de datos y generación del producto.

1.6.2. Centros de Distribución de Datos (Clearinghouses)

Los metadatos son importantes en la creación de Clearinghouses para datos geoespaciales, donde los usuarios pueden encontrar la información que necesitan para sus propósitos.

Un centro de distribución de datos geoespaciales es una localidad, a la cual se accede a través de un sitio web para buscar conjuntos de datos espaciales. Una Clearinghouse puede contener conjuntos de datos específicos que pueden ser descargados o que pueden contener información acerca de conjuntos de datos que ayuden al usuario a tomar una decisión sobre si vale o no la pena obtener el conjunto de datos y como hacerlo.

Hasta ahora, hemos expuesto el significado de metadatos, y cómo este cambia de acuerdo al contexto dentro del cual se utiliza. También hemos explicado la importancia de los metadatos. A partir de ahora, entraremos al uso de los metadatos para sistemas de información geoespaciales, el esquema de metadatos en el que nos enfocaremos es el FGDC.

Capítulo 2

Estándar de Contenido para Metadatos: Geoespaciales Digitales

En el capítulo anterior, dimos una explicación sobre lo que son los metadatos, sus aplicaciones, además de una breve descripción de varios esquemas que existen para trabajar con ellos en distintas áreas. Entre los estándares existentes, uno que está siendo muy utilizado es el Estándar de Contenido para Metadatos Geoespaciales Digitales (Content Standard Digital Geospatial Metadata, CSDGM) [3] desarrollado por el Comité Federal de Datos Geográficos (Federal Geographic Data Committee, FGDC).

En el presente capítulo, profundizamos en el tema, exponiendo como ha sido el desarrollo de este estándar, sus objetivos, sus alcances y explicamos algunas de sus secciones principales que son de interés para la mejor comprensión de esta tesis.

2.1. Desarrollo del Estándar

El Comité Federal de Datos Geográficos comenzó a trabajar en la primera versión del estándar en Junio de 1992, por medio de un foro sobre metadatos espaciales. De este foro surgió borrador que fue publicado a partir de Octubre de 1992 y hasta Abril de 1993 para

una revisión pública. En el verano de 1993 se prueba el borrador revisado como parte del prototipo del Centro Distribuidor de Metadatos (Clearinghouse). Otros borradores se proporcionaron en Enero y Marzo de 1994. La primera versión fue aprobada el 8 de Junio de 1994.

La versión más reciente fue creada en Junio de 1998, y es la que tomamos como base para el desarrollo de la tesis. A continuación describimos los objetivos, alcances, organización y algunas de las secciones y elementos que conforman el estándar y que resultan de interés para este trabajo.

2.2. Objetivo

El objetivo de éste estándar es proveer un conjunto de definiciones y de terminología común para la documentación de datos geospaciales digitales. Establece los nombres y las definiciones de los elementos de datos y de los elementos compuestos (un conjunto de elementos de datos) que se usarán, así como los valores que podrán tomar dichos elementos.

La información incluida en el estándar fue seleccionada basada en 4 roles que los metadatos juegan, que se repiten continuamente y cuya importancia importancia varía dependiendo del usuario. Los roles son los siguientes:

- Disponibilidad. La información necesaria para determinar los conjuntos de datos que existen para una ubicación geográfica.
- Adaptabilidad para su uso. La información necesaria para determinar si un conjunto de datos satisface una necesidad específica.
- Acceso. Información necesaria para adquirir un conjunto de datos identificado.
- Transferencia. Información necesaria para procesar y usar un conjunto de datos.

2.3. Alcance del Estándar

Este estándar tiene el propósito de apoyar la colección y el procesamiento de los metadatos geoespaciales. Y tiene la intención de que sea utilizado a todos los niveles tanto del sector privado como del gobierno.

El estándar fue desarrollado desde la perspectiva de definir la información requerida por los futuros usuarios, para que puedan determinar la disponibilidad de un conjunto de datos geoespaciales; establecer si son o no aplicables al uso planeado; determinar la manera en la cual acceder al conjunto de datos y lograr de manera efectiva su transferencia.

2.4. Organización del Estándar

El estándar está organizado en una jerarquía de elementos de datos y elementos compuestos que definen la información de contenido para metadatos para documentar un conjunto de datos geoespaciales. El punto de partida es el elemento Metadato, el cual es un elemento que se forma de otros elementos compuestos para representar diferentes conceptos sobre el conjunto de datos. Cada uno de los elementos compuestos tiene una sección numerada en el estándar; cada sección empieza con el nombre y la definición del elemento compuesto que la define. En la figura 2.1 tenemos una visión general de la organización del estándar.

- Elemento compuesto: Es un grupo de elementos de datos y otros elementos compuestos. Los elementos compuestos son descritos por los elementos de datos, ya sea directa o indirectamente a través de otros elementos compuestos. La definición de los elementos compuestos se forma con: nombre del elemento compuesto, definición, un nombre corto y un tipo que es compuesto.
- Elemento de dato: Es un objeto primitivo de dato. Incluye el nombre del elemento de dato, su definición, una descripción de los valores que puede tomar (dominio) y un nombre corto del elemento.

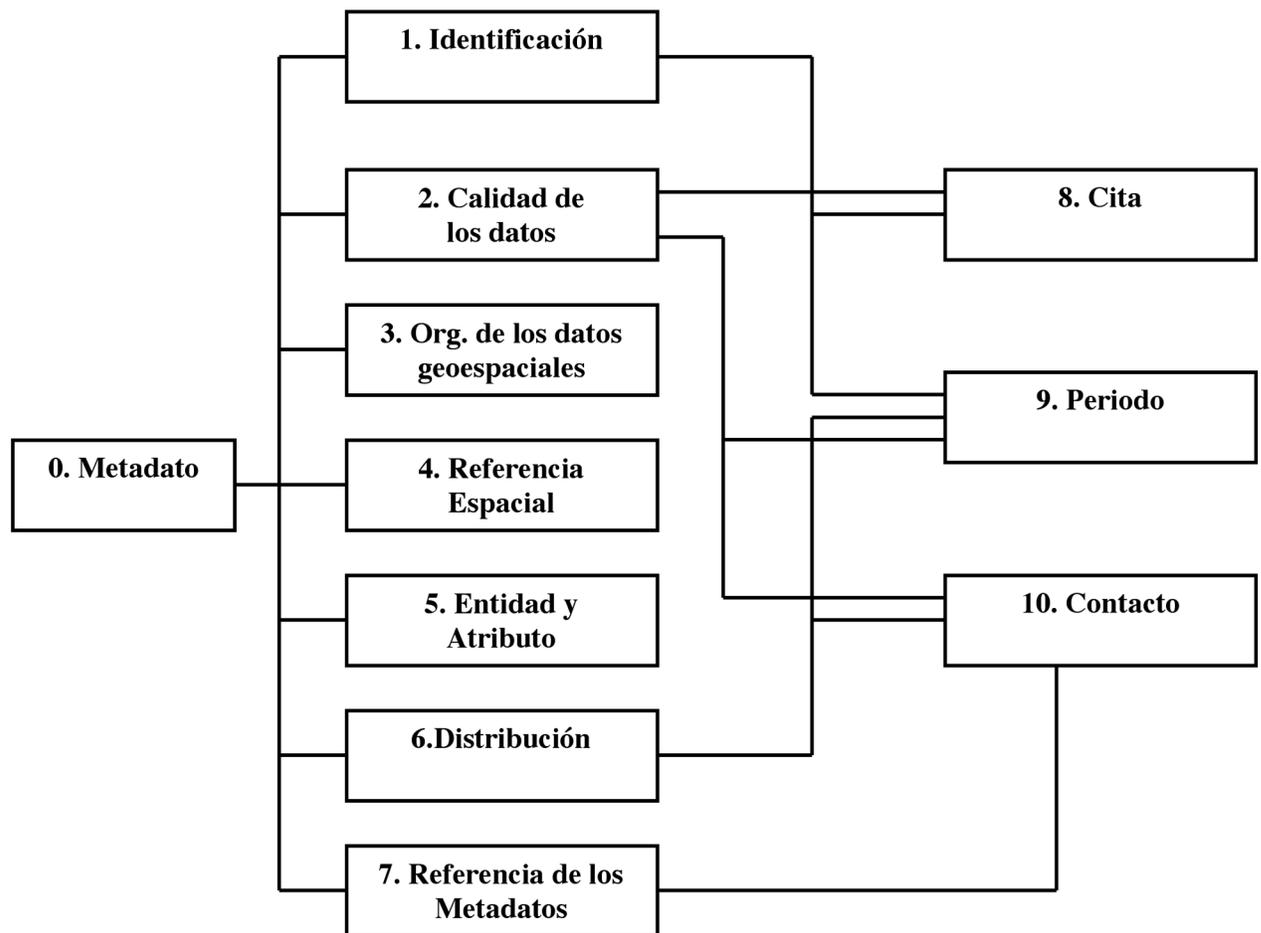


Figura 2.1: Organización del Estándar para Contenido de Metadatos Geospaciales Digitales

De estas definiciones, tenemos que, los tipos de dato pueden ser: entero, real, texto, fecha y hora. Los nombres cortos consisten de un máximo de 8 caracteres ASCII. Además, la existencia de los elementos puede ser:

- Obligatoria.
- Obligatoria si aplica. Este concepto significa que si el conjunto de datos presenta (o, para los conjuntos de datos en los que se trabaja o se tienen planeados, se conoce que presentará) un cierta característica, entonces el productor deberá proveer la información necesaria para describir esa característica.
- Opcional. Los elementos opcionales son proporcionados a discreción por el proveedor.

2.5. Secciones del Estándar

A continuación se presenta información más detallada de las secciones principales y de las secciones de apoyo del estándar. El punto de partida es el elemento compuesto Metadato (sección 0), y a partir de él existen 7 secciones principales y 3 secciones más de apoyo.

2.5.1. Metadato

Metadato. Contiene información acerca del contenido, calidad, condición y otras características del dato.

- Tipo: Compuesto.
- Nombre corto: metadata

Se compone de los siguiente elementos compuestos:

- Información de identificación (Sección 1). Contiene la información básica del conjunto de datos. Su nombre corto es **idinfo**. Este elemento es obligatorio.

- Información de calidad de los datos (Sección 2). Da una explicación general acerca de la calidad del conjunto de datos. Su nombre corto es **dataqual**. Esta sección es opcional.
- Información de organización de los datos espaciales (Sección 3). El mecanismo usado para representar información espacial en el conjunto de datos. El nombre corto de la sección es **spdoinfo**. La sección es de carácter opcional.
- Información de referencia espacial (Sección 4). Es la descripción del marco de referencia para codificar los datos y las coordenadas del conjunto de datos. Su nombre corto es **spref**. Esta sección es opcional.
- Información de entidad y atributo (Sección 5). Proporciona detalles del conjunto de datos, tales como, tipos de entidades, sus atributos, y los dominios de los cuales estos atributos pueden tomar sus valores. Esta sección es opcional y su nombre corto es **eainfo**.
- Información de distribución (Sección 6). Información acerca del distribuidor y de las opciones que hay para obtener los datos. Esta sección es opcional. El nombre corto es **distinfo**.
- Información de referencia de los metadatos (Sección 7). Contiene la información del estado actual de la información del metadato y la parte responsable. Esta sección es obligatoria y su nombre corto es **metainfo**.

2.5.2. Identificación

Esta sección proporciona información básica para identificar a un conjunto de datos. Se compone de varios elementos que a continuación se listan, al lado del nombre de cada elemento se encuentra entre paréntesis el nombre corto que le corresponde.

- Cita (citation). Elemento obligatorio que consiste de la información contenida en la sección de apoyo llamada Información de Cita.

-
- Descripción (descript). Este elemento proporciona una caracterización de los datos, incluyendo a que uso está destinado y sus limitaciones. Es obligatorio y se compone de:
 - Resumen (abstract). Una breve explicación del conjunto de datos. Este elemento es obligatorio.
 - Propósito (purpose). Elemento obligatorio que es un resumen de las intenciones con las cuales se desarrolló el conjunto de datos.
 - Información adicional (supplinf). Su existencia es opcional, se trata de otra información descriptiva acerca del conjunto de datos.
 - Periodo del contenido (timeperd). Se conforma por:
 - Referencia de actualidad (current). Es la base con la cual se determina el periodo de la información.
 - Estado (status). Se refiere al estado de la información y de su mantenimiento con respecto al conjunto de datos. Es obligatorio. Este elemento compuesto consiste de los elementos siguientes:
 - Progreso (progress). Es el estado actual del conjunto de datos. Obligatorio. Puede tomar los siguientes valores: completo, en desarrollo o en planeación.
 - Frecuencia de mantenimiento y Actualización (update). Es la frecuencia con la cual se realizan cambios al conjunto de datos después de que el conjunto de datos inicial está completo. Obligatorio. Puede tomar uno de los siguientes valores: continuamente, diariamente, semanalmente, mensualmente, anualmente, conforme se necesite, irregular, no planeado, desconocido.
 - Dominio espacial (spdom). Es el dominio del área geográfica del conjunto de datos y es obligatorio. Se forma con los siguientes elementos:
 - Coordenadas límite (bounding). Son los límites de cobertura de un conjunto de datos. Obligatorio. Estos límites se expresan como valores de latitud y longitud en el orden siguiente:

- Coordenada del límite Oeste (westbc). Se expresa en longitud, su gama de valores es:
–180,0 <= Coordenada Límite Oeste < 180,0. Obligatoria.
- Coordenada del límite Este (eastbc). Expresada en longitud, dentro del intervalo:
–180,0 <= Coordenada Límite Este < 180,0. Obligatoria.
- Coordenada del límite Norte (northbc). Se expresa en latitud dentro del intervalo:
–90,0 <= Coordenada Límite Norte < 90,0. Obligatoria.
- Coordenada del límite Sur (southbc). Expresada en latitud dentro del intervalo:
–90,0 <= Coordenada Límite Sur < 90,0. Obligatoria.
- Polígono-G del conjunto de datos (dsgpoly). Son las coordenadas que definen el contorno del área cubierta por el conjunto de datos. Este elemento es opcional y se compone de los elementos siguientes:
 - Anillo-G externo del Polígono-G del conjunto de datos. (grnpoin). Es el límite cerrado de un área. Obligatorio. Definido por un mínimo de 4 elementos Punto del Anillo-G o por un solo elemento Anillo-G:
 - ◇ Punto del Anillo-G (grngpoin). Una sola ubicación geográfica definida por una latitud: –180,0 <= Longitud Anillo-G < 180,0
y una longitud: –90,0 <= Latitud Anillo-G < 90,0
 - ◇ Anillo-G (gring). Es un conjunto de pares ordenados, separados por comas, en el cual el primer punto en cada par es la longitud y el otro es la latitud.
 - Anillo-G de exclusión del Polígono-G del conjunto de datos (dsgpolyx). Esta es la frontera cerrada de un área vacía o un hoyo en un área interior. Obligatorio si aplica. Se expresa usando los mismos elementos que el Anillo-G externo del conjunto de datos
- Palabras clave (keywords). Son las palabras o frases que resumen un aspecto del

conjunto de datos. Consisten de:

- Tema (theme). Contiene asuntos que son tratados por el conjunto de datos. Es obligatorio. Se compone de:
- Lugar (place). Ubicaciones geográficas caracterizadas por el conjunto de datos. Es obligatorio si aplica.
- Estrato (stratum). Capa o posición vertical caracterizada por el conjunto de datos. Obligatorio si aplica.
- Temporal (temporal). Periodo caracterizado por el conjunto de datos. Obligatorio si aplica.

Cada uno de los elementos anteriores se compone a su vez de:

- Sinónimo de la palabra clave. Es una referencia a una diccionario de sinónimos formalmente registrado de las palabras clave.
 - Palabra clave. Es una palabra o frase que se usa comúnmente para describir la información correspondiente a cada punto.
-
- Restricciones de Acceso (acconst). Prerequisitos legales para acceder a los datos. Estos incluyen cualquier restricción de acceso para asegurar la propiedad intelectual y cualquier limitación para obtener el conjunto de datos. Esta información es obligatoria. En caso de no existir alguna restricción, se debe especificar: ninguna.
 - Restricciones de Uso (useconst). Cualquier requisito legal para usar el conjunto de datos una vez que el acceso fue concedido. Este elementos es obligatorio. Si no existe restricción, especificar: ninguna.
 - Punto de Contacto (ptcontac). Información de contacto de un individuo u organización que se conozca sobre el conjunto de datos. Es de carácter opcional; su contenido se basa en la sección de apoyo Información de Contacto.
 - Gráficas de navegación (browse). Una gráfica que provee una ilustración del conjunto de datos. La gráfica debe incluir una leyenda para poder interpretarla. Estas gráficas

de navegación son opcionales, contienen: un nombre de archivo de la gráfica, una descripción de la ilustración y el tipo de archivo de la gráfica, este último puede ser: cgm, eps, emf, gif, jpeg, pbm, ps, tiff, wmf o xwd.

- Crédito del conjunto de datos (datacred). Es el reconocimiento para aquellos que contribuyeron al conjunto de datos.
- Información de seguridad (secinfo). Contiene las restricciones impuestas al conjunto de datos por cuestiones de seguridad nacional, privacidad, etc. Esta información es opcional. Se conforma de lo siguiente: nombre del sistema de clasificación de seguridad, nombre de la restricción, este puede ser: muy secreto, secreto, confidencial, restringido, sin clasificar, sensible u otro. Por último contiene una descripción adicional acerca del manejo de las restricciones.
- Ambiente nativo del conjunto de datos (native). Una descripción del ambiente de procesamiento del productor del metadato; esto incluye, el nombre del software y la versión, el sistema operativo, el nombre del archivo incluyendo el servidor, la ruta, el nombre del archivo y el tamaño del conjunto de datos. Es opcional.
- Referencia cruzada. Aquí se coloca información sobre otros conjuntos de datos que estén relacionados o, que pueden resultar de interés. Es opcional. La información contenida aquí se basa en la sección de apoyo Información de Cita.

2.5.3. Organización de los datos espaciales

Esta sección proporciona información sobre los mecanismos usados para representar la información espacial en el conjunto de datos. Su nombre corto es **spdoinfo**. Esta sección es obligatoria si aplica. Su contenido es el siguiente:

- Referencia espacial indirecta (indspref). Es el nombre de los tipos de características geográficas, esquemas de direccionamiento u otros medios a través de los cuales se hace referencia a una ubicación en el conjunto de datos. Entenderemos como referencia espacial indirecta, cualquier manera de describir una ubicación sin el uso

de coordenadas, por ejemplo, el nombre de un estado, de un municipio, o de un camino, para identificar a un lugar de manera única. El elemento referencia espacial indirecta es, obligatorio si aplica.

- Referencia espacial directa (*direct*). Es el sistema de objetos utilizado para representar espacio en el conjunto de datos. Este elemento es obligatorio. Su dominio es conformado por: punto, vector o raster.
- Información del objeto espacial. Este elemento consiste de dos elementos mutuamente excluyentes:
 - Información del objeto punto y vector (*ptvcinf*). Es el tipo y número de objetos espaciales vector o de puntos fuera de la malla que están en el conjunto de datos. Obligatorio. Este elemento contempla dos terminologías para describir a los objetos vector y punto, una de estas terminologías toma como base el Estándar para Transferencia de Datos Espaciales [10]; y la otra, se basa en Formato del Producto Vector [11], publicado por el Departamento de defensa de los Estados Unidos de América. Sólo debemos utilizar una de ellas.
 - En caso de usar la terminología del Estándar para Transferencia de Datos Espaciales, tenemos que especificar el tipo de objeto vector o punto, esto es, el nombre de los objetos espaciales usados para localizar un lugar ya sea de 0, 1 o 2 dimensiones dentro del conjunto de datos. Los tipos posibles son: punto, punto entidad, punto etiqueta, punto área, “nodo, gráfica planar”, “nodo, red”, cadena, enlace, cadena completa, “cadena área”, “cadena red, gráfica planar”, “cadena red, gráfica no-planar”, “arco circular, centro de tres puntos”, arco elíptico, B-Spline uniforme, polígono-G, anillo de composición mixta, anillo compuesto de cadenas, anillo compuesto de arcos, conjuntos de Bezier, polígono-GT compuesto de anillos, polígono-GT compuesto de cadenas, polígono universal compuesto de anillos, polígono universal compuesto de cadenas, polígono vacío compuesto de anillos, polígono vacío compuesto de cadenas. Opcionalmente además del tipo, se tiene el número total de objetos punto o vector en el conjunto de datos.

- Si la terminología usada es la establecida en el Formato Producto Vector, entonces debemos especificar el nivel de la topología, es decir, el grado de completitud de la topología del conjunto de datos [11], el cual debe estar dentro del intervalo: $0 \leq \text{Nivel de topología} \leq 3$; también debemos especificar el tipo de objeto punto o vector, en este caso, los valores posibles son: nodo, borde, cara o texto. Y finalmente, de forma opcional, el número total de objetos punto o vector.
- Información del objeto raster (rastinfo). Se refiere al tipo y número de objetos espaciales raster en el conjunto de datos. Esta información es de carácter obligatorio. Se compone de:
 - Tipo de objeto raster. Es el objeto usado para localizar lugares de 0, 1 y 2 dimensiones en el conjunto de datos. Los valores posibles son: punto, pixel, celda de malla, voxel.
 - Información de conteo. Esta información es opcional y consiste de: el número máximo de objetos raster a lo largo de los ejes x e y , usado para objetos raster rectangulares y en caso de objetos volumétricos, entonces también se especificará el máximo número de objetos a lo largo del eje z .

2.5.4. Referencia Espacial

La información de referencia espacial es una descripción del marco de referencia y los medios para codificar las coordenadas del conjunto de datos. El nombre corto de este elemento es **spref**, y es de carácter obligatorio si aplica. Lo conforman los siguientes puntos:

- Definición del sistema de coordenadas horizontal (horizsys). Es el marco de referencia o sistema a partir del cual, cantidades lineales o angulares son medidas y asignadas a la posición que ocupa un punto. Este sistema de coordenadas debe ser uno de los 3 siguientes: geográfico, planar o local; en caso de aplicar, también debe contener un elemento más, llamado modelo geodésico. Estos sistemas son descritos a continuación:

- Geográfico (geograph). Consiste de la latitud y longitud que definen un punto sobre la superficie de la Tierra con respecto a un esferoide de referencia.
 - Resolución de latitud (latres). La mínima diferencia entre 2 valores de latitud adyacentes expresados en unidades de coordenadas geográficas.
 - Resolución de longitud (lonres). Es la mínima diferencia entre 2 longitudes adyacentes expresados en unidades de coordenadas geográficas.
 - Unidades de coordenadas geográficas (geounit). Unidades de medida usados por la latitud y la longitud.
- Planar (planar). Son las distancias, o distancias y ángulos que definen la posición de un punto sobre un plano de referencia en el cual se ha proyectado la superficie de la Tierra.
 - Proyección del mapa (mapproj). Es la representación sistemática de toda o de una parte de la superficie de la Tierra sobre un plano. La proyección del mapa posee un nombre y un conjunto de parámetros acordes a la proyección elegida.

Las proyecciones o representaciones, pueden ser una de las siguientes: área cónica equivalente de Albers, Azimutal equidistante, cónica equidistante, equirectangular, perspectiva vertical general del lado más cercano, gnomonico, área equivalente azimutal de Lambert, cónica conforme a Lambert, Mercator, estereográfica modificada para Alaska, cilíndrica Miller, Mercator oblicuo, ortográfico, estereográfico polar, policónico, Robinson, senooidal, Mercator oblicuo del espacio, estereográfico, Mercator transversal, Van der Grinten.

En lo que se refiere al conjunto de parámetros, se debe incluir sus nombres, valores y la descripción de la relación matemática entre la Tierra y el plano.
 - Sistema de coordenadas de la malla (gridsys). Es un sistema de coordenadas en un plano rectangular, basado y ajustado matemáticamente como una proyección del mapa, de tal manera que resulte fácil el transformar las

coordenadas hacia y desde dicho plano.

- Planar local (localp). Es cualquier sistema de coordenadas planar de mano-derecha, donde el eje z coincide con una plomada a través del origen que es alineado localmente con la superficie de la Tierra.
- Local (local). Es la descripción de cualquier sistema de coordenadas que no está alineado con la superficie de la Tierra.
 - Descripción local (localdes). Una descripción del sistema de coordenadas y su orientación a la superficie de la Tierra.
 - Información de georeferencia local. Una descripción de la información proveída para registrar el sistema local a la Tierra (esto es, puntos de control, datos efímeros de satélites, datos de navegación inercial).
- Modelo geodésico (geodetic). Este modelo contiene los parámetros para la superficie de la Tierra. Estos parámetros son:
 - Nombre del dato horizontal (horizdn). Es la identificación del sistema de referencia usado para definir las coordenadas de los puntos.
 - Nombre del elipsoide (ellips). Identificación de las representaciones de la forma de la Tierra. Pueden ser: Clarke 1866, Sistema de Referencia geodésico 80, u otro sistema.
 - Eje semi-mayor (semiaxis). Es el radio del eje ecuatorial del elipsoide.
 - Denominador de razón de aplanamiento (denflat). El denominador de la razón de la diferencia entre los radios ecuatorial y polar del elipsoide cuando el numerador es puesto a 1.
- Definición del sistema de coordenadas vertical (vertdef). Es el marco de referencia o sistema a partir del cual se miden distancias (altitudes o profundidades).
 - Definición del sistema de altitud. Es el sistema o marco de referencia en base al cual se miden altitudes.

- Nombre del datum ¹ de altitud (altdatum). Es la identificación de la superficie tomada como referencia para la medición de altitudes.
- Resolución de altitud (altres). Es la mínima distancia posible entre 2 valores de altitud adyacentes, expresado en unidades de distancia de altitud.
- Unidad de distancia de altitud (altunits). Unidades en las cuales se registran altitudes. Pueden ser metro, pies u otra.
- Método de codificación de altitud (altenc). Los medios utilizados para codificar altitudes. Pueden ser: Coordenadas explícitas de elevación incluidas con coordenadas horizontales, coordenadas implícitas, valores de atributos.
- Definición del sistema de profundidad (depthsys). Es el marco de referencia o sistema en base al cual se miden profundidades.
 - Nombre del datum de profundidad (depthdn). Es la identificación dada a la superficie o referencia desde donde se miden las profundidades. Algunos de los valores que puede tener este elemento son: superficie local, “gráfica datum, datum para la reducción del sonido”, marea astronómica más baja, marea astronómica más alta, promedio aguas bajas, promedio aguas altas, promedio de nivel del mar, datum del reconocimiento de tierra, promedio de fuentes de aguas bajas, promedio de fuentes de aguas altas, “sin corrección”, también es posible especificar algún otro tipo.
 - Resolución de profundidad (depthres). Es la mínima distancia posible entre 2 valores de profundidad adyacentes, expresado en unidades de distancia de profundidad. Debe ser mayor que 0.0.
 - Unidades de distancia de profundidad (depthdu). Son las unidades en las cuales se registran las profundidades. Pueden ser: metros, pies o, alguna otra.
 - Método de codificación de profundidad (depthem). Son los medios utilizados para codificar profundidades. Estos puede ser: coordenadas explícitas

¹Datum: Un conjunto de parámetros y puntos de control usados para definir de forma precisa la forma de la Tierra (i.e. como un esferoide). El datum es la base para un sistema de coordenadas planar. <http://www.kam.to/kam/services/gisdictionary.cfm>

de profundidad junto con coordenadas horizontales, coordenadas implícitas o, valores de atributos.

2.5.5. Referencia de Metadatos

Esta sección nos proporciona información sobre el estado actual de los metadatos y de la parte responsable. Se compone de los siguientes elementos:

- Fecha de los metadatos (metd). Es la fecha en la cual se crearon los metadatos o de la última actualización. Esta es obligatoria.
- Fecha de revisión de los metadatos (metrd). Es la fecha de la última revisión de los metadatos y es opcional.
- Fecha de siguiente revisión de los metadatos (metfrd). Es la fecha en la que se efectuará una nueva revisión a los metadatos, es opcional.
- Contacto del metadato (metc). Se refiere a la parte responsable por la información del metadato. Este elementos es obligatorio y su contenido está basado en la sección de apoyo Información de Contacto.
- Nombre estándar de los metadatos (metstdn). Es obligatorio. Puede tomar el valor, para este caso de: Estándar de Contenido para Metadatos Geoespaciales Digitales.
- Versión del estándar de los metadato (metstdv). Esta se debe especificar de manera obligatoria.
- Convención de la hora del metadato (mettc). Esta es utilizada si la información del conjunto de datos incluye la hora del día. Este elemento es obligatorio si aplica. Puede tomar los valores: horario local, horario local con factor de diferencia u horario universal.
- Restricciones de acceso a los metadatos (metac). Prerequisitos legales para acceder a los metadatos. Estos incluyen cualquier restricción de acceso para asegurar

la propiedad intelectual y cualquier limitación para obtener los metadatos. Estas restricciones son opcionales.

- Restricciones de uso de los metadato (metuc). Cualquier requisito legal para usar los metadatos una vez que el acceso fue concedido. Estas restricciones son opcionales.
- Información de seguridad del metadato (metsi). Contiene las restricciones impuestas a los metadatos por cuestiones de seguridad nacional, privacidad, etc. La existencia de este elemento es opcional. Su contenido se forma de los siguientes elementos: nombre del sistema de clasificación de seguridad, nombre de la restricción, este ultimo puede ser: muy secreto, secreto, confidencial, restringido, sin clasificar, sensible u otro. Por último contiene una descripción adicional acerca del manejo de las restricciones.
- Extensiones a los metadatos (metextns). Esta es una referencia a elementos extendidos del estándar definidos por un productor de metadatos o por un usuario de la comunidad. Son obligatorias si aplica. Se compone de: el nombre del documento que describe la aplicación del estándar a una comunidad de usuarios específica y de una dirección electrónica que contenga la información de la extensión de los metadatos para el conjunto de datos.

2.6. Secciones de Apoyo

Como se mencionó en la sección anterior, el estándar contiene además de las 7 secciones principales, 3 secciones de apoyo. Estas últimas son utilizadas por varias de las secciones principales pero nunca son usadas de manera independiente.

2.6.1. Cita

La información de cita es la referencia recomendada para ser usada por el conjunto de datos. Su nombre corto es **citeinfo**. Es un elemento compuesto que contiene:

- Originador (origin). Es el nombre de la organización o individuo que desarrolló el conjunto de datos. Obligatorio.

- Fecha de publicación (pubdate). Es la fecha en la cual el conjunto de datos es publicado. Obligatorio. Puede tomar los siguientes valores: desconocida, material sin publicar, o una fecha.
- Hora de publicación (pubtime). Es la hora del día en la cual el conjunto de datos fue publicado. Opcional.
- Título (title). Es el nombre del conjunto de datos. Obligatorio.
- Edición (edition). Es la versión del título. Obligatorio si aplica.
- Forma de presentación de los datos geoespaciales (geoform). Es la forma en la cual los datos geoespaciales son representados. Obligatorio si aplica. Los valores posibles son: atlas, audio, diagrama, documento, globo, mapa, modelo, presentación multimedia, perfil, dato digital raster, video, vista, otro.
- Información de la serie (serinfo). Es la identificación de la publicación serial de la cual el conjunto de datos forma parte. Obligatorio si aplica. Este elemento está conformado por: nombre de la publicación serial y por un identificador del ejemplar de la publicación.
- Información de la publicación (pubinfo). Detalles de la publicación para los conjuntos de datos publicados. Obligatorio si aplica. Se compone de:
 - Lugar de publicación. Nombre de la ciudad, estado o provincia y país donde el conjunto de datos fue publicado o liberado.
 - Editor. El nombre de la organización o individuo que publicó el conjunto de datos.
- Otros detalles de la cita (othercit). Otra información requerida para completar la cita. Obligatoria si aplica.
- Enlace en línea (onlink). Dirección electrónica que contenga al conjunto de datos. Es opcional.

- Cita más larga (lworkcit). Es la información que identifica un trabajo más grande en el cual el conjunto de datos es incluido. Obligatoria si aplica.

2.6.2. Periodo

Esta sección contiene información acerca de la fecha y hora de un evento. Es obligatoria. Puede tomar 1 de 3 formas:

- Fecha / hora simples (sngdate). Es la manera de codificar una sola fecha y hora. Es obligatoria. Se forma de:
 - Fecha de calendario. Es obligatoria. Consiste de el año y opcionalmente el mes, o el mes y el día.
 - Hora del día. Es opcional, se forma de: la hora y opcionalmente los minutos o, los minutos y segundos.
- Múltiples fechas y horas (mdattim). Es la manera de codificar varias fechas y horas. Es obligatoria. Se compone de dos o más elementos de tipo fecha / hora simples.
- Intervalo de fechas (rngdates). Manera de codificar un intervalo de fechas y hora. Es obligatoria. Se compone de una fecha de inicio y una fecha de fin del evento. Y opcionalmente hora de inicio y hora de fin del evento.

2.6.3. Contacto

La sección Contacto cuyo nombre corto es **cntinfo**, proporciona la identidad y los medios para comunicarse con la(s) personas u organizaciones asociadas con el conjunto de datos. Obligatorio. Está constituido por:

- Contacto principal. Este puede tomar una de 2 formas:
 - Persona principal de contacto (cntperp). La persona y la afiliación de la persona, asociada con el conjunto de datos. Obligatorio. Utilizado en casos donde la asociación de la persona con los datos es más importante o significativa que

la asociación entre la organización y el conjunto de datos. En este caso tiene 2 elementos: el nombre de la persona de contacto y opcionalmente, el nombre de la organización

- Organización principal de contacto (cntorgp). La organización y el miembro de la organización asociados con el conjunto de datos. Obligatorio. Es usado en casos donde la asociación de la organización con el conjunto de datos es más significativa que la asociación de la persona con el conjunto de datos. Contiene el nombre de la organización y opcionalmente el de la persona de contacto.
- Posición del contacto (cntpos). Se refiere al cargo que ocupa el individuo. Es un elemento opcional.
- Dirección del contacto (cntaddr). La dirección de la organización o del individuo. Obligatorio. Consta de lo siguiente:
 - Tipo de dirección. Obligatorio. Puede tomar los siguientes valores: de correo, física, física y de correo, o también es posible especificar otro tipo.
 - Dirección. Una línea para la dirección. Obligatorio si aplica.
 - Ciudad. Obligatorio.
 - Estado o provincia. Obligatorio.
 - Código postal. Obligatorio.
 - País. Opcional.
- Teléfono de voz del contacto (cntvoice). El número de teléfono por el cual se puede hablar con la organización o con el individuo. Obligatorio.
- Teléfono TDD/TTY del contacto (cnttdd). El número de teléfono por el cual, personas con problemas de audición pueden contactar a la organización o individuo. Opcional.
- Teléfono de fax (cntfax). Opcional.

- Dirección de correo electrónico (entemail). Opcional.
- Horas de servicio (hours). Periodo en el cual es posible hablar con la organización o individuo. Opcional.
- Instrucciones de contacto (entinst). Información adicional sobre cuando y como se puede contactar a la organización o individuo. Opcional.

En este capítulo hemos presentado los objetivos, alcances y algunas de las secciones que conforman el Estándar de Contenido para Metadatos Geoespaciales Digitales, ya que es con base en él que decidimos desarrollar nuestra aplicación. Pusimos especial énfasis en las secciones del estándar que son de carácter obligatorio y en las secciones de apoyo. Explicamos también las secciones de organización de datos espaciales y la de referencia espacial, porque son de interés para futuros trabajos. Si se desea mayor detalle del resto del estándar, se puede consultar directamente en [3].

Capítulo 3

Sistema Meta-X

A lo largo de los capítulos anteriores, tratamos los conceptos que, a nuestra consideración, resultan de relevancia al trabajar con metadatos. Empezamos explicando el concepto mismo de metadatos, vimos que no hay una definición única para éste término, sino que, esta varía de acuerdo al entorno en el cual estemos o tengamos planeado trabajar. Es esta diversidad de entornos, la que ha dado pie al surgimiento de distintos estándares o esquemas de metadatos; entre estos esquemas, encontramos el estándar para contenido de metadatos digitales geoespaciales digitales, cuyas características fueron explicadas en el capítulo anterior.

También en la sección 1.4.1, observamos que existen varias herramientas para crear metadatos que trabajan sobre el estándar ya mencionado. Y como se pudo observar, pocas son las que trabajan en el sistema operativo UNIX y, algunas otras basadas en web operan en cualquier plataforma; el resto de las herramientas (la mayoría), trabajan sobre el sistema operativo Windows en sus distintas versiones. Dentro de este conjunto de herramientas, algunas tienen un costo, otras son de libre distribución pero requieren de otros sistemas para trabajar, los cuales sí tienen un costo.

En esta tesis proponemos el sistema Meta-X, el cual apoyará varias actividades involucradas en el desarrollo de aplicaciones de bases de datos, en un nivel que denominamos metadiseño.

En este capítulo presentamos la arquitectura general del sistema, y posteriormente expli-

camos la arquitectura de la herramienta para creación de metadatos.

3.1. Arquitectura del Sistema

La arquitectura del sistema Meta-X, sigue el modelo cliente-servidor. Consta de 2 capas:

- Herramienta de creación de metadatos.
- Servidor de bases de datos.

Lo anterior se observa en la figura 3.1:

3.1.1. Herramienta de creación de metadatos

La herramienta de creación Meta-X consta de tres módulos, cada uno de ellos relacionados con su respectivo repositorio. Así tenemos que hay una herramienta para poder realizar la recolección de requerimientos, otra para edición de reglas y un tercer módulo para creación de metadatos.

En esta tesis, nos encargamos del tercer módulo, es decir, del de creación de metadatos. Este módulo se encarga de la captura de metadatos, la información que capturamos se basa en el estándar para metadatos, y nos enfocamos a sus secciones que son de carácter obligatorio: identificación y referencia de metadatos. Y también a las secciones de apoyo: cita, periodo y contacto.

La herramienta trabaja de forma independiente (standalone), y los datos capturados son salvados en formato XML.

3.1.2. Servidor de Bases de Datos

En esta capa, tal como se observa en la figura 3.1, manejamos varias bases de datos, a las cuales dividimos en dos grupos, cada una de ellas destinadas a almacenar distinta información relacionada con el proyecto que se va a desarrollar. Estas bases de datos son:

- Base del proyecto.

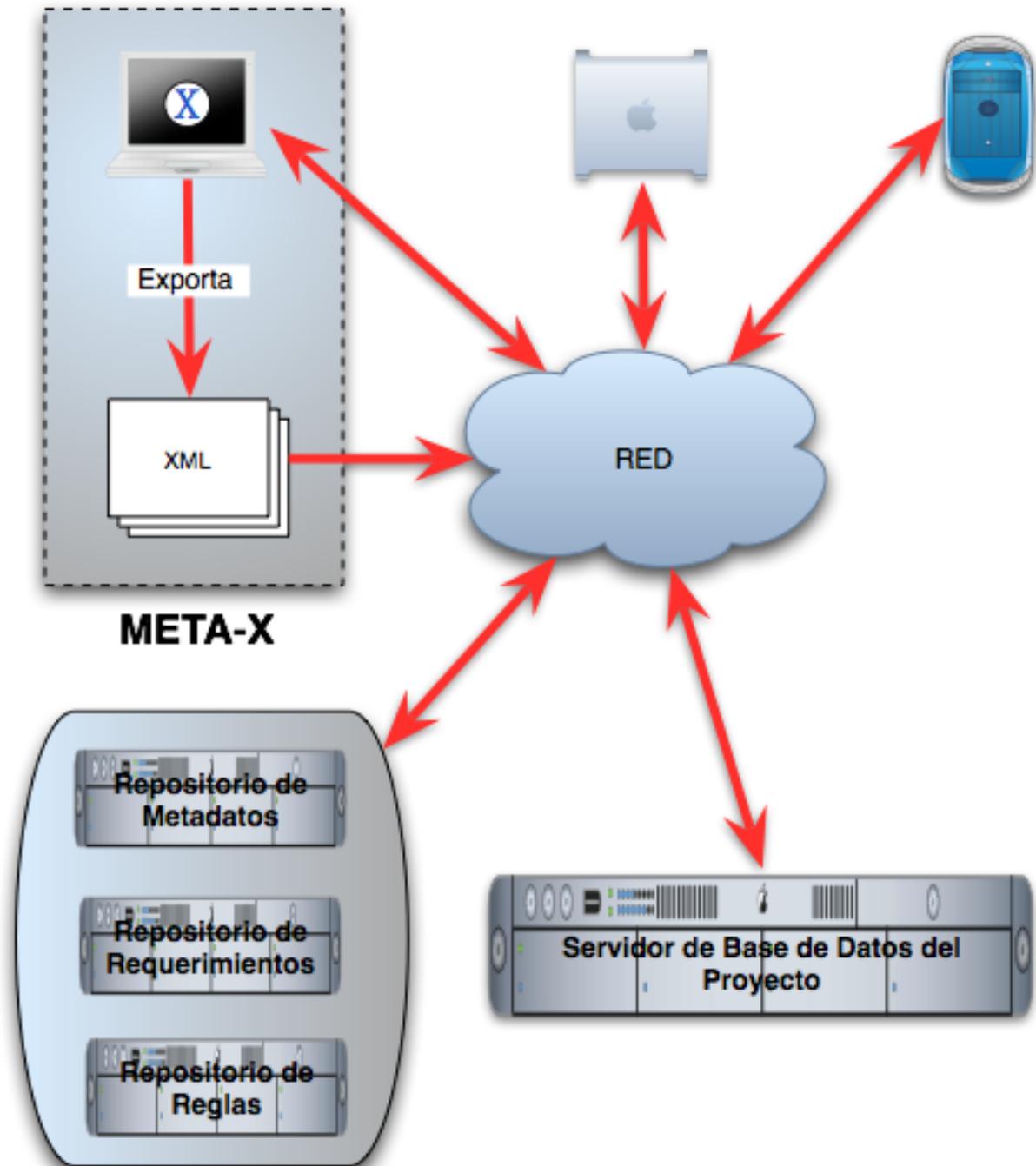


Figura 3.1: Arquitectura General del Sistema

- Repositorios de metadiseño. Este grupo se compone de:
 - Base del Proyecto.
 - Base de Requerimientos.
 - Base de Reglas.
 - Base de Metadatos.

Base del proyecto

Podemos considerar a la base de datos del proyecto, como la base principal, puesto que contiene la información exclusiva del sistema que se está desarrollando.

Repositorio de requerimientos

Este repositorio está destinado a almacenar los requerimientos del sistema, para posteriormente poder extraer una especificación de requerimientos. Obtener los requerimientos, es todo un proceso que, a través de entrevistas y otras técnicas de recolección, nos lleve a determinar lo que un sistema debe hacer y que propiedades debe poseer [15]. Así, este repositorio contendrá los siguientes tipos de requerimientos:

- Funcionales. Aquí estableceremos lo que el sistema debe hacer.
- No-Funcionales. Estos requerimientos son las propiedades que el sistema que estemos desarrollando debe tener.
- Restricciones. Son las limitaciones que se imponen al proyecto. Por ejemplo, las razones para desarrollar el sistema, o el vocabulario que dicho sistema usará.

Uno de los factores relevantes para mantener una base de requerimientos es que, conforme pasa el tiempo, surgirán nuevos requerimientos y será necesario llevar un seguimiento del estado del sistema, necesitaremos saber, cuando y en que forma ha sufrido modificaciones, y probablemente, quién estaba a cargo cuando dichas modificaciones fueron realizadas. Si tenemos una base de requerimientos podremos realizar consultas a esta para recuperar esa información.

Repositorio de reglas

En este repositorio se almacenarán las reglas que se encargarán de mantener la semántica de la base. La semántica de la base se refiere al significado de los datos almacenados en ella. La razón por la cual resulta importante tomar en cuenta la semántica, es porque nos da la oportunidad de determinar si los datos que estamos almacenando son correctos, es decir, consistentes con los elementos del mundo real a los que representan. A estas reglas, se les conoce como reglas semánticas ó reglas de integridad [5].

Proponemos que las reglas almacenadas en este repositorio sigan el paradigma Evento-Condición-Acción. Tal como su nombre lo dice, se componen de tres elementos:

- **Evento.** Describe un suceso al cual una regla es capaz de responder. Este puede ser una petición a la base de datos para inserción, borrado o actualización de tuplas.
- **Condición.** Una condición es una expresión que evalúa el contexto en el cual se presentó el evento.
- **Acción.** La acción consiste de una función que especifica la manera de responder al evento una vez que la condición resultó verdadera.

Repositorio de metadatos

Este repositorio contendrá los metadatos del sistema que se esté desarrollando. El objetivo de tener este repositorio es que sea posible compartir la información que contiene, con quien la necesite para elaborar otro proyecto. Tomará como base el estándar de contenido para metadatos geográficos.

3.2. Diseño e Implementación de Meta-X

En esta sección explicaremos la manera en la que fue desarrollada la herramienta para creación de metadatos Meta-X. Las clases que la componen, sus métodos y atributos.

Desarrollamos el sistema Meta-X siguiendo el paradigma orientado a objetos y utilizando el lenguaje Objective-C y Cocoa para el sistema operativo Mac OS X.

3.2.1. ¿Qué es Cocoa?

Cocoa es un conjunto de objetos de software que implementa casi todas las características que son comunes en aplicaciones para Mac OS X. A Cocoa la componen varios frameworks que contienen bibliotecas de objetos, funciones, tipos de datos y documentación. Los frameworks más importantes para Cocoa son: el Application Kit y el Foundation. Tal como se ve en la figura 3.2

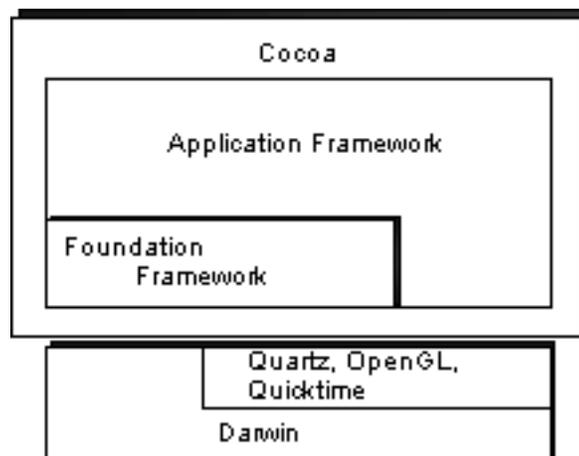


Figura 3.2: Cocoa y sus componentes

El Foundation contiene los objetos no gráficos que son útiles en cada aplicación de Cocoa, mientras que el Application Kit es el que provee objetos gráficos y los elementos para las interfaces gráficas de usuario.

3.2.2. Meta-X, aplicación multidocumentos

Las aplicaciones que manejan múltiples documentos se han vuelto comunes hoy en día, lo vemos en procesadores de texto, hojas de cálculo, programas de dibujo, entre otras. Primero, definamos lo que es un documento, esto lo podemos ver desde 2 perspectivas, la del usuario y la del programador.

Desde el punto de vista del usuario, un documento es un solo cuerpo de información que está en su propia ventana. El usuario entonces puede crear muchos documentos y salvar

cada uno de ellos a un archivo.

Desde la perspectiva del programador, un documento esta formado por los objetos que son extraídos de un archivo de interfaz (archivo nib) y el objeto controlador que carga y administra estos objetos [9].

Cocoa, a través del Application Kit, nos ofrece un framework que facilita la mayor parte del trabajo para el manejo de múltiples documentos. Nuestra aplicación Meta-X sigue esta arquitectura, conocida como basada en documentos. La cual exponemos en la figura 3.3.

Las aplicaciones basadas en documentos siguen el paradigma llamado Modelo-Vista-Controlador (MVC), el cual permite que exista una completa separación entre los datos y la interfaz de usuario. El MVC separa la aplicación en 3 capas tal como lo indica su nombre, esto es, modelo, vista y controlador [1]. La función que desempeña cada una de ellas es la siguiente:

- Modelo. Es aquí donde residen los datos de la aplicación, también encontramos el comportamiento relacionado directamente con los datos.
- Vista. Aquí encontramos la interfaz gráfica de usuario.
- Controlador. Su rol es ser un intermediario entre la vista y el modelo.

Lo anterior lo observamos con más claridad en la figura 3.4.

En el framework que nos ofrece Cocoa, hay tres clases que interoperan para ayudarnos a implementar aplicaciones basadas en documentos, estas son: `NSDocumentController`, `NSDocument` y `NSWindowController`. Estas clases tienen una relación uno a muchos entre sí, esto es, una aplicación basada en documentos tiene un `NSDocumentController`, el cual puede crear y manejar muchos objetos `NSDocument`. A su vez, los objetos `NSDocument` pueden crear y manejar varios objetos `NSWindowController`.

En la figura 3.5 observamos las clases que participan en Meta-X. Ahora describiremos cada una de estas clases.

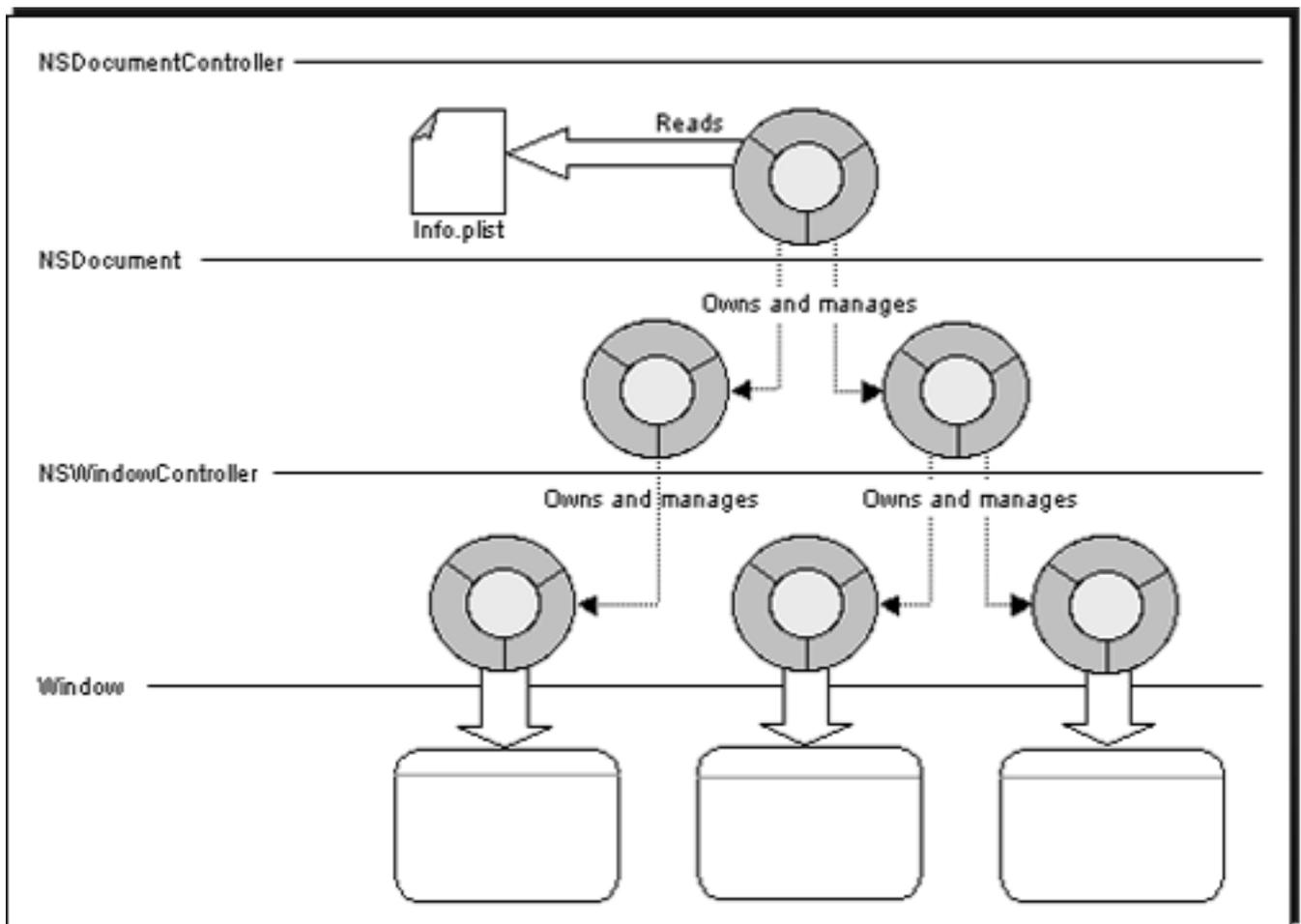


Figura 3.3: Arquitectura basada en documentos

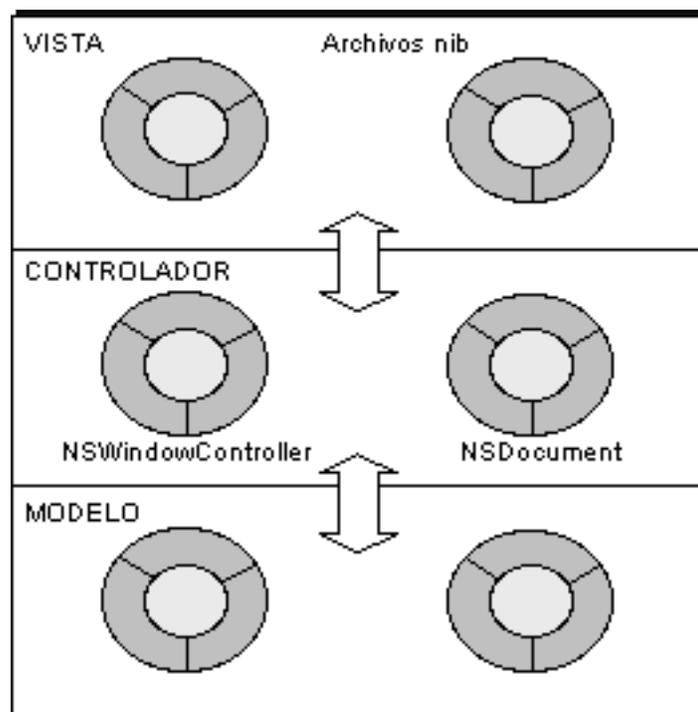


Figura 3.4: Paradigma MVC

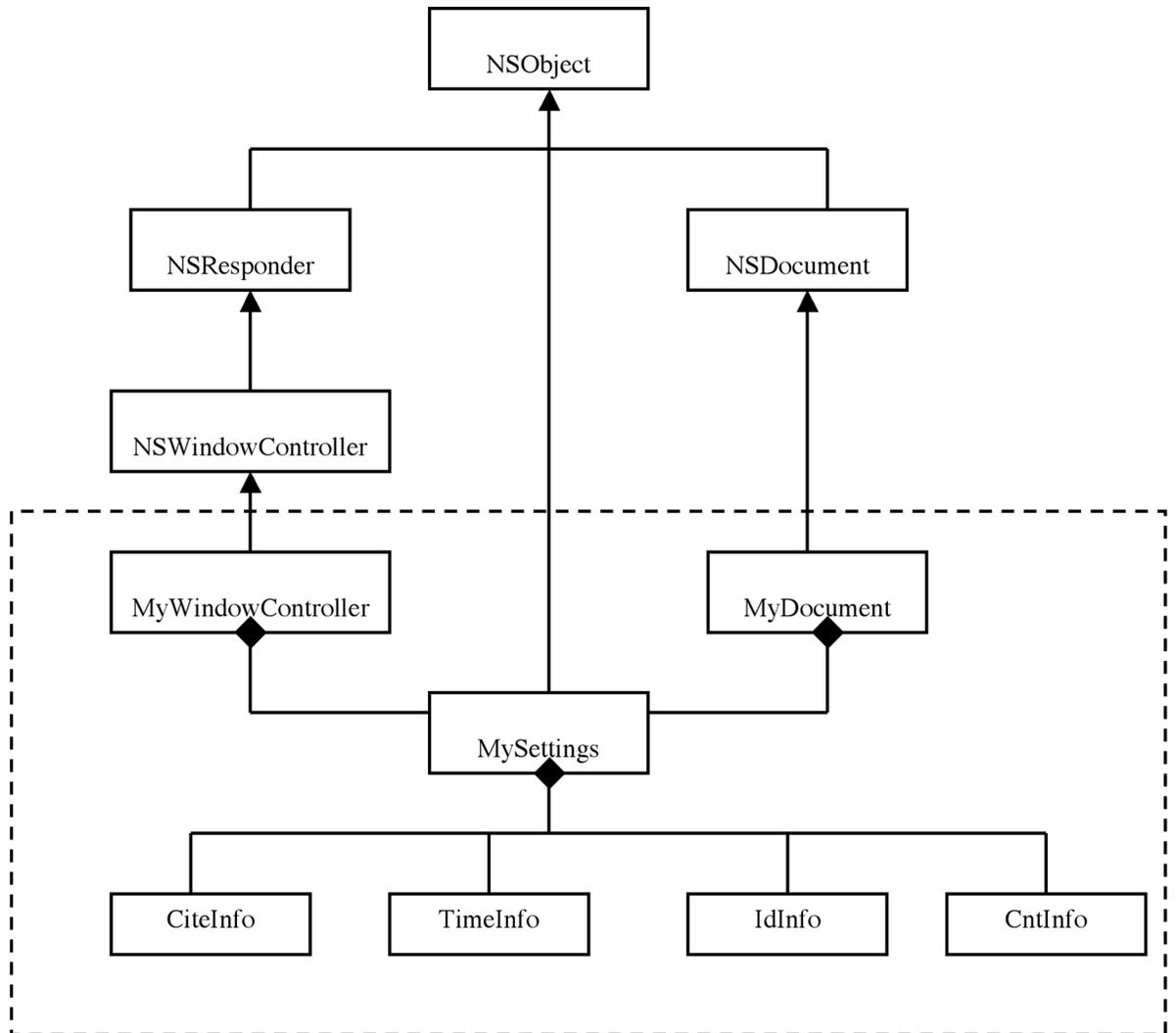


Figura 3.5: Meta-X y su arquitectura multidocumento

NSDocument

La clase NSDocument es una clase abstracta. Las subclases de NSDocument son usados para encapsular los datos asociados con el documento. Esta clase implementa las acciones estándar que son enviadas por los objetos del menú Salvar, Salvar Como, Deshacer y Cerrar.

NSDocumentController

Esta clase ayuda a la creación de nuevos documentos y a abrir documentos ya existentes. Sólo debe haber una instancia de esta clase en cualquier aplicación que programemos. NSDocumentController se encarga de crear una nueva instancia de la subclase de NSDocument, cada vez que desde el menú de usuario elegimos Nuevo.

NSWindowController

La clase NSWindowController se encarga de manejar una ventana asociada con un documento, la cual es comunmente almacenada en un archivo nib. Si tuvieramos múltiples ventanas, cada una tendría su propio controlador de ventana.

3.2.3. Vista en Meta-X

La interfaz gráfica de Meta-X consta de dos archivos: MainMenu.nib y MyDocument.nib. El primero es la barra de menú del usuario. El segundo contiene la ventana principal de nuestra aplicación.

Como observamos en la figura 3.6, en vez de tener varias ventanas para capturar la información, nosotros decidimos hacer uso de pestañas (tabs), lo cual permite capturar mayor cantidad de información relacionada y mantenerla en una única ventana, sin tener que llenar la pantalla con ventanas. Nuestra ventana principal, contiene 2 pestañas: información de identificación y referencia de metadatos. Estas tienen los campos necesarios para capturar a los metadatos.

The screenshot displays the 'Meta-X Main Window' interface, version 1.0.0. The window title bar includes standard OS window controls (minimize, maximize, close) and the text 'Meta-X Main Window' and 'Meta-X Version 1.0.0'. Below the title bar, there are two tabs: 'Identification' (selected) and 'Metadata Reference'. The main content area is titled 'Identification Information' and contains a sub-menu with five tabs: 'Contact Information', 'Identification' (selected), 'Cite Information', 'Period Information', and 'Spatial Domain'. The 'Identification' tab is active, showing a form with the following fields:

- Abstract:** A large text input area.
- Purpose:** A large text input area.
- Supplemental Information:** A large text input area.
- Status:** A text input field.
- Keywords:** A text input field.
- Access Constraints:** A large text input area.
- Use Constraints:** A large text input area.

Figura 3.6: Vista de la interfaz gráfica de Meta-X

3.2.4. Controladores en Meta-X

En esta sección explicaremos los controladores de nuestra aplicación, empezaremos describiendo a la clase controladora de ventanas, a la cual llamamos `MyWindowController`, posteriormente tratamos con la clase `MyDocument`.

`MyWindowController`

Esta clase hereda de la `NSWindowController`. `MyWindowController` es el que designamos como dueño del archivo de interfaz `MyDocument.nib`. Esto lo hacemos porque esta aplicación manejará en un futuro otras ventanas, tal como lo observamos en el esquema general de la aplicación, así a cada ventana se le asignará su propio controlador.

En esta clase declaramos todas las variables instancia necesarias para comunicarse con la interfaz gráfica, es decir, indicarle cuando cambiar el estado visible de los objetos que contiene o, por el otro lado, obteniendo el estado de los mismos cuando necesitamos esa información. A estas variables se les conoce como *outlets*. Cada una de estas outlets, corresponde con cada uno de los objetos que colocamos en la interfaz gráfica con la ayuda de Interface Builder.

En `MyWindowController` también tenemos un conjunto de métodos de acceso a los outlets. Estos métodos los usamos para poder consultar el estado de una variable, así como también para que sea posible modificar dicho estado. Por cada outlet que tenemos en `MyWindowController`, tenemos un método de acceso, este tiene el mismo nombre de la variable instancia y su función es recuperar el valor que tenga esa variable. Las variables instancias de esta clase, las podemos separar de acuerdo a los objetos que encontramos en cada una de las pestañas de la interfaz gráfica, tal como se ve a continuación:

- Identification. Esta pestaña está dividida a su vez en las siguientes secciones:
 - Contact Information. Las outlets correspondientes a esta sección son: `cntPosTextField`, `cntPerTextField`, `cntOrgTextField`, `telTextField`, `faxTextField`, `emailTextField`, `countryTextField`, `cityTextField`, `addrTextField`, `addrTypeTextField`, `stateTextField`.

- Identification: Las variables instancia relacionadas a esta pestaña son: abstractTextView, purposeTextView, supInfTextView, statusTextField, keyTextField, accConsTextView, useConsTextView.
 - Cite Information: Las outlets que corresponden a esta parte son: originTableView, originTextField, pubDateTextField, titleTextField, serNameTextField, serIssueTextField.
 - Period Information: Los outlets que intervienen en esta parte son: singDateTextField, singTimeTextField, beginDateTextField, endDateTextField, beginTimeTextField, endTimeTextField.
 - Spatial Domain: Esta sección hace referencia a los siguientes outlets: westBCTextField, eastBCTextField, northBCTextField and southBCTextField.
- Metadata Reference. Para la parte de la referencia de metadatos, los outlets que intervienen son: metaDateTextField, metaStdVerTextField, metaStdNameTextField,

Además de los métodos de acceso, en esta clase tenemos los siguientes dos grupos:

- Métodos para actualización genéricos. Estos métodos son llamados genéricos porque su objetivo es que puedan ser llamados por muchos métodos de actualización específicos. Estos métodos genéricos reciben 2 parámetros: un objeto de control y el valor de la variable. Los métodos genéricos contenidos en MyWindowController son:
 - updateTextField.
 - updateTableView.
 - updateTextView.
- Métodos para actualización específicos. Estos métodos se encargan de actualizar objetos específicos de la interfaz gráfica. Por lo tanto, tenemos tantos métodos de este tipo como el número de objetos a actualizar.
Estos métodos, son invocados por lo que se conoce como notificaciones, las cuales

son registradas por el centro de notificaciones.

El AppKit nos proporciona por defecto, un centro de notificaciones. Este centro puede ser usado para enviar una notificación global (broadcast) a cualquier objeto que esté registrado para recibirla. Estamos haciendo uso del patrón conocido como observer, de acuerdo al cual, un objeto puede tener muchos observadores.

Otros métodos importantes en esta clase son:

- `windowDidLoad`. Es en este método donde, registramos el controlador de ventanas en el centro de notificaciones, y este registro se realiza en el método `windowDidLoad`.
- `registerNotificationObservers` Con este método, damos de alta a los elementos que pueden disparar notificaciones, las cuales serán observadas por el objeto controlador de ventanas ha sido registrado para poder hacerlo.
- `windowWillReturnUndoManager`. Con este método delegado, le indicamos a la ventana que devuelva un gestor de deshacer (undo manager) que nosotros elegimos y no el que es creado por defecto.

MyDocument

La clase `MyDocument` es una subclase de `NSDocument`. Es necesario tener esta subclase, porque así será posible crear una nueva instancia de ella cada vez que el usuario abra un documento ya existente, o decida crear uno nuevo.

Esta clase, es la encargada de controlar la clase modelo de datos de nuestra aplicación, por lo que la ubicamos también en la capa Controlador del paradigma MVC.

`MyDocument` se encargará de crear una instancia de la clase `MySettings` por cada documento que sea creado, y así mismo se hará cargo de que sea liberada de la memoria una vez que el documento se cierre. Por eso creamos una variable instancia llamada `mySettings` y creamos igualmente su respectivo método de acceso, que lleva el mismo nombre.

Otros métodos que se encuentran en esta clase son:

- `makeWindowController`. Este método es el que se encarga de crear los objetos controladores de ventana.

- `setupDictionaryFromMemory`. Este método usa un conjunto de pares llave-valor para configurar un diccionario mutable (mutable dictionary) para que este sirva como un almacenamiento intermedio antes de salvarlo en un medio persistente.
- `writeToFile`. Este método, es el que se encarga de almacenar el objeto diccionario (que creamos con el método `setupDictionaryFromMemory`) al disco y lo almacena en formato XML usando la codificación UTF-8.
- `loadDataRepresentation`. A través de este método obtenemos los datos del disco cuando abrimos un documento, y también cuando ejecutar el comando revertir, del menu de usuario.
- `restoreFromStorage`. Este método es para poder recuperar el diccionario que está en forma de un flujo de bytes y convertirlo de nuevo a un objeto diccionario.
- `setupDictionaryFromStorage`. Este método es el que permite que, el flujo de bytes que recupera el método `restoreFromStorage`, sea convertido a un objeto diccionario.

3.2.5. Modelo en Meta-X

Aquí entramos a la capa Modelo de nuestra aplicación. Esta capa se encuentra formada de varias clases: `CiteInfo`, `CntInfo`, `IdInfo`, `MetaInfo`, `MySettings`.

CiteInfo

Las variables instancia de esta clase contienen los valores que son capturados por los objetos de la interfaz, en la pestaña señalada como Cite Information. En la tabla 3.1 podemos ver el conjunto de variables instancia que componen a esta clase y los métodos que son utilizados para acceder a ellas.

CntInfo

Esta clase contiene las variables necesarias para trabajar con la información de contacto de los metadatos. Sus variables instancia y sus respectivos métodos de acceso son vistos

con más claridad en la tabla 3.2.

TimeInfo

En esta clase tenemos las variables instancia necesarias para almacenar la información que proporcionan los metadatos referente al tiempo. Estas variables se listan en la tabla 3.3. Instancias de esta clase son usadas por la clase IdInfo.

IdInfo

En la siguiente tabla podemos observar las variables instancia que la componen y los métodos de acceso a dichas variables, que contiene. Entre las variables instancia que

VARIABLES INSTANCIA	Método de acceso (recupera)	Método de acceso (asigna)
origin	origin	setOrigin
pubDate	pubDate	setPubDate
title	title	setTitle
serName	serName	setSerName
serIssue	serIssue	setSerIssue

Cuadro 3.1: Variables y métodos de acceso de la clase CiteInfo

VARIABLES INSTANCIA	Método de acceso (recupera)	Método de acceso (asigna)
cntPerson	cntPerson	setCntPerson
cntOrg	cntOrg	setCntOrg
cntPos	cntPos	setCntPos
tel	tel	setTel
fax	fax	setFax
email	email	setEmail
country	country	setCountry
city	city	setCity
addr	addr	serAddr
addrType	addrType	setAddrType
state	state	setState

Cuadro 3.2: Variables y métodos de acceso de CntInfo

conforman esta clase, destacan las variables instancia de tipo TimeInfo, CiteInfo y CntInfo, que son las clases que describimos antes.

MetaInfo

Esta clase posee las variables y los métodos necesarios para contener la información de referencia de los metadatos. Sus variables y respectivos métodos para acceder a ellas se ven en la tabla 3.5. Junto con las variables que se observan en la tabla, la clase MetaInfo también hace uso de una variable instancia de tipo CntInfo, cuya estructura hemos observado ya en la tabla 3.2.

VARIABLES INSTANCIA	Método de acceso (recupera)	Método de acceso (asigna)
singleDate	singleDate	setSingleDate
singleTime	singleTime	setSingleTime
beginDate	beginDate	setBeginDate
endDate	endDate	setEndDate
beginTime	beginTime	setBeginTime
endTime	endTime	setEndTime

Cuadro 3.3: Variables y métodos de acceso de TimeInfo

VARIABLES INSTANCIA	Método de acceso (recupera)	Método de acceso (asigna)
abstract	abstract	setAbstract
purpose	purpose	setPurpose
supInf	supInf	setSupInf
status	status	setStatus
keyword	keyword	setKeyword
accConst	accConst	setAccConst
useConst	useConst	setUseConst
westBC	westBC	setWestBC
eastBC	eastBC	setEastBC
northBC	northBC	setNorthBC
southBC	southBC	setSouthBC

Cuadro 3.4: Variables y métodos de acceso de la clase IdInfo

Variabes Instancia	Método de acceso (recupera)	Método de acceso (asigna)
metaDate	metaDate	setMetaDate
metaStdName	metaStdName	setMetaStdName
metaStdVer	metaStdVer	setMetaStdVer

Cuadro 3.5: Variables y métodos de acceso de la clase MetaInfo

MySettings

Esta clase la podemos considerar como la clase principal del modelo, ya que es en esta donde se llevan acabo varias acciones básicas de nuestra aplicación.

Esta clase contiene variables instancia del tipo MyDocument, IdInfo y MetaInfo. Y cada una de estas variables tiene también, como en las otras clases que describimos con anterioridad, métodos para recuperar y asignar valores a ellas (métodos de acceso).

Al igual que el resto de las clases del modelo, esta clase hereda directamente de NSObject, y contiene ya un método para inicializar una instancia, llamado init. Sin embargo, este método sólo devuelve self. Lo cual no es suficiente para dar un estado inicial a todas las variable, razón por la cual creamos otro método llamado initWithDocument. Este método es el inicializador designado, y es el encargado de asignar valores iniciales a cada una de las variables que forman parte de la clase. En otras palabras, establece el estado inicial completo del objeto.

Además de los métodos de inicialización, tenemos los siguientes:

- undoManager. Este método es necesario para que MySettings pueda comunicarse con el gestor para deshacer (undo manager) asociado con el documento.
- convertToDictionary. Este método nos sirve para que, en caso de que aumentáramos variables en la clase MySettings, sólo tengamos que añadir estas variables a este diccionario y una vez aquí, se pasa el diccionario al objeto documento.
- restoreFromDictionary. Sirve para recuperar la información almacenada en el diccionario creado por el método convertToDictionary.

Capítulo 4

Caso de Estudio

En este capítulo establecemos a modo de caso de estudio, proponemos primero un modelo conceptual del repositorio de metadatos, con la intención de que sirva como base para la posterior implementación del mismo en el sistema de gestión de base de datos PostgreSQL. En la segunda parte de este capítulo, hacemos uso de nuestra aplicación Meta-X para capturar algunos metadatos.

4.1. Modelo conceptual

Para llegar a la etapa de diseño físico de un sistema de bases de datos, primero tenemos que pasar por el diseño conceptual y después por el diseño lógico. En esta sección, hacemos una propuesta que aplica a la primera parte de este proceso de desarrollo, es decir, el diseño conceptual. Para lo cual, hicimos uso de la metodología entidad-relación [17] .

4.1.1. Entidad - Relación

Los elementos principales en esta metodología son: entidades, relaciones y atributos.

- Entidades. Son una abstracción de objetos del mundo real acerca de los cuales deseamos capturar información. Son los elementos principales de la metodología.
- Relaciones. Son las asociaciones que existen entre una o varias entidades.

- Atributos. Son las características o propiedades describen a las entidades y a las relaciones.

La notación de los diagramas que realizamos sigue el enfoque conocido como “pata del cuervo” (crow’s foot). La figura 4.1 muestra los símbolos que conforman dicho enfoque.

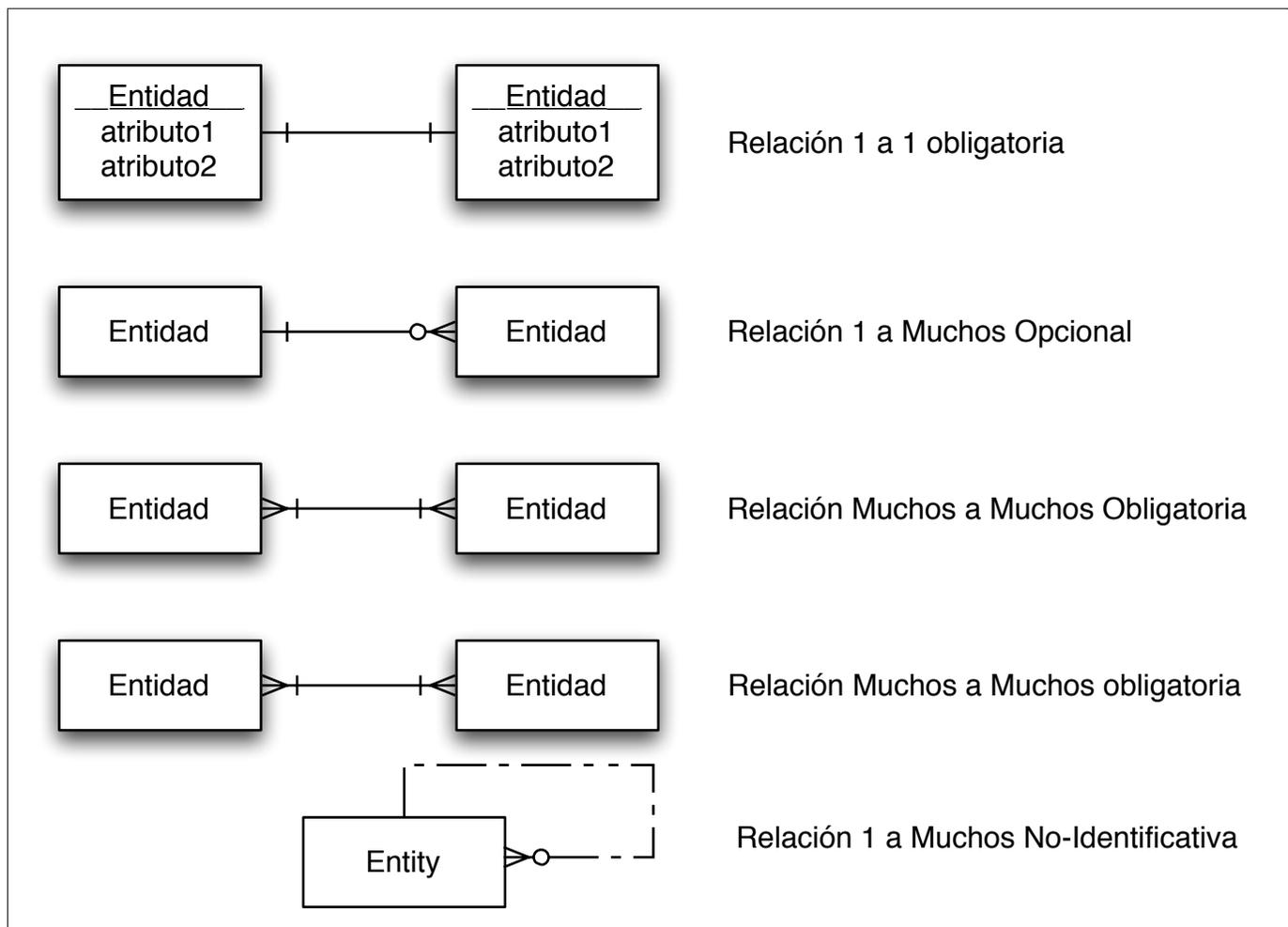


Figura 4.1: Notación “Crow’s Foot” del Modelo ER

4.1.2. Repositorio de Metadatos

El repositorio de metadatos lo podemos dividir de forma general en 3 grupos de entidades. El primer grupo se conforma de dos entidades: Colección y Metadato, las cuales nos sirven para identificar a un conjunto de metadatos almacenado y, para identificar a cada metadato dentro de una colección. El segundo y tercer grupo de entidades se basan en la información que nos proporciona el estándar para metadatos que estudiamos en capítulos anteriores.

Lo anterior nos lleva a tener 7 entidades principales y 3 entidades de apoyo. A continuación explicamos las distintas entidades que componen este modelo.

- Colección (Collection): Esta entidad tiene como objetivo identificar contiene 2 atributos: nombre e identificador del conjunto de metadatos que será almacenado.
- Metadato (Metadata): Es la entidad que identifica al metadato dentro de la colección. A partir de esta entidad se relacionan el resto de las entidades que contendrán la información de cada metadato. Dichas entidades las listamos a continuación:
- Entidades principales:
 1. IdInfo. Contiene la información básica del conjunto de datos. Se relaciona de manera uno a uno, y su existencia es obligatoria.
 2. DataQual. Es la entidad que representa la sección de Calidad de los Datos. Da una explicación general acerca de la calidad del conjunto de datos que se va a almacenar.
 3. SpatialDataOrg. Organización de los Datos Geoespaciales es la sección representada por esta entidad. La entidad Metadata se relaciona de manera uno a uno con existencia opcional del lado de SpatialDataOrg.
 4. SpatialRef. Es la descripción del marco de referencia para codificar los datos y las coordenadas del conjunto de datos. Se encargará de guardar la información de Referencia Espacial.

5. Entidad y Atributo (EntityAttribute). Proporciona detalles del conjunto de datos, tales como, tipos de entidades, sus atributos, y los dominios de los cuales estos atributos pueden tomar sus valores.
 6. Distribución (Distribution). Información acerca del distribuidor y de las opciones que hay para obtener los datos. Se relaciona en forma muchos a uno con la entidad Metadata. La existencia de la entidad Distribution es opcional.
 7. MetaInfo. Contiene la información del estado actual de la información del metadato y la parte responsable. Esta entidad tiene una relación uno a uno con la entidad principal Metadata, y la existencia de MetaInfo es obligatoria.
- Entidades de apoyo:
 1. CiteInfo. Es la referencia recomendada para utilizarse con el conjunto de datos.
 2. TimeInfo. Información acerca de la fecha y hora de un evento.
 3. Contact. Proporciona la identidad y los medios para comunicarse con la(s) persona(s) y organizaciones asociadas con el conjunto de datos.

En la figura 4.2 mostramos el esquema general de nuestro modelo.

De las 7 entidades principales que observamos en el diagrama general de la figura 4.2, solo desarrollamos las secciones de Información (IdInfo) y la de referencia de metadatos, además

Submodelo ER para la información de Identificación

La entidad principal en esta sección es IdInfo. Contiene los siguientes atributos:

Mientras que puede contener opcionalmente la siguiente información:

- Punto de Contacto
- Exploración de gráficas
- Créditos del Conjunto de Datos

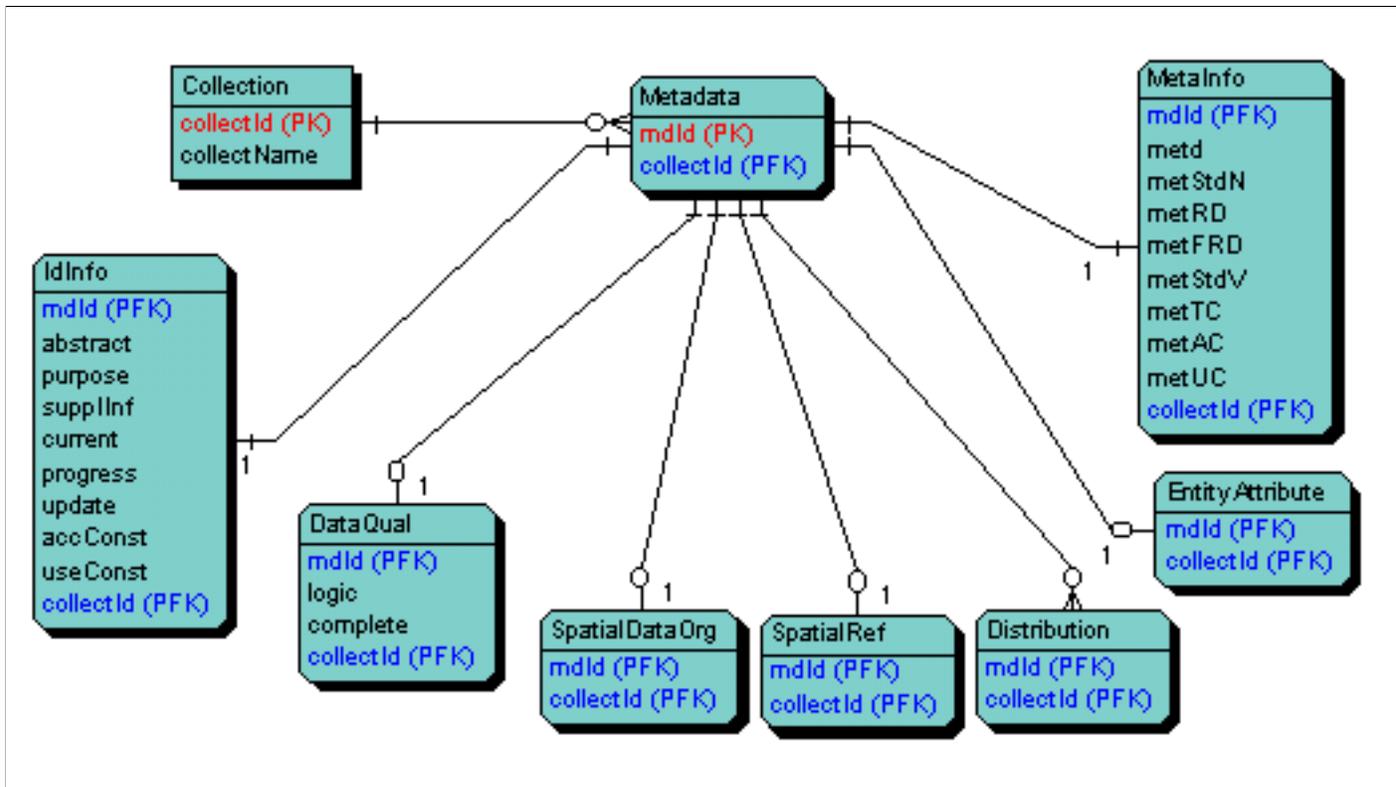


Figura 4.2: Diagrama ER General del Repositorio de Metadatos

- Información de seguridad
- Ambiente nativo del conjunto de datos.
- Referencia cruzada.

El esquema ER de esta sección lo mostramos en la figura 4.3

Submodelo ER para la información de Referencia del Metadato

Las entidades que intervienen para modelar esta sección son:

- MetaInfo. Esta es la entidad principal de este submodelo. Contiene los atributos que son de carácter obligatorio. Como son, nombre del estándar, fecha del estándar y la versión entre otros. Se relaciona con las siguientes entidades:
- MetSI. Esta entidad está destinada a almacenar información relacionada con la seguridad. Su existencia es opcional. y se relaciona de uno a uno con la entidad MetaInfo.
- MetExtns. Esta entidad contendrá en sus atributos referencias a otros documentos de interés.

Y es necesario que se relacione con las entidades del submodelo de contacto, el cual lo explicaremos más adelante. El submodelo ER para esta sección se observa en la figura 4.4

Submodelo ER para la información de Cita

En el submodelo correspondiente a la información de la cita, intervienen las entidades que se listan a continuación.

- CiteInfo. Esta es la entidad principal de este submodelo, sus atributos contienen información como el originador del conjunto de datos, el título del mismo, la edición, la dirección electrónica para acceder, el nombre de la publicación serial y el

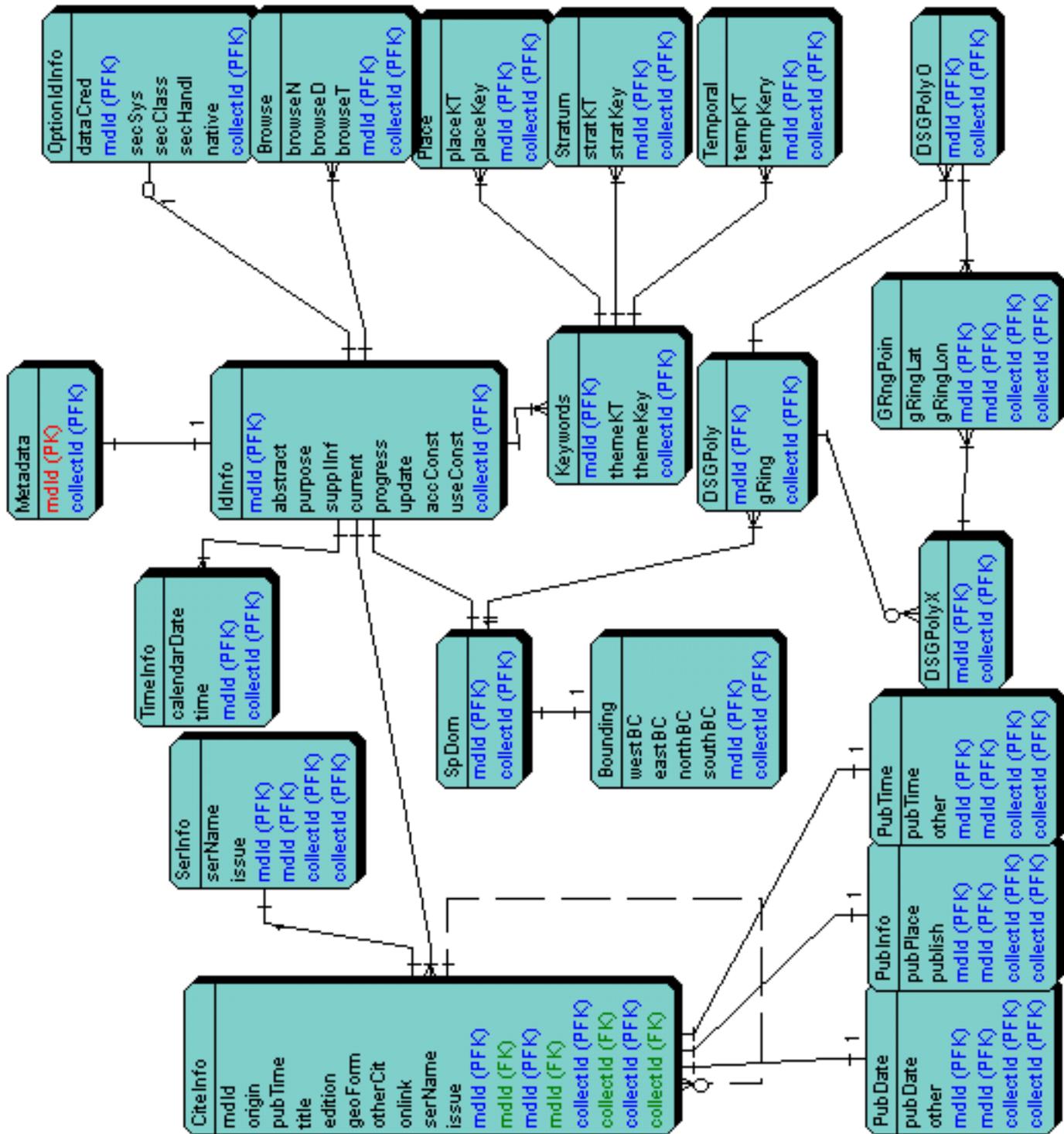


Figura 4.3: Submodelo ER de la sección de identificación

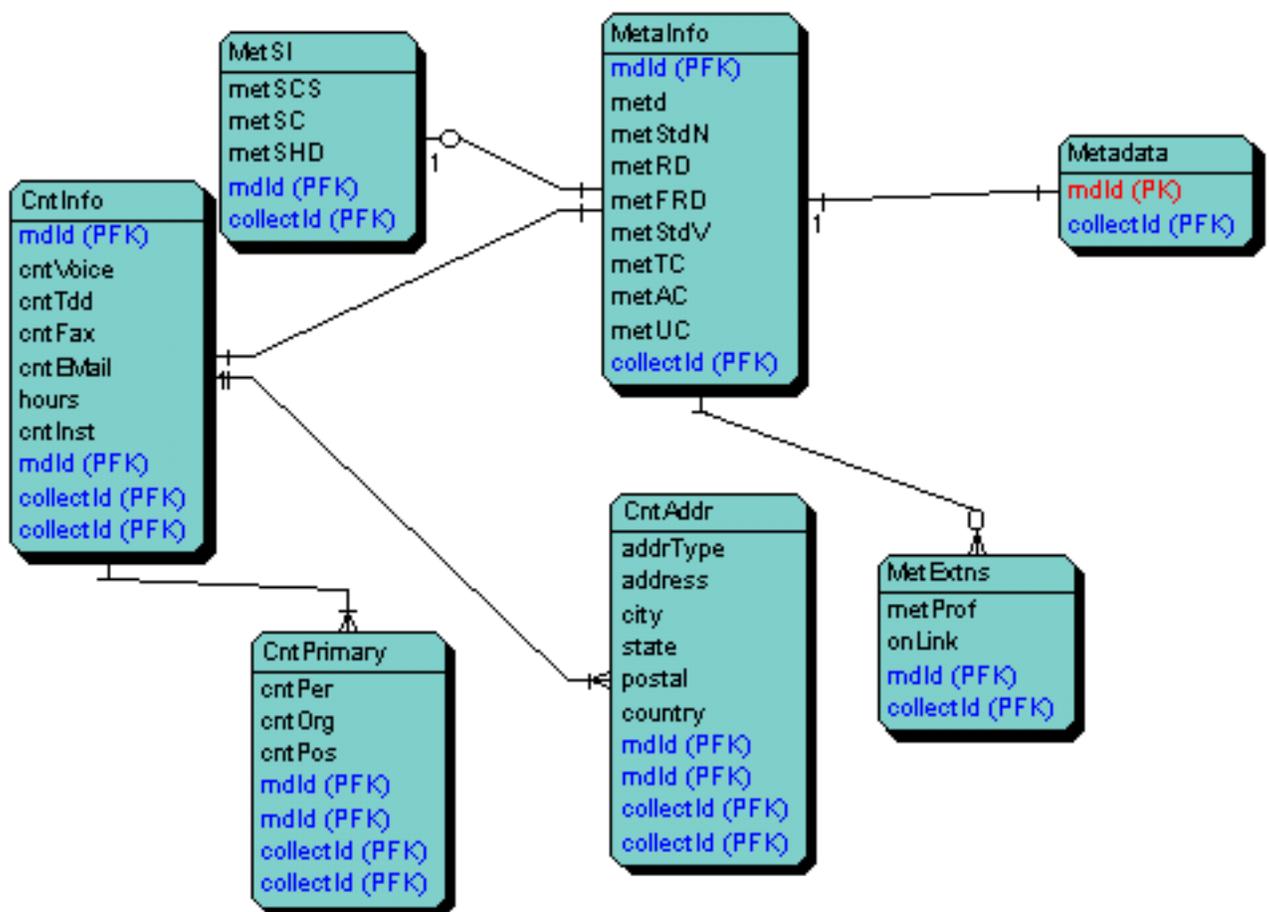


Figura 4.4: Submodelo ER de la sección de Información de Referencia del Metadato

ejemplar de la misma. Esta tiene una relación uno a uno con las entidades PubTime y PubDate. Y tiene una relación unaria de uno a muchos, la cual es de carácter opcional, porque solo se utilizará en caso de que una instancia de cita, contenga más instancias de este mismo tipo.

- PubDate. Aunque es de carácter obligatorio se optó por hacerla entidad dado, que en caso de no existir fecha, puede tomar un valor distinto al formato de fecha.
- PubTime. Al igual que en la entidad anterior, esta se separó en otra entidad porque aparte de poder tomar un valor de tiempo, si este no existiera, entonces se especificaría otro valor distinto, que no necesariamente podría representarse con un nulo en la base de datos física.

El modelo entidad-relación es el que se observa en la figura 4.5.

Submodelo ER para la información de Contacto

- CntInfo. Esta es la entidad principal en este modelo, y está destinada a almacenar la información para poder contactar al creador del conjunto de datos que se esté trabajando.
- CntPrimary. Esta entidad está relacionada de forma muchos a uno con respecto a CntInfo, porque cabe la posibilidad de que exista más de un contacto, ya sea persona u organización.
- CntAddr. Esta entidad también se relaciona de manera muchos a uno con CntInfo, esto porque puede haber más de una dirección de contacto.

Así el submodelo resultante para esta sección de contacto se observa en la figura 4.6.

4.2. Caso de estudio de Meta-X

En esta sección, utilizaremos la aplicación que desarrollamos para capturar la información concerniente a varios metadatos.

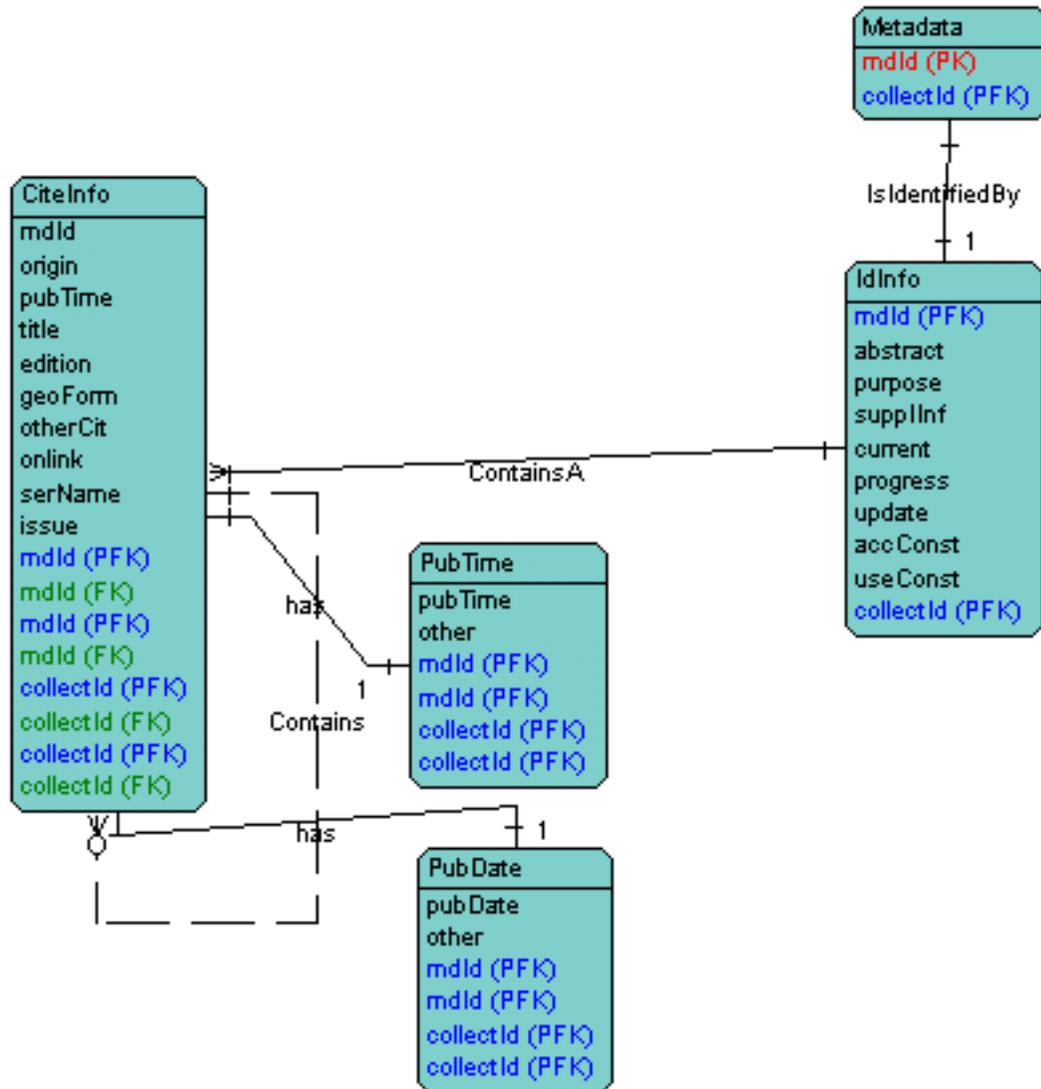


Figura 4.5: Submodelo ER de la sección de Información de Cita

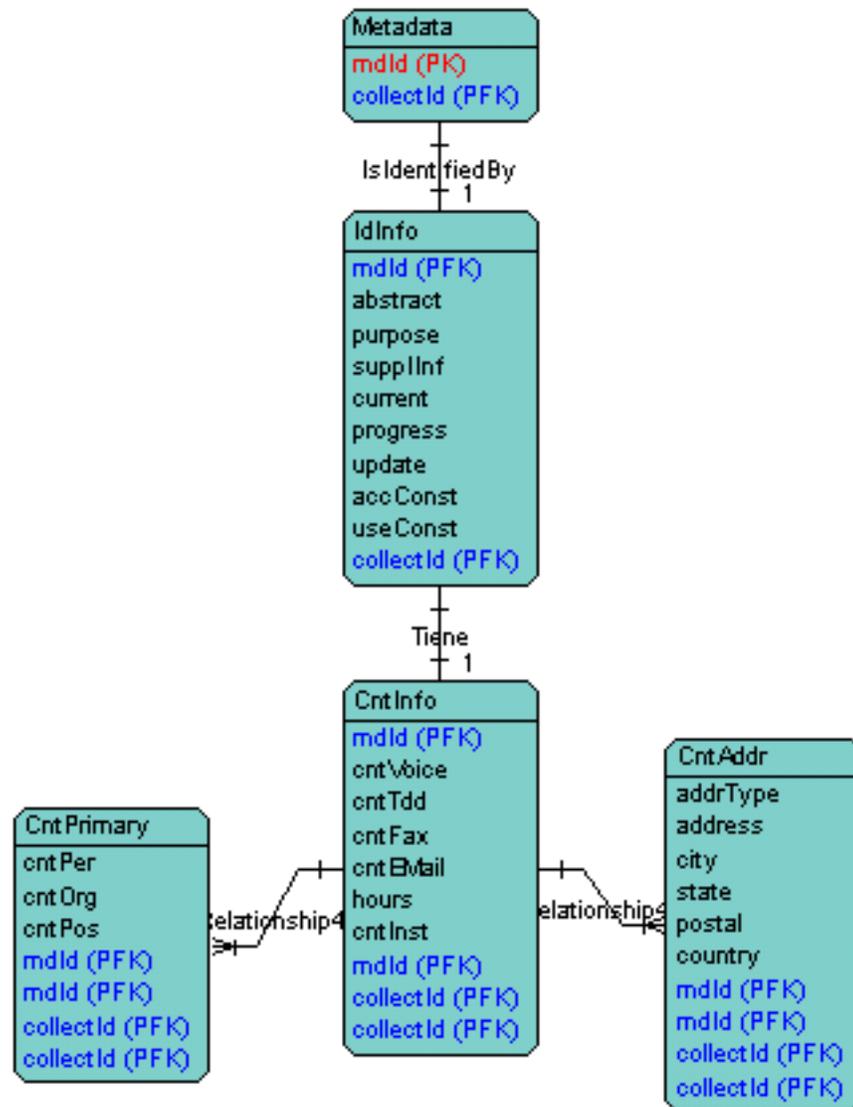


Figura 4.6: Submodelo ER de la sección de Información de Contacto

4.2.1. INEGI

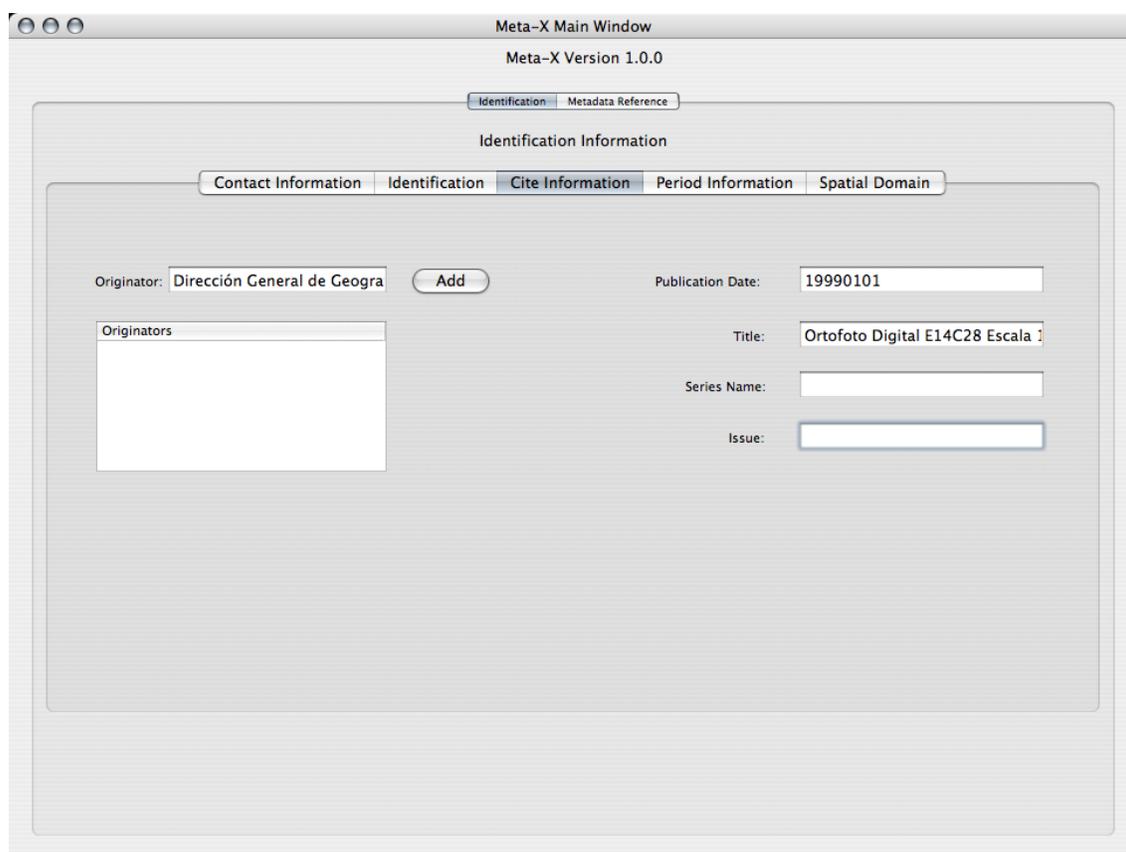
El primer caso lo proporciona el INEGI (Instituto Nacional de Estadística, Geografía e Informática). Y nos da la siguiente información:

Información de Identificación

- Información de cita:
 - Originador: Dirección General de Geografía
 - Fecha Publicación: 19990101
 - Tiempo de Publicación: Desconocido
 - Título: Ortofoto Digital E14C28 Escala 1:20,000 (CHILPANCINGO) Ortofoto(s)
 - Edición: 1
 - Forma de Presentación de Datos Geoespaciales: Imagen de Percepción remota
 - Lugar de Publicación: Aguascalientes, Ags., México
 - Editor: INEGI
 - Liga en Línea: <http://www.inegi.gob.mx/>

La información anterior se observa en la imagen 4.7 de la pantalla de captura de la información de cita.

- Descripción.
 - Resumen: Las Ortofotos combinan las características de la imagen de una fotografía aérea de cualquier escala con las cualidades geométricas de un mapa. Son imágenes de las que se han eliminado las deformaciones causadas por la óptica de la cámara y por el desplazamiento aparente de los objetos del terreno en el momento de la toma fotográfica.



The screenshot displays the 'Meta-X Main Window' interface, version 1.0.0. The window title bar includes standard OS window controls. The main content area is titled 'Identification Information' and features a tabbed interface with the following tabs: 'Identification', 'Metadata Reference', 'Contact Information', 'Cite Information' (selected), 'Period Information', and 'Spatial Domain'. The 'Cite Information' tab contains several input fields and a list:

- Originator:** A text field containing 'Dirección General de Geogra' and an 'Add' button.
- Originators:** An empty list box.
- Publication Date:** A text field containing '19990101'.
- Title:** A text field containing 'Ortofoto Digital E14C28 Escala :'.
- Series Name:** An empty text field.
- Issue:** An empty text field.

Figura 4.7: Vista de la pantalla de captura de información de cita

- Propósito: Se pueden usar para una gran variedad de propósitos, que van desde la generación de mapas intermedios hasta su uso como referencia de campo para investigaciones y análisis de las ciencias de la tierra, como son recursos naturales (erosión, agua, bosques), en agricultura de riego, de temporal; en infraestructura, tal como aeropuertos, presas, carreteras; en urbanismo vial, urbanismo industrial, urbanismo recreativo, urbanismo habitacional. La Ortofoto en su modalidad digital es útil como cubierta en los Sistemas de Información Geográfica (SIG), así mismo son una herramienta muy útil en la actualización de gráficos a línea y mapas topográficos (cartografía).
- Información Suplementaria: Las Ortofotos Digitales, por estar georreferidas, es posible ubicar y determinar las dimensiones de los rasgos visibles en ellas. Las Ortofotos están producidas con una resolución en el terreno de 2 metros por pixel. La dimensión de cada una es de 6 minutos 40 segundos de longitud por 7 minutos 30 segundos de latitud. Este formato está diseñado para que al ensamblar seis Ortofotos escala 1 a 20,000 (A, B, C, D, E y F) se obtenga la cobertura de una carta topográfica 1 a 50,000. Tienen la Proyección Universal Transversa de Mercator (UTM). Datum NAD27 o ITRF92. Archivo binario, crudo, con extensión .bil. Archivos de texto con metadatos, con extensión .txt con dimensiones de la Ortofoto, proyección, datum, georreferencia y fecha de toma de la foto, archivos "World", como son .hdr, .blw, .prj. 45 MB son el volumen promedio de datos por Ortofoto. La cobertura de todo el país es más de 12,000 Ortofotos. Seis Ortofotos cubren el área de una hoja escala 1 a 50,000. La clave de la Ortofoto se forma con la nomenclatura de la hoja escala 1 a 50,000 en la que está incluida y una letra de la A a la F. Las Ortofotos pueden ser distribuidas a través de cinta de 8mm, disco compacto o vía internet (ftp). Fechas de toma de las fotos, de febrero de 1995 a marzo de 1995.

En la figura 4.8 observamos la captura de los campos de descripción de los metadatos.

- Período de Tiempo de Contenido:

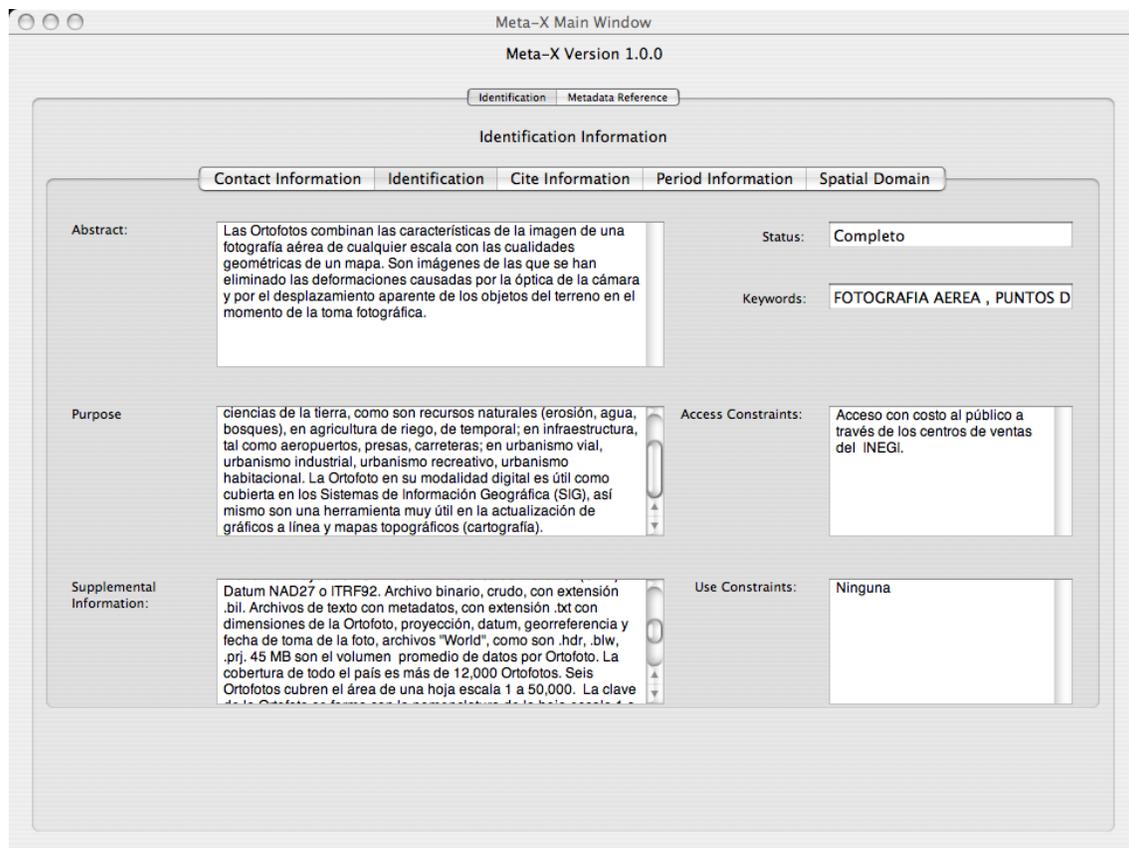


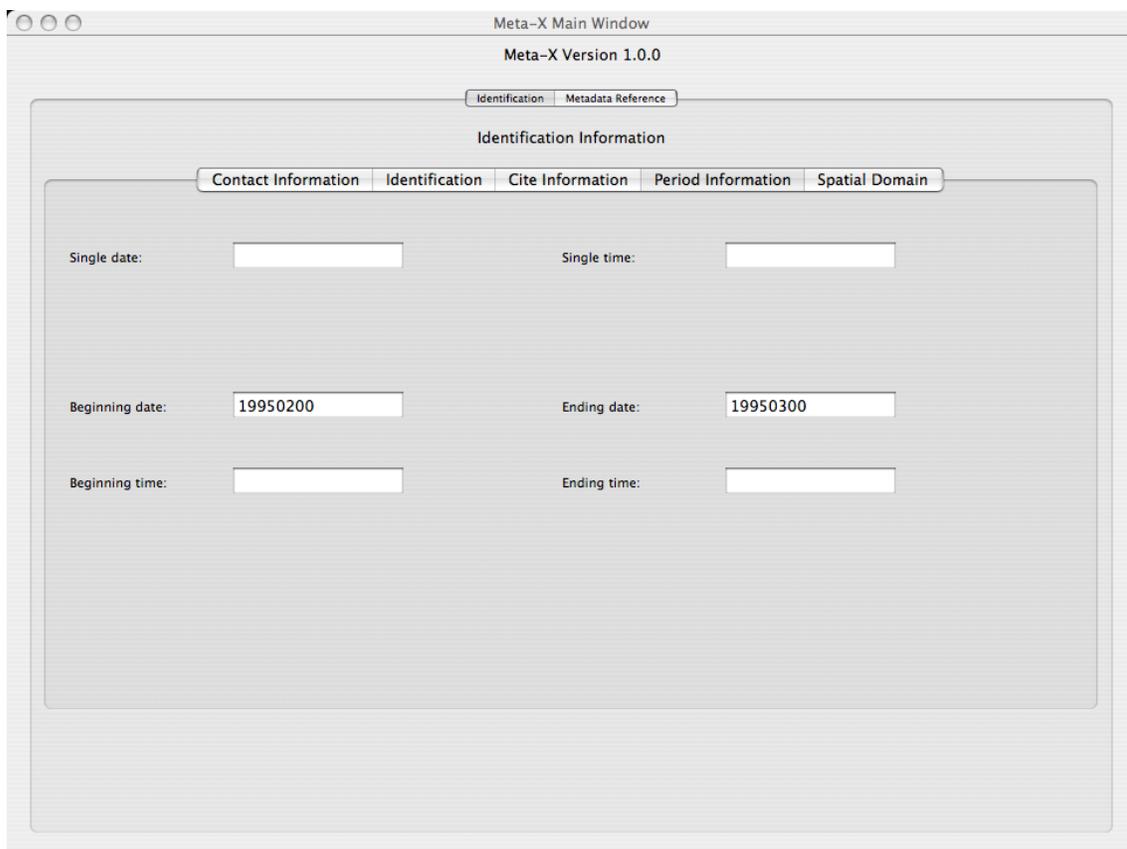
Figura 4.8: Vista de la pantalla de captura de la descripción de metadatos

Información de Período de Tiempo:

Intervalo de Fechas/Tiempos:

- Fecha de Inicio: 19950200
- Fecha Final: 19950300

La información sobre periodo, o de tiempo de los metadatos también es capturada en la interfaz gráfica, como solo existe un intervalo de fechas, entonces solo esos campos son llenados. En la figura 4.9 podemos observar lo dicho.



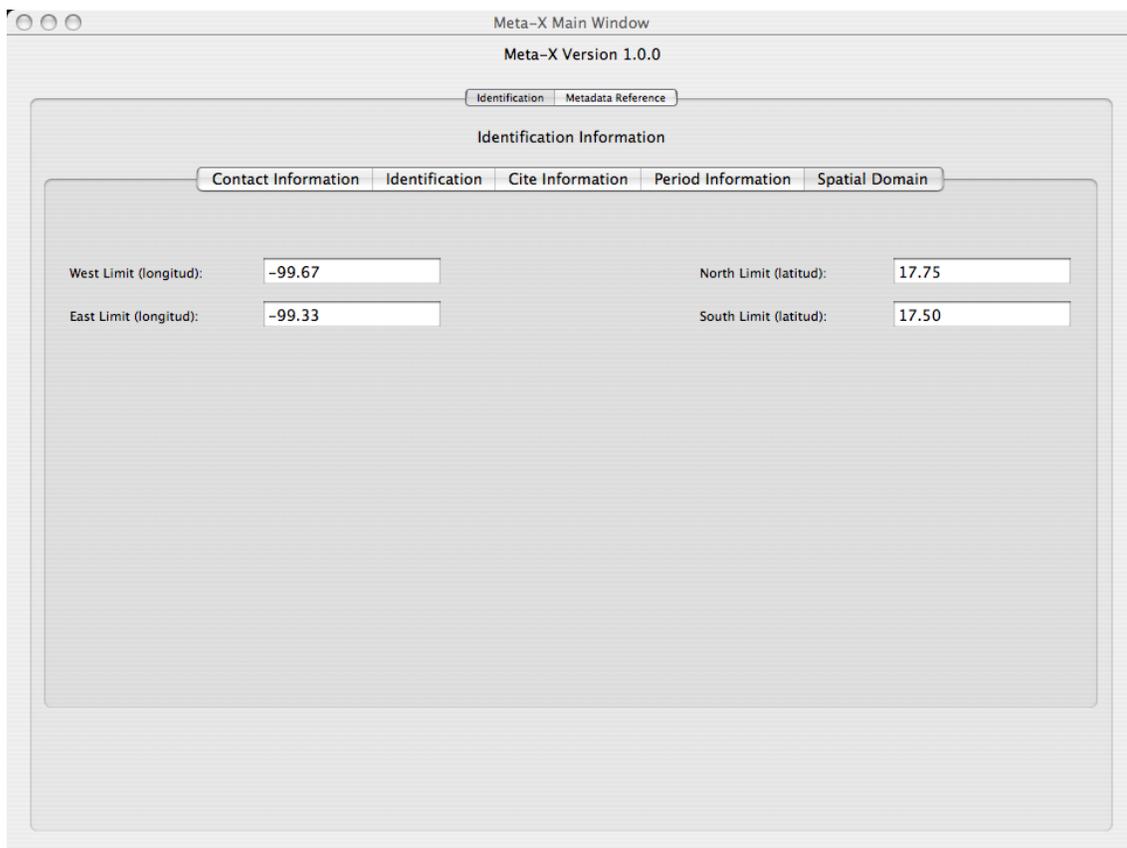
The screenshot shows a window titled "Meta-X Main Window" with the version "Meta-X Version 1.0.0". The main content area is divided into two tabs: "Identification" (selected) and "Metadata Reference". Under the "Identification" tab, there is a sub-section titled "Identification Information" with five sub-tabs: "Contact Information", "Identification", "Cite Information", "Period Information" (selected), and "Spatial Domain". The "Period Information" sub-tab contains six input fields arranged in a 3x2 grid:

Single date:	<input type="text"/>	Single time:	<input type="text"/>
Beginning date:	<input type="text" value="19950200"/>	Ending date:	<input type="text" value="19950300"/>
Beginning time:	<input type="text"/>	Ending time:	<input type="text"/>

Figura 4.9: Vista de la pantalla de captura de la descripción de metadatos

- Estado

- Progreso: Completo
 - Frecuencia de Mantenimiento y Actualización: Desconocida
- Dominio Espacial. Esta información es capturada en la pestaña correspondiente Spatial Domain, de la interfaz gráfica, tal como se observa en la figura 4.10
- Coordenada Límite Oeste: -99.67
 - Coordenada Límite Este: -99.33
 - Coordenada Límite Norte: 17.75
 - Coordenada Límite Sur: 17.50



The screenshot shows the 'Meta-X Main Window' interface. At the top, it displays 'Meta-X Version 1.0.0'. Below this, there are two tabs: 'Identification' (selected) and 'Metadata Reference'. Under the 'Identification' tab, there is a sub-tabbed area with five options: 'Contact Information', 'Identification', 'Cite Information', 'Period Information', and 'Spatial Domain'. The 'Spatial Domain' tab is active, showing four input fields for spatial limits:

West Limit (longitud):	<input type="text" value="-99.67"/>	North Limit (latitud):	<input type="text" value="17.75"/>
East Limit (longitud):	<input type="text" value="-99.33"/>	South Limit (latitud):	<input type="text" value="17.50"/>

Figura 4.10: Vista de la pantalla de captura de información de Dominio Espacial

- Palabras clave: FOTOGRAFIA AEREA , PUNTOS DE CONTROL
- Restricciones de Acceso: Acceso con costo al público a través de los centros de ventas del INEGI.
- Restricciones de Uso: Ninguna

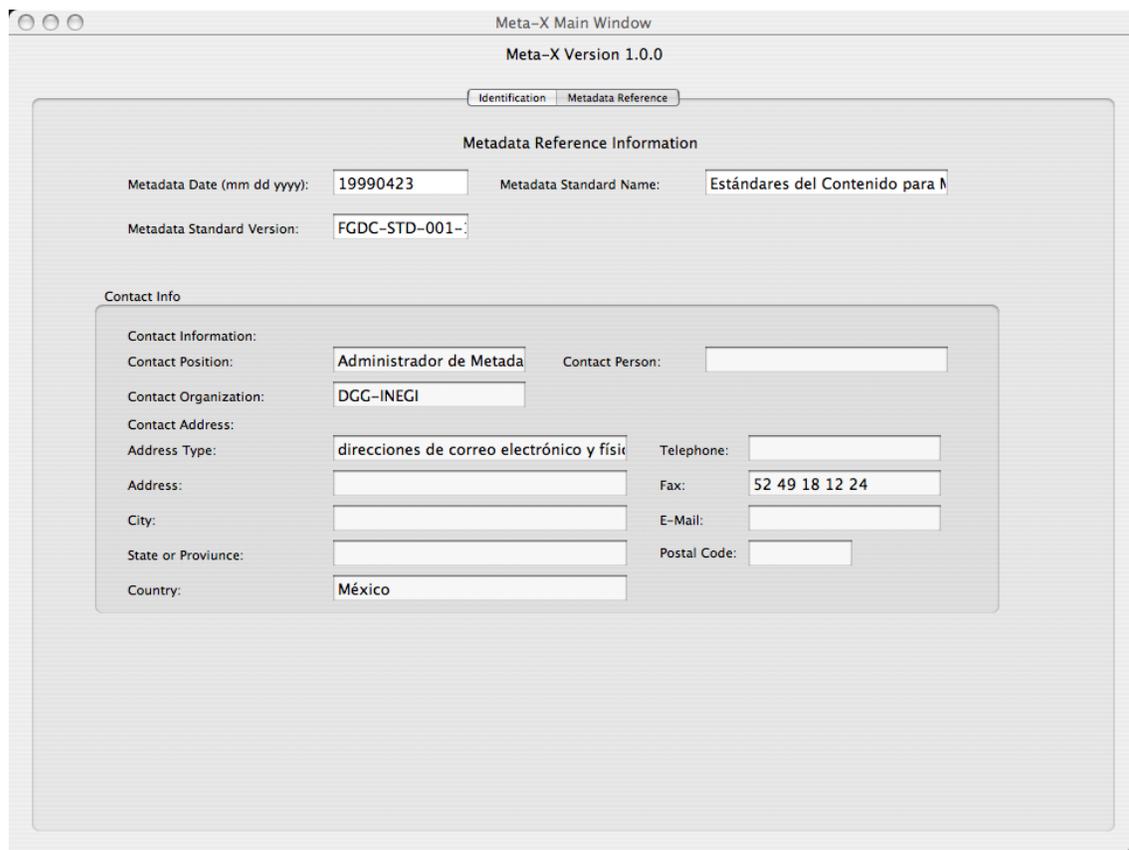
Información de Referencia de Metadatos

En el documento del INEGI se nos proporciona la siguiente información:

- Fecha de Metadatos: 19990423
- Fecha de Revisión de Metadatos: 19990608
- Fecha de Revisión de Características de Metadatos: 19990608
- Contacto de Metadatos:
Información de Contacto:
Principal Persona de Contacto:
 - Persona: Administrador de Metadatos
 - Organización: DGG-INEGI
 - Tipo de Dirección: direcciones de correo electrónico y física
 - Dirección: Av. Héroe de Nacozari # 2301 Sur. Fracc. Jardines del Parque
 - Ciudad: Aguascalientes
 - Estado o Municipio: Aguascalientes
 - Código Postal: 20270
 - País: México
 - Teléfono: 52 49 10 53 00 Ext. 5856
 - Facsimil: 52 49 18 12 24
 - Dirección de Correo Electrónico: rarioja@dgg.inegi.gob.mx

- Nombre del Estándar de Metadatos: Estándares del Contenido para Metadatos Geoespaciales Digitales del FGDC
- Versión del Estándar de Metadatos: FGDC-STD-001-1998

De esta información , se capturó la siguiente usando Meta-X. En la figura 4.11 podemos observar la apariencia de la interface para capturar la información del contacto y de la referencia de metadatos.



The screenshot displays the 'Meta-X Main Window' interface, version 1.0.0. It features two tabs: 'Identification' and 'Metadata Reference'. The 'Metadata Reference Information' section contains the following fields:

- Metadata Date (mm dd yyyy): 19990423
- Metadata Standard Name: Estándares del Contenido para M
- Metadata Standard Version: FGDC-STD-001-

The 'Contact Info' section is enclosed in a rounded rectangle and includes the following fields:

- Contact Information:
- Contact Position: Administrador de Metada
- Contact Person:
- Contact Organization: DGG-INEGI
- Contact Address:
- Address Type: direcciones de correo electrónico y físic
- Address:
- City:
- State or Province:
- Country: México
- Telephone:
- Fax: 52 49 18 12 24
- E-Mail:
- Postal Code:

Figura 4.11: Vista de la pantalla de captura de información de contacto

Capítulo 5

Conclusiones y Trabajo Futuro

5.1. Conclusiones

A lo largo de esta tesis hemos visto que trabajar con metadatos no resulta algo trivial, esto se ve claramente desde la propia definición de los mismos, la cual como vimos al inicio, varía de un ambiente a otro y, la definición más común de ellos es, información acerca de la información, esta los hace parecer más simples de lo que son realmente.

La complejidad de trabajar con metadatos radica, en determinar que información es importante capturar acerca de los datos con los que estamos trabajando. Tan es así, que han surgido y se siguen desarrollando estándares que ayuden a establecer la manera en la que los metadatos serán tratados, es decir, la manera en la que trabajaremos con ellos, que información contendrán, y dos de los puntos más importantes, la manera en que debemos crear y compartir esos metadatos con la comunidad.

Es en estos dos últimos aspectos a los que traté de enfocarme al desarrollar esta tesis. Para empezar, decidí trabajar en base a un estándar, porque creo q es la mejor manera de que podamos trabajar con otros grupos o comunidades interesadas en metadatos, hablando en los mismos términos.

5.1.1. Aportaciones

En lo que respecta a la creación de los metadatos, decidimos que era necesaria una herramienta que trabajara sobre la plataforma Mac OS X, ya que actualmente no hay alguna herramienta que trabaje en esta plataforma. Además, nuestra herramienta salva los metadatos capturados en formato XML, lo cual hace más factible que sea compartido con la comunidad vía internet.

Cabe destacar también que, Meta-X no solo fue ideada para la creación de metadatos, sino que participará en otras facetas del metadiseño de sistemas de bases de datos, como son la recolección de requerimientos y la edición de reglas semánticas.

En lo que respecta a la arquitectura propuesta, tenemos un conjunto de repositorios que almacenarán distinta información relacionada con un proyecto, esta tiene la ventaja de que al tener bases de datos separadas para almacenar los distintos módulos de un proyecto, podemos reutilizar la información que nos ofrecen los metadatos para futuros trabajos. Evitamos de este modo la duplicidad de la información, es decir, volver a capturar los datos; además de que almacenar los archivos en formato XML, nos da la oportunidad de compartirlos con el resto del mundo al publicarlos en internet.

Nuestra herramienta fue desarrollada en C-Objetivo para Mac OS X pensando en que interactuará con otras aplicaciones que trabajan sobre la misma plataforma, formando parte del proyecto para CONACyT llamado CADBD dirigido por el Dr. Sergio Chapa Vergara.

5.2. Trabajo Futuro

En lo que respecta a nuestro sistema Meta-X, existen varios puntos que se pueden desarrollar:

- Módulo de reglas semánticas. Para este módulo, se puede extender Meta-X para capturar la información que se requiera para estas reglas, que pueden seguir el paradigma evento-condición-acción. Para su análisis, podría adaptarse una herramienta llamada Mx-Graf la cual trabaja con grafos.

- Módulo de requerimientos. Se puede extender Meta-X creando un conjunto de formas para capturar los metadatos clasificándolos en distintos tipos como : funcionales, no-funcionales. Un documento que resulta de interés para esta sección es el de Prácticas Recomendadas de la IEEE para Especificaciones de Requerimientos de Software [12].
- Repositorios. En lo concerniente a los repositorios falta por implementar el modelo conceptual para metadatos, y realizar un desarrollo de los otros repositorios: requerimientos y reglas semánticas. Se planea que dicha implementación se haga sobre el sistema de gestión de bases de datos PostgreSQL.

Finalmente, yendo a un punto más ambicioso, una vez que tengamos implementado el repositorio de metadatos, basado en el estándar de contenido para metadatos, será posible seguir el trabajo por esta línea para poder llegar a implantar un Clearinghouse de nuestro país.

Bibliografía

- [1] Scott Anguish, Erik M. Buck, and Donald A. Yacktman. *Cocoa Programming*. Sams Publishing, 2002.
- [2] Amy Brand, Frank Daly, and Barbara Meyers. *Metadata Demystified: A Guide for Publishers*. Sheridan Press & NISO Press, July 2003.
- [3] Federal Geographic Data Committee. Fgdc-std-001-1998. content standard for digital geospatial metadata, June 1998.
- [4] User Guide Committee, Gail Clement, and Pete Winn. Dcmi glossary. <http://dublincore.org/documents/usageguide/glossary.shtml>, May 2005.
- [5] Adoración de Miguel and Mario Piattini. *Fundamentos y Modelos de Bases de Datos*. Alfaomega, 2da. edition, 1999.
- [6] Kenneth E. Foote and Margaret Lynch. Geographic information systems as an integrating technology: Context, concepts, and definitions. <http://www.colorado.edu/geography/gcraft/notes/intro/intro.html>, 2000.
- [7] NISO Framework Advisory Group. A framework of guidance for building good digital collections, 2004.
- [8] David Hart and Hugh Phillips. Metadata primer - a "how to" guide on metadata implementation. <http://www.lic.wisc.edu/metadata/metaprim.htm>, June 1998.
- [9] Apple Computer Inc. *Learning Cocoa*. O'Reilly, 2001.

-
- [10] Department of Commerce. Spatial data transfer standard (sdts), November 1997.
 - [11] Department of Defense. Vector product format (mil-std-600006), 1992.
 - [12] Software Engineering Standards Committee of the IEEE Computer Society. Ieee recommended practice for software requirements specifications, 1998.
 - [13] National Information Standard Organization. Understanding metadata. 4733 Bethesda Avenue, Suite 300. Bethesda, MD 20814 USA, 2004.
 - [14] Sudha Ram, Jinsoo Park, and George L. Ball. Semantic-model support for geographic information systems. *Computer*, 32(5):74–81, 1999.
 - [15] Suzanne Robertson and James Robertson. *Mastering the Requirements Process*. Addison-Wesley, 1999.
 - [16] Vijay P. Singh and M. Fiorentino. *Geographical Information Systems in Hydrology*. Kluwer Academic Publishers, 1996.
 - [17] Toby J. Teorey. *Database Modeling and Design*. Morgan Kaufmann Publishers, Inc., 1990.