

Desarrollo de Sistemas de Información bio-climática

Rodrigo Torréns Heeren

torrens@cecalc.ula.ve

Tutores:

Magdiel Ablan

mablan@ula.ve

Jacinto A. Dávila Quintero

jacinto@ula.ve

**Trabajo de Grado Presentado ante la ilustre Universidad de Los
Andes como requisito final para optar al título de:
MAGISTER SCIENTIAE EN COMPUTACIÓN**

**UNIVERSIDAD DE LOS ANDES
CONSEJO DE ESTUDIOS DE POSTGRADO
FACULTAD DE INGENIERÍA
Postgrado de Computación**

Julio 2003

Resumen:

La presente investigación trata sobre sistemas de información ecológicos, biológicos y ambientales, haciendo especial énfasis en el manejo de datos y metadatos que permiten formar colecciones de datos científicos. En este trabajo se realizó un estudio de las tecnologías, estándares y herramientas usadas en el manejo de colecciones de datos científicos. En particular, se estudiaron los diferentes estándares de manejo de información estructurada, haciendo énfasis en el lenguaje *EML* como estándar de intercambio y representación de datos. *EML* esta siendo desarrollado por varias instituciones en los Estados Unidos, y se esta comenzando a usar dentro de diferentes comunidades científicas. Como resultado se presentan una serie de lineamientos y recomendaciones para el desarrollo de sistemas de información ecológicos, biológicos y ambientales. Dichos lineamientos fueron usados en la implementación de un prototipo de sistema de información para manejo de colecciones de datos que pretende servir como base de desarrollo para sistemas de este tipo. El prototipo final implementado permite crear y consultar colecciones de datos, concentrándose en documentar los datos, a través de la creación de archivos de metadatos conformes al estándar *EML*. Se espera usar esta investigación como punto de partida de otros desarrollos en esta área

Palabras clave: Sistemas de Información, metadatos, colecciones de datos científicos, *EML*.

Tabla de Contenido

DESARROLLO DE SISTEMAS DE INFORMACIÓN BIO-CLIMÁTICA.....	1
1 MANEJO DE DATOS ECOLÓGICOS, BIOLÓGICOS Y AMBIENTALES.....	7
1.1 Objetivo general de la investigación.....	7
1.2 Sistemas de información para datos ecológicos, biológicos y ambientales.	7
1.2.1 Definiciones.....	7
1.2.2 Redes de Información Ecológica y Ambiental y Sistemas de Información de Red.....	8
1.2.3 Casos de estudio modelo.....	9
1.2.3.1 LTER Network (Long-Term Ecological Research Network).....	9
1.2.3.2 GCMD (Global Change Master Directory).....	10
1.2.3.3 Eco-Red Venezuela.....	10
1.2.3.4 Red de Estaciones Bio-climáticas del Edo. Mérida.....	12
1.3 Estándares y lenguajes de manejo e intercambio de datos y metadatos.....	13
1.3.1 Recursos de Información y metadatos.....	13
1.3.1.1 Recursos de Información.....	13
1.3.1.2 Metadatos.....	14
1.3.1.3 Importancia y razones para definir y usar metadatos.....	15
1.3.2 Modelos, Estándares y lenguajes para metadatos.....	16
1.3.2.1 Modelos para metadatos.....	16
1.3.2.2 Algunos modelos o esquemas de metadatos.....	18
1.3.2.3 Modularidad y extensibilidad de metadatos: los perfiles de aplicación de metadatos (Application Profiles).....	21
1.3.2.4 Ciclo de vida de los metadatos.....	22
1.3.2.5 Codificación de metadatos.....	23
1.3.3 Estándares para datos ambientales, biológicos, ecológicos y geo-espaciales.....	25
1.3.3.1 FGDC.....	25
1.3.3.2 NBII.....	26
1.3.3.3 Darwin Core (DwC).....	26
1.3.3.4 GCMD-DIF.....	27
1.3.4 Estándar de metadatos usado en proyecto LTER.....	27
1.3.4.1 Estándar de Metadatos para datos no-geoespaciales.....	27
1.3.4.2 Descriptores Genéricos de Metadatos para las ciencias ecológicas (LTER).....	28
1.3.5 Ecological Metadata Language: EML.....	30
1.3.5.1 Principios de diseño de EML.....	30
1.3.5.2 Diseño y estructura de EML (2.0.0).....	31
1.3.5.3 Creación de Metadatos EML.....	33
1.3.5.4 Implantación de EML en LTER.....	33
2 PROPUESTAS PARA EL MANEJO DE DATOS AMBIENTALES Y ECOLÓGICOS.....	35
2.1 Objetivos de la investigación y definición del problema.	35
2.2 Conceptos básicos y recomendaciones generales para "Gerentes de Información ".....	35
2.2.1 Definición de roles y acciones en el manejo de datos y metadatos.....	36
2.2.1.1 Gerentes de información.....	36
2.2.1.2 Actores.....	37

2.2.1.3	Acciones	38
2.2.1.4	Sistema de incentivos.....	39
2.2.2	Definición, selección y uso de estándares y herramientas	40
2.2.2.1	Estándares para metadatos	40
2.2.2.2	Herramientas para metadatos.....	41
2.2.3	Prueba, validación, control y aseguramiento de la calidad (QA/QC) de los datos	42
2.2.4	Educación y entrenamiento	43
2.3	Propuestas para desarrolladores de sistemas de información	44
2.3.1	Arquitecturas de Sistemas de Información para datos ambientales y ecológicos.....	44
2.3.1.1	Arquitecturas básicas para datos y metadatos.....	45
2.3.1.2	Arquitecturas básicas de un “Clearinghouse” de datos	45
2.3.2	Metadatos estructurados y estándares de intercambio de datos	48
2.3.2.1	Codificación y formato de metadatos	48
2.3.2.2	EML como lenguaje de representación e intercambio de metadatos.....	50
2.3.3	Tecnologías y herramientas para manejo de colecciones de datos	51
2.3.3.1	Herramientas relacionadas con XML	51
2.3.3.2	Herramientas relacionadas con EML.....	52
2.4	Sistemas prototipo para manejo de datos y metadatos	53
2.4.1	Descripción de prototipos desarrollados	53
2.4.2	Metodología y herramientas utilizadas.....	54
2.4.2.1	Metodologías.....	54
2.4.2.2	Lenguajes de programación, modelado y representación de datos.....	55
2.4.3	Definición de requerimientos para sistema de manejo de colecciones de datos.....	56
2.4.3.1	Características y descripción general para el sistema	56
2.4.3.2	Proceso general de creación y consulta de colecciones de datos.....	57
2.4.3.3	Diagramas de casos de uso.....	57
2.4.4	Arquitectura del prototipo de manejo de colecciones de datos.....	59
2.4.5	Descripción del sistema prototipo desarrollado.	60
2.4.5.1	Partes del sistema prototipo	60
2.4.6	Características e interfaz del prototipo.....	63
2.4.6.1	Interfaz de sistema de creación de colecciones de datos:	63
2.4.6.2	Interfaz de consulta de colecciones.....	64
3	APLICACIÓN A CASOS DE ESTUDIO: ECO-RED VENEZUELA Y RED DE ESTACIONES DEL EDO. MÉRIDA.....	69
3.1	Similitudes y diferencias con LTER-EUA y los casos de estudio. Aspectos del manejo de datos en Venezuela.....	69
3.1.1	Comparación de características de LTER con casos de estudio de este trabajo	69
3.1.2	Algunos aspectos de manejo de datos en proyectos locales:	70
3.2	Actividades realizadas en proyectos Eco-Red Venezuela y Red de Estaciones de Mérida.....	71
3.3	Plan general para manejo de colecciones de datos para Eco-Red Venezuela y Red de Estaciones del Edo. Mérida.....	72
3.3.1	Tareas organizacionales.....	72
3.3.2	Creación de colecciones de datos: procesos recomendados y escenarios posibles.....	73
3.3.2.1	Proceso general propuesto para creación de colecciones de datos	73
3.3.2.2	Escenarios posibles para el manejo de colecciones de datos.....	75
3.3.3	Entrenamiento: Talleres de Manejo de información para científicos y gerentes de información....	76
4	CONCLUSIONES Y TRABAJO FUTURO	78

5 BIBLIOGRAFÍA Y REFERENCIAS.....	81
6 ANEXOS	86
6.1 Tabla de contenidos detallada de taller básico de manejo de datos y metadatos	86
INDICE.....	88

Índice de Ilustraciones

FIG.1. EJEMPLO DE DEGRADACIÓN NORMAL DE LOS CONTENIDOS DE INFORMACIÓN ASOCIADOS CON DATOS Y METADATOS A TRAVÉS DEL TIEMPO	16
FIG.2. EJEMPLO DE METADATOS DE UNA COLECCIÓN DE DATOS CLIMÁTICOS EXPRESADOS EN DUBLIN CORE DENTRO DE UNA PÁGINA HTML.	20
FIG.3. CICLO DE VIDA DE UN RECURSO DE INFORMACIÓN O METADATOS.	23
FIG.4. ESTRUCTURA DE LA CLASE EML-DATASET	32
FIG.5. ROLES EN EL MANEJO DE METADATOS.	36
FIG.6. MODELO SIMPLIFICADO DEL CICLO DE VIDA DE LOS DATOS Y METADATOS	37
FIG.7. MODELOS BÁSICOS PARA SISTEMAS DE METADATOS	45
FIG.8. MODELO BÁSICO DE UN SISTEMA “CLEARINGHOUSE” DE DATOS.	46
FIG.9. MODELO TÍPICO DE UN CLEARINGHOUSE DE DATOS	46
FIG.10. CLEARINGHOUSE CENTRALIZADO DE METADATOS.	47
FIG.11. LOCALIZANDO FUENTES DE METADATOS: (RED DE SERVIDORES DE METADATOS).....	48
FIG.12. INTERFAZ GLOBAL DE BÚSQUEDA DE DATOS GEOESPACIALES DE GSDI.	48
FIG.13. ESQUEMA DE TRANSFORMACIONES XML-XSL	49
FIG.14. EJEMPLO DE ARCHIVO EML (EML-DATASET).....	50
FIG.15. MUESTRA DE INTERFAZ DE PROTOTIPOS 1 Y 2.....	54
FIG.16. ESQUEMA DEL MODELO DE PROTOTIPOS	55
FIG.17. DIAGRAMAS DE CASOS DE USO PARA DELIMITAR EL SISTEMA COMPLETO	58
FIG.18. CASOS DE USO PARA PROVEEDOR DE SERVICIOS Y PROVEEDOR DE DATOS	58
FIG.19. CASOS DE USO PARA USUARIO FINAL	59
FIG.20. DIAGRAMA DE ESTRUCTURA DEL SISTEMA	59
FIG.21. COMPONENTES DEL SISTEMA PROTOTIPO DE MANEJO DE COLECCIONES DE DATOS	60
FIG.22. SISTEMA DE CREACIÓN DE COLECCIONES	61
FIG.23. SISTEMA DE ALMACENAMIENTO DE COLECCIONES	62
FIG.24. SISTEMA DE CONSULTA DE COLECCIONES	62
FIG.25. SISTEMA DE FORMATO DE COLECCIONES	63
FIG.26. INTERFAZ WEB DE CAPTURA DE METADATOS	64
FIG.27. INTERFAZ WEB DE CONSULTA DE COLECCIONES DE DATOS.....	65
FIG.28. SALIDA DE CONSULTA DE ARCHIVO XML DE COLECCIÓN DE DATOS	66
FIG.29. SALIDA HTML DE CONSULTA DE COLECCIÓN DE DATOS.....	67
FIG.30. SALIDA EN FORMATO TEXTO A PARTIR DE INTERFAZ DE CONSULTA DE COLECCIONES	67
FIG.31. ESQUEMA DE ENTRADA DE DATOS Y CREACIÓN Y PUBLICACIÓN DE COLECCIONES.....	75

Índice de Tablas

TABLA 1. DESCRIPTORES DE METADATOS DE MICHENER (1997)	30
---	----

TABLA 2. HERRAMIENTAS PARA MANEJO DE COLECCIONES DE DATOS EML.....	53
TABLA 3. COMPARACIÓN DE CARACTERÍSTICAS DE LTER CON CASOS DE ESTUDIO	70
TABLA 4. PROCESO DE CREACIÓN DE COLECCIONES DE DATOS.....	74
TABLA 5. INFORMACIÓN SOBRE EL TALLER BÁSICO DE MANEJO DE DATOS Y METADATOS	77

1 Manejo de datos ecológicos, biológicos y ambientales

Las cantidades excepcionalmente grandes y complejas de información envueltas en la investigación del medio ambiente representan un reto significativo para las organizaciones que usan y manejan esta información. Los problemas de almacenamiento, recuperación y análisis de estos contenidos necesitan de un cuidadoso estudio de los métodos y herramientas más adecuadas para resolverlos.

Este trabajo se centra en el estudio del manejo, identificación, y preservación en el tiempo, de recursos de información relacionados con las ciencias ecológicas, biológicas y ambientales.

1.1 Objetivo general de la investigación

Este trabajo tiene como objetivo general, establecer recomendaciones y un grupo de métodos de trabajo para el manejo de datos y metadatos, que pueden ser aplicados en diferentes proyectos, desarrollos e investigaciones relacionadas con el manejo de información proveniente de la ecología, biología o disciplinas que estudian el clima y medio ambiente. Estas recomendaciones tienen que ver con la selección y uso de estándares para manejo de metadatos estructurados; con la selección de las herramientas informáticas más adecuadas y con la metodología de trabajo de las personas encargadas de aplicar dichos estándares y herramientas.

Los resultados de este trabajo de investigación podrán ser usados como una especificación de desarrollo de sistemas de información para manejo de datos y metadatos para las ciencias ecológicas, biológicas y ambientales, así como para el manejo de información proveniente de otras disciplinas de la ciencia y el conocimiento.

1.2 Sistemas de información para datos ecológicos, biológicos y ambientales.

1.2.1 Definiciones

Los Sistemas de información para datos ecológicos, biológicos y ambientales o *Sistemas de Información Ambientales* (EIS por sus siglas en inglés) se han desarrollado para el manejo de los recursos de información asociados con el medio ambiente (información climática, biológica, ecológica, meteorológica, etc.), y son usados principalmente para la investigación científica y como apoyo en la toma de decisiones y manejo de situaciones donde estén involucrados los tipos de datos mencionados. (UNDP-UNSO, 2002).

En esta investigación, nos referiremos indistintamente a los *Sistemas de Información Ambientales* como *Sistemas de Información Ecológicos, Biológicos y Ambientales*, *Sistemas de Información Bio-climática* o simplemente como *Sistemas de Información Ecológicos*.

Los *Sistemas de Información Ambientales*, permiten llevar a cabo estudios que abarcan periodos temporales grandes y espacios geográficos extensos, así como usar los datos para propósitos diferentes a los originales, por parte de personas distintas a las que recolectaron dichos datos (a esto se le llama uso secundario de los datos).

En (Torréns, 2001) se mencionaron algunos otros aspectos que tienen que ver con los beneficios que trae la implementación y uso de Sistemas de Información para las ciencias ecológicas. Se mencionan allí, los obstáculos que implica que la información o los datos no sean accesibles y que, peor aún, no se conozca siquiera su existencia, obstáculo que un sistema de información ayudaría a evitar. También se destaca que a través de la implementación de información altamente estructurada y sistemas asociados,

se pueden realizar análisis y descubrir correlaciones entre conjuntos disjuntos de información, lo que ayuda en la toma de decisiones e investigación científica.

Los Sistemas de Información Ambientales evolucionaron desde la pura adquisición de datos hasta el manejo de los mismos para soporte en la toma de decisiones. En este aspecto, los Sistemas de Información Ambientales han evolucionado a la par con la mayoría de los tipos de Sistemas de Información existentes. En [Torréns, \(2001\)](#) se detallan las diferentes fases por las que han pasado los sistemas de información ecológicos y ambientales desde los años 80 hasta ahora.

Por otro lado existe una tendencia entre grupos de investigación de todo el mundo, a crear redes internacionales de trabajo que permiten mantener contacto y conectarse entre ellos, así como tener acceso a los datos producidos por cualquiera de ellos. Esto se ha facilitado en los últimos años con la difusión del uso de redes de computadoras (Internet y el Web). Esto hace que los científicos y poseedores de datos, pongan énfasis en aspectos como el manejo de los datos, su formato y maneras de compartirlos, en un intento de maximizar el potencial y usos de las tecnologías de redes que están surgiendo ([Spedding, 2000](#)).

1.2.2 Redes de Información Ecológica y Ambiental y Sistemas de Información de Red

Como mencionábamos al final de la sección anterior, los científicos que manejan datos sobre un aspecto particular de las ciencias ambientales, o sobre un lugar geográfico determinado, necesitan intercambiar datos entre sí, debido a que los ecosistemas, climas y ambientes del mundo en realidad están estrechamente relacionados e interconectados y forman un sistema global. Esto lleva naturalmente a que los científicos intenten obtener datos distribuidos geográficamente y pertenecientes a varias disciplinas y áreas del conocimiento.

La toma de decisiones en cuanto a políticas y gerencia ambiental requiere datos a escalas regionales, nacionales y mundiales. Normalmente estos datos son recolectados a escalas mucho más pequeñas, lo que hace difícil su integración y análisis. El establecimiento de redes de información ambientales por lo tanto, se hace necesario para que los actores involucrados tengan acceso a datos en todos los niveles, locales, regionales, nacionales y mundiales, y los utilicen de una manera integrada e interconectada. Esto nos lleva a mencionar a continuación otro concepto, el de *Sistema de Información de Red (Network Information System - NIS)*, revisado aquí desde el punto de vista de las ciencias ecológicas y ambientales.

Cuando trabajamos con información útil para varias disciplinas, grupos y personas, se nos presentan algunas interrogantes como las siguientes ([Porter, 2001b](#)):

- ¿Que hacer para que la información y datos estén disponibles para los científicos, ecólogos, meteorólogos, entes gubernamentales, etc.?
- ¿Qué mecanismos se usaran para que puedan localizar la información que cada uno necesita, en formatos adecuados para diferentes usos, o que se puedan recuperar fácilmente?
- ¿Cómo se puede asegurar que la información esta actualizada y sea precisa y confiable?

Según el proyecto LTER, *Long Term Ecological Research Network*, (del que se hablara en la siguiente sección, 1.2.3), un *Sistema de Información de Red* (NIS por sus siglas en inglés) es aquel que soporta la investigación y manejo de datos ambientales tanto en el ámbito de los sitios de captura de datos como a nivel de la red a la que pertenecen muchos sitios ([Porter, 2001b, p.9](#)). Este concepto se vuelve más importante y a la vez factible debido al desarrollo creciente de las redes información y a las telecomunicaciones. En este contexto, se deben hacer esfuerzos en la practica para contestar cada una de

las interrogantes planteadas anteriormente, lo que contribuirá a establecer una relación más estrecha entre las ciencias ecológicas y ambientales y las tecnologías de la información y comunicación (Porter, 2001b, p.7).

1.2.3 Casos de estudio modelo

Se mencionan a continuación algunos proyectos internacionales que fueron considerados en esta investigación como casos de estudio modelo (LTER y GCMD). Se mencionan también dos proyectos que se están desarrollando actualmente en Venezuela¹, a los que en particular están dirigidas las recomendaciones propuestas en este trabajo.

1.2.3.1 LTER Network (Long-Term Ecological Research Network)²

La “Red de Investigación Ecológica a Largo Plazo” (LTER Network), es un esfuerzo colaborativo que envuelve más de 1000 científicos y estudiantes que investigan procesos ecológicos sobre escalas temporales y espaciales de gran magnitud (Brunt, 2001, p.1). LTER es una red con las características mencionadas anteriormente en 1.2.2.

La red LTER es un programa soportado por la *National Science Foundation (NSF)* de los Estados Unidos, e incluye 24 sitios o estaciones de investigación principales ubicadas en diversos ambientes desde bosques hasta ambientes marinos pasando por desiertos y humedales (Michener, 2001, p.1). También posee una *Oficina de Red (Network Office)*, que coordina la comunicación entre los sitios y otras actividades. Durante sus comienzos en los años 80, se hizo énfasis en satisfacer las necesidades y el manejo de datos e información a nivel individual de cada estación; aparte de eso cada sitio tenía iniciativas propias en cuanto al manejo de sus datos. A medida que el programa LTER maduró y la cantidad de sitios miembros aumentó, el tamaño y complejidad de las bases de datos específicas de cada sitio se incrementó grandemente, y la síntesis e integración de datos creció en importancia. Consecuentemente, se hizo también un gran énfasis en las actividades involucradas en la coordinación de las actividades de los sitios, y el desarrollo de un sistema de información de red (en la sección anterior, 1.2.2, se mencionaron características de este tipo de sistemas de información).

El programa LTER fue diseñado desde su concepción para considerar e incorporar el manejo y gestión de datos como una componente integral de la investigación científica. Dentro de su objetivo global de dar facilidad a la investigación, en los años 80 los gerentes de datos (*data managers*) de cada sitio, dedicaron significativos esfuerzos en desarrollar métodos para manejar documentación, y estudiaron todo lo que tiene que ver con formatos y custodia de datos, desarrollo, y mantenimiento de códigos y aplicaciones informáticas, etc. En los 90 se hicieron desarrollos basados en los avances anteriores, utilizando a su vez las tecnologías asociadas a Internet que estaban en rápida expansión para el desarrollo de sistemas de información accesibles globalmente. Los objetivos que se plantea para el futuro LTER (y que esta atacando actualmente), tienen que ver con levantar la infraestructura que permita pasar de administración, gerencia o manejo de datos a la gerencia del (Brunt, 2001, p.2).

SDSC-LTER Collaborative Research³

El Centro de Supercomputación de San Diego (SDSC) y la Organización LTER han establecido vínculos de trabajo y colaboración, con el objetivo de establecer un puente entre el mundo de la supercomputación y los retos científicos involucrados con el manejo de datos provenientes de la red de sitios de LTER, así como facilitar el uso de recursos de computación de alto rendimiento como soporte de los objetivos de la comunidad LTER. (SDSC-LTER, 2002)

¹ Este trabajo de tesis está enmarcado dentro de los dos proyectos mencionados.

² <http://lternet.edu/>

³ <http://www.sdsc.edu/sdsc-lter/>

Entre los objetivos de esta asociación se pueden mencionar:

- Promover y diseminar nuevas tecnologías relevantes a la investigación ecológica a largo plazo a través de la red LTER.
- Proveer a los científicos de las herramientas de última tecnología relativas al manejo y diseminación de datos a la comunidad LTER y a otras comunidades científicas.
- Promover y mejorar la interacción de la red LTER con otras comunidades ambientales y científicas.

Dentro de esta iniciativa se están desarrollando varios proyectos que tienen que ver con el manejo y procesamiento de información biológica y de biodiversidad, que incluye la creación de bases de datos, sistemas inteligentes de búsqueda de información, sistemas de trabajo colaborativo en esa área, etc.

1.2.3.2 GCMD (*Global Change Master Directory*)⁴

La NASA de los EUA, lleva a cabo un proyecto que consiste en reunir y distribuir gran cantidad de datos sobre el medio ambiente mundial, con el objetivo primario de facilitar la labor de los científicos que estudian los efectos de los cambios climáticos a nivel global.

A través de su sitio Web, se pueden buscar y obtener datos de diferentes tipos relacionados con el medio ambiente, organizados de forma temática, por su localización, por el tipo de instrumentación usada para obtenerlos, o por proyectos. Una característica interesante del sitio Web del *NASA-GCMD* es que se puede obtener datos y metadatos formateados siguiendo varios estándares internacionales. Se pueden pedir los datos en diferentes formatos (*DIF, GILS, Dublín Core, ANZLIC, etc.*); así como también recuperar siguiendo el estándar del FGDC. En la sección 1.xxx se hablara de los formatos y modelos para metadatos.

1.2.3.3 *Eco-Red Venezuela*⁵

El gobierno de Venezuela, a través del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICIT, ahora FONACIT) y estimulado por el éxito de la “*US Long Term Ecological Research (LTER) Network*”, decidió en 1996 establecer y soportar la Red Venezolana de Estaciones de Investigación Ecológica a Largo Plazo (*Eco-Red Venezuela*), en cooperación con instituciones científicas del país. En diciembre de 1996 el CONICIT creó la Comisión para la *Eco-Red Venezuela*, la cual condujo un proceso de análisis de 14 sitios potenciales, con base en criterios definidos. Finalmente 10 sitios fueron seleccionados. En el sitio Web de la *Eco-Red* se encuentra una lista y descripción de dichos sitios⁵ (*ECO-RED, 1998*).

La *Eco-Red Venezuela* se estableció el 29 de septiembre de 1997. Los representantes de cada estación se reunieron en ese momento para iniciar sus actividades de cooperación. Un segundo taller se realizó en Mérida, El Cobre, en marzo de 1998, con el apoyo del CONICIT, Fundacite-Mérida y la Universidad de Los Andes (ULA). En el mismo se trataron importantes aspectos relacionados con las organización de la red, incluyendo los conceptos, estructura y funcionamiento de la red en correspondencia con las necesidades del país, la misión y objetivos de la misma, áreas comunes de investigación, propuestas para desarrollar una Oficina Coordinadora y un proyecto para el fortalecimiento de las estaciones y su conectividad. Un tercer taller se realizó en febrero de 2000, en el Centro de Investigación Forestal El Meray (Edo. Monagas), con el apoyo de CVG-Proforca, CONICIT, Universidad de Los Andes y Fundación La Salle.

⁴ <http://gcmd.gsfc.nasa.gov>

⁵ <http://cesimo.ing.ula.ve/LA-ECO-RED/ECORED/>

La Eco-Red Venezuela está trabajando en las iniciativas siguientes: a) Plan de fortalecimiento de las capacidades de las estaciones de la red; b) Programa de conexión entre las estaciones de la red; c) Creación de bases de datos para cada estación, así como un directorio de investigadores y proyectos; d) Creación de la Oficina Coordinadora de la Eco-Red Venezuela y d) Vinculación a la red internacional (ILTER) y al proyecto regional de la red latinoamericana (LA-ILTER)⁶.

Se listan a continuación las reuniones efectuadas de la Eco-Red Venezuela:⁷

- Reunión. Mérida, Venezuela. Septiembre 2002.
- Simposio Investigación Ecológica a Largo Plazo en el Neotrópico. Mérida. Octubre 2001.
- IV Taller. Mérida, Venezuela. Octubre 29, 2001.
- IV Taller. Mérida, Venezuela. Octubre 29, 2001.
- III Taller. El Merey, 11 y 12 de febrero de 2000.
- II Taller. Mérida - El Cobre, 02 - 06 Marzo de 1998.

Es de destacar la reunión efectuada en Octubre del 2001 (*IV Taller Nacional de la Eco-Red Venezuela*), enmarcada en el IV Congreso Venezolano de Ecología realizado en Mérida, con el propósito de estimular el proceso de consolidación de la Eco-Red. Se plantearon los siguientes objetivos generales para dicho taller: evaluar el avance de la Eco-Red Venezuela, fortalecer la organización y conducción de la misma y definir la participación de la Eco-Red en la iniciativa para Latinoamérica (LA-ILTER). En dicha reunión participaron tanto miembros de las estaciones de la Eco-Red, como investigadores de la ULA y otras organizaciones. Se trataron allí importantes temas entre los que destacan: seguimiento de planes de trabajo, proyecto de conectividad de la Eco-Red, manejo de información, bases de datos y organización de metadatos, estándares para las estaciones, comunicación entre las estaciones; así como temas relativos al establecimiento de la red latinoamericana ILTER .

La última reunión importante de la Eco-Red, se efectuó en Mérida, el 20 y 21 de septiembre de 2002. Se pueden consultar las memorias, programas y otros documentos generados en estas reuniones, en la sección de eventos del sitio Web de la Eco-Red⁸.

Por último es importante mencionar que la presente investigación esta motivada y enmarcada en parte, dentro de los esfuerzos que lleva adelante la Eco-Red Venezuela para el manejo y gestión de datos y metadatos para los sitios miembro de esta red; aunque las recomendaciones y acciones generadas, pueden aplicarse a otros proyectos donde estén involucrados el manejo de información ecológica, biológica, climática, ambiental o científica.

1.2.3.4 Red de Estaciones Bio-climáticas del Edo. Mérida ⁹

Debido a necesidades particulares surgidas dentro de varias organizaciones, instituciones y personas del Edo. Mérida, se comenzó a estudiar y proyectar la instalación de una red de estaciones bioclimáticas para el Estado y el desarrollo de los sistemas de comunicación e información necesarios.

Uno de los problemas que se pretende abordar y que impulsa la búsqueda de respuestas y soluciones, es el que presentan los productores de algunos rubros agrícolas (plátano especialmente) de la Zona Sur del Lago de Maracaibo, donde eventos climáticos y plagas de varios tipos afectan la producción agrícola de la que dependen económicamente. La zona esta afectada por frecuentes fenómenos micro-climáticos (tornados y chubascos intensos de corta duración) que causan la pérdida o destrucción de algunos

⁶ <http://www.ilternet.edu/>

⁷ Tomado de: <http://cesimo.ing.ula.ve/LA-ECO-RED/ECORED/>

⁸ <http://cesimo.ing.ula.ve/LA-ECO-RED/ECORED/eventos.shtml>

⁹ <http://www.cecalc.ula.ve/investigacion/proyectos/2000/bioclimatehtml/redbioclimate.htm>

productos agrícolas. También la zona se ve afectada por plagas como la Sigatoka Negra, que ataca las especies de musáceas cultivadas en la región.

También hace falta el planteamiento de sistemas de alerta climática orientados a la prevención o aviso temprano de fenómenos climáticos que pueden causar problemas y desastres que afectan a la población del Estado y al país (como los fenómenos ocurridos en el Edo. Vargas en el año 1999, o las frecuentes inundaciones locales que ocurren en cada temporada de lluvias).

Cualquier iniciativa que pretenda dar solución o entender cualquiera de los dos casos mencionados anteriormente, necesita disponer de datos a varias escalas espacio-temporales, que son la base de cualquier investigación científica básica o de un sistema de apoyo a la toma de decisiones (como en el caso de un sistema de alerta climática contra desastres naturales).

Actualmente existen diferentes tareas pendientes en este proyecto (en algunas de ellas, esta investigación de tesis hará aportes importantes y directos), entre las que se destacan: adquisición, instalación y consolidación de estaciones de medición de parámetros climáticos y biológicos en diferentes sitios del estado Mérida, establecimiento de mecanismos y sistemas de gestión de datos y metadatos provenientes de estaciones (en la propuesta de tesis a estos sistemas se les dio el nombre de "Sistemas de Información Bio-climática") (Torréns, 2001), desarrollo de proyectos que ocupen los datos recolectados como fuente primaria de información, identificación de proyectos que necesiten hacer reuso secundario de datos recolectados, etc.

Se destacan los siguientes objetivos generales para el proyecto:

- Establecer una infraestructura de datos ambientales.
- Establecer una unidad inicial de coordinación para la red, desde el punto de vista del manejo, custodia y distribución de los datos generados.
- Asegurar la distribución de los datos generados.
- Asegurar la interoperabilidad de los sistemas desarrollados con otros sistemas (similares o no).
- Asegurar la permanencia de los datos por periodos largos del tiempo.

Entre los objetivos específicos definidos para este proyecto, podemos mencionar: crear mecanismos y herramientas que faciliten el acceso a los datos generados por cualquier estación participante en la red o científicos que los necesiten, usar estándares internacionales (de formato y contenidos) para estructurar y almacenar los datos (para permitir intercambio con otros sistemas), generar diferentes productos de información, entre otros (Torréns, 2001).

Se puede encontrar una lista detallada de objetivos, así como otros detalles referentes a este proyecto en el documento de definición del proyecto (Ablan, 2000, p.13).

1.3 Estándares y lenguajes de manejo e intercambio de datos y metadatos

Se tratan en esta sección algunas definiciones básicas, que son importantes para el desarrollo de este trabajo: "recurso de información" y "metadatos para seguir luego discutiendo, su naturaleza e importancia dentro del contexto de esta investigación. Se describen los diferentes modelos para el manejo de recursos de información y metadatos, y los lenguajes para representar y codificar la estructura y elementos de dichos modelos.

1.3.1 Recursos de Información y metadatos.

1.3.1.1 Recursos de Información

Según el Comité Europeo para la Estandarización (CEN, 1997) (CEN por sus siglas en francés)¹⁰,

" un recurso de información es una unidad lógica de información que puede ser almacenada, procesada, transmitida o reproducida y manejada como unidad por los creadores de la información, proveedores de servicio y usuarios. Puede tener un conjunto de metadatos asociados con él" ...

...(más adelante en este capítulo se dan definiciones del término "metadatos").

Uno de los efectos del crecimiento de la llamada "Sociedad de la Información" es que una enorme cantidad de valiosa información en formato digital está disponible en todo el mundo gracias a las tecnologías de la información y la comunicación modernas (TICs). Esos recursos de información tienen un gran potencial para ser combinados, re-usados, y aprovechados. Un problema clave es que dichos recursos están distribuidos a través de ambientes (técnicos) heterogéneos, que hacen que su descubrimiento (*discovery*) y uso sea una tarea difícil. Otro problema es que la calidad, confiabilidad y exactitud de estos recursos varía dependiendo de la fuente de información. A su vez, mecanismos estandarizados son necesarios para permitir la identificación, localización, uso y re-uso de recursos de información digitales. El desarrollo y adopción de esquemas de metadatos estándares es una parte clave de la solución del problema. En las secciones 1.3.2 y 1.3.3 se tratan en detalle los diferentes estándares creados para diferentes tipos de recursos de información, y en particular los estándares que tienen que ver con manejo de información proveniente de las ciencias ecológicas.

Como ejemplo de recursos de información podemos mencionar: catálogos bibliográficos y los libros-publicaciones a los que se refieren, catálogos y objetos de un museo (pinturas o cualquier tipo de objetos), colecciones de datos geoespaciales como datos de mediciones oceanográficas o climáticas de la atmósfera terrestre, etc.

Es importante aquí también proponer la definición del término "colección de datos" (en inglés se usa frecuentemente el término *dataset*):

"Una colección de datos es una serie de observaciones recolectadas con la misma metodología" (LTER-Luquillo, 2001).

Cada colección de datos debería tener documentación suficiente para que alguien no familiarizado con la investigación pueda replicar el estudio. Una colección de datos puede ser dividida en sub-colecciones (archivos de datos) que son discretos en el espacio y el tiempo. La documentación para una colección de datos debería incluir todas las subdivisiones espaciales y temporales de los datos.

Existen dos aspectos clave que tienen que ser discutidos con cuidado. El primero consiste en la forma de "describir" los recursos de información o los objetos asociados a esta, permitiendo su uso efectivo (organización de la información). El segundo aspecto sería como "encontrar" los recursos de información u objetos de información que alguien necesita (recuperación de la información). En ambos aspectos está involucrado el término definido a continuación: "metadatos".

1.3.1.2 Metadatos.

La definición básica de metadatos y que al parecer es aceptada universalmente es la siguiente:

¹⁰ <http://www.cenorm.be>

Metadatos: "datos sobre datos".

Otra definición frecuentemente citada proviene de T. Berners Lee ¹¹, uno de los padres del World Wide Web (WWW):

... "la información sobre información es generalmente conocida como Metadatos. Específicamente, en el área del WWW, metadatos son información procesable por computadoras acerca de un recurso en el Web u otras cosas".

Otras definiciones generales serían:

"Información estructurada que describe, explica, localiza, o de alguna manera facilita recuperar, usar o gerenciar un recurso de información" (NISO, 2001, p.3).

"Todos los datos y conocimientos que proporcionan información acerca de procesos, datos y entidades" (Ahmed, 2001, p.344).

En otros casos algunas definiciones tienden a restringir o reflejar algún dominio de aplicación particular. La siguiente definición es proporcionada por uno de los miembros principales del proyecto LTER, y asocia el término "metadatos" con el de "colección de datos" que sería en este caso el equivalente a "recurso de información", "dato" o "entidad" mencionados en las anteriores definiciones:

"Nivel superior de la información, o instrucciones que describen el contenido, contexto, calidad, estructura, y accesibilidad de una colección de datos específica" (Michener, 1997, p.331).

Por lo general los metadatos están asociados con un objeto"; este objeto es un recurso de información. El objeto en sí mismo puede o no ser digital (por ejemplo el libro en una biblioteca al cual se refiere el catálogo de la biblioteca puede ser electrónico o físico). Los metadatos también existen en ambas formas digital y no digital. En (CEN, 1999, p.10) se describen las diferentes formas de relacionar los metadatos con el recurso de información (objeto) descrito. En la sección 1.3.2 se explicarán también las diferentes formas que toman los metadatos (modelos o tipos de metadatos).

Los metadatos han sido usados principalmente en conexión con datos discretos como datos alfanuméricos y numéricos almacenados en una base de datos convencional. Es frecuente generar a partir de este tipo de datos definiciones conocidas como "diccionarios de datos" (archivo donde se define la organización básica de una base de datos)¹².

Los Metadatos pueden ser usados para dar información acerca del contenido de un recurso de información, con el objetivo de facilitar el "descubrimiento" (*discovery*), identificación, localización, o detección de uso ilegítimo (o prevenirlo), de dicho contenido (CEN, 1999, p.9).

1.3.1.3 Importancia y razones para definir y usar metadatos.

Los metadatos permiten a usuarios diferentes al investigador principal hacer uso efectivo de los datos y determinar sus aplicaciones más apropiadas (uso secundario). Es frecuente encontrar donde

¹¹ <http://www.w3.org/DesignIssues/Metadata.html>

¹² Tomado de diccionario de términos de computación
"Webopedia": http://www.webopedia.com/TERM/d/data_dictionary.html

colecciones de datos han tenido que ser descartadas debido a que los investigadores no proporcionaron la documentación adecuada para el uso por parte de terceras personas (UNDP-UNSO, 2001, p.8). Resulta importante para los investigadores entender que el incremento en los costos de incluir la documentación necesaria en el momento de recolectar los datos es pequeña en comparación con el costo de intentar reconstruirlos al final del proyecto o tiempo después que este ha culminado, puede ser prohibitivo o imposible de realizar.

Una de las principales razones que justifican considerar el uso de metadatos en un sistema de información, es la de evitar la pérdida de información en el tiempo. Esta pérdida ocurre por muchas causas: degradación de los medios de almacenamiento (papel, cintas y discos de computadora, etc.), retiro o muerte del científico que recolectó los datos, obsolescencia de la tecnología y medios de almacenamiento, destrucción de dichos medios por eventos catastróficos, etc.

La pérdida de datos también ocurre a través de la ausencia de los modelos conceptuales usados para interpretarlos. Algunos conjuntos de datos están asociados a hipótesis y relaciones no lineales que solo pueden ser representadas por complejos modelos de simulación. Por lo tanto preservar una colección de datos también incluye la conservación del modelo de simulación y sus respectivos archivos de entrada y salida. Por otra parte, los científicos han reconocido la necesidad de preservación de los datos (Michenner, 1997, p.332), pero a menudo solo se enfocan en el resguardo de resultados sintéticos a través de la “publicación”.

En la Fig.1 tomada de (Michenner, 1997, p.332) se da un ejemplo de la degradación de la información asociada con datos y metadatos a través del tiempo.

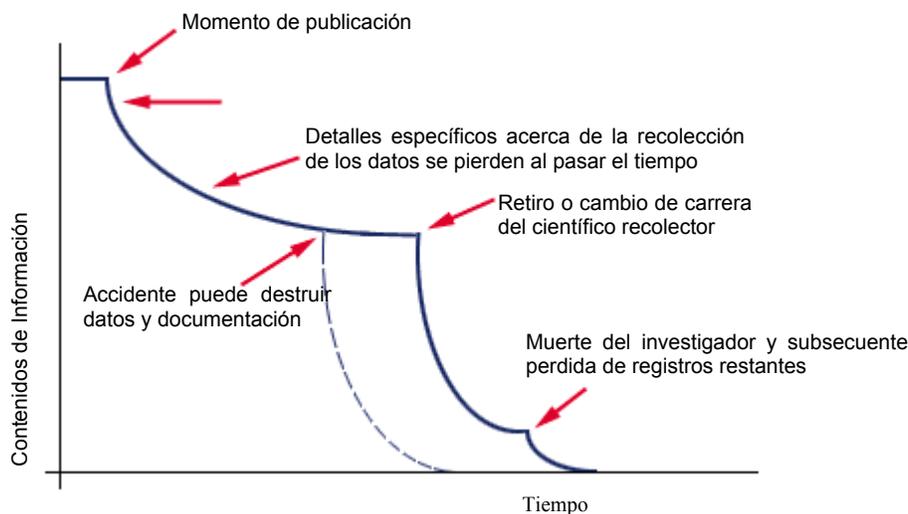


Fig.1. Ejemplo de degradación normal de los contenidos de información asociados con datos y metadatos a través del tiempo

Según Michener (1997, p.331) algunos objetivos específicos de la implementación de sistemas de información que hagan uso de metadatos serían:

- Facilitar la identificación y adquisición de datos para un tema determinado, y para un periodo de tiempo o localización geográfica específica.

- Facilitar la determinación de la conveniencia de los datos para cumplir con un objetivo específico.
- Facilitar el procesamiento, análisis y modelado de los datos.
- Incluir dentro de los datos conocimiento semántico acerca de los mismos.

Una adecuada documentación sobre el muestreo, procedimientos analíticos, anomalías y calidad de los datos, y estructura de las colecciones de datos ayudara a asegurarse que esos datos puedan ser correctamente interpretados o reinterpretados más adelante en el tiempo. A menudo se establecen 20 años como el objetivo para mantener datos usables por científicos o personas no familiarizadas con dichos datos o con su recolección (Michener, 1997, p.332)

Al existir una creciente necesidad de intercambio de datos entre diferentes entes recolectores (lo que aquí llamaremos colaboración inter-sitios), la ampliación de los estándares de metadatos para abarcar, aparte de la documentación sobre los datos, la información sobre el sitio recolector, abre nuevas posibilidades de colaboración y compartición de la información.

Otro beneficio importante es que los metadatos proveen información crítica para expandir las escalas en las que los científicos ambientales trabajan. Estudios que incluyen comparaciones temporales entre sitios, y comparaciones dentro de cada sitio, dependen de la disponibilidad de suficientes metadatos (Michener, 1997, p.332). Con las herramientas tecnológicas actuales se pueden hacer estudios a escalas regionales y globales, que hace unos pocos años hubieran sido impensables; y de hecho existen varios proyectos que están haciendo ese tipo de estudios (por ejemplo el proyecto “*Global Change Master Directory*” (GCMD)¹³ dirigido por la NASA, mencionado en la sección 1.2.3.2).

Los metadatos también son críticos para lograr combinar conjuntos de datos físicos, químicos y biológicos que contienen diferentes parámetros pero comparten dominios temáticos, espaciales o temporales comunes. (Michener, 1997, p.333)

1.3.2 Modelos, Estándares y lenguajes para metadatos.

1.3.2.1 Modelos para metadatos

Cuando comunidades con intereses comunes se ponen de acuerdo para intercambiar metadatos, crean una taxonomía y jerarquías estándares de ítems o elementos; esto es a lo que se llama “*modelo de metadatos*” (CM-LSAL, 2001). A estas agrupaciones de elementos creados para un propósito particular también se les llama “*esquemas de metadatos*”.

Para intercambiar información sobre un objeto en particular, se crea un registro de metadatos; el registro es específico para un objeto y un modelo particular de metadatos. Generalmente, un registro de metadatos puede ser intercambiado con otro sistema que entiende ese modelo de metadatos, pero no lo puede hacer con sistemas que entienden otros modelos diferentes.

Según un reporte del “*Centro Europeo de Estándares*” (CEN, 1999, p.10), la mayoría de los modelos para metadatos están basados en tres conceptos principales: 1. *Clases de metadatos*, 2. *Roles*, 3. *Acciones*. A continuación se describe brevemente cada uno de estos conceptos.

1. Clases de Metadatos.

¹³ <http://gcmd.gsfc.nasa.gov>

Las estructuras de metadatos pueden ser arbitrariamente simples o complejas y la información en los metadatos puede ser clasificada de acuerdo a diferentes criterios. Estas estructuras (clases) de metadatos están compuestas por un cierto número de "elementos" o ítems, asociados a definiciones semánticas descriptivas de los atributos de un recurso de información. Cada comunidad puede, dentro de su modelo, definir de manera diferente una propiedad o elemento de un recurso de información, o pueden ponerse de acuerdo en intercambiar o usar definiciones. Por ejemplo, la iniciativa *Dublin Core* (de la cual se darán detalles más adelante) especifica un conjunto base de 15 elementos que pueden ser aplicados para describir gran cantidad de recursos de información. El modelo de LOM (*Learning Objects Metadata*)¹⁴ que está siendo desarrollado por la *IEEE* y otras organizaciones para describir recursos en ambientes de enseñanza-aprendizaje tiene cerca de 100 elementos.

2. Roles.

Tres roles principales se pueden identificar en un modelo de metadatos: *Creador*, *Proveedor de Servicios* y *Usuario*. Estos roles se pueden aplicar a los recursos de información (datos) o a los metadatos. El *Creador* (autor) genera el recurso, el *Proveedor de Servicios* ofrece servicios a los usuarios, relacionados con el uso de los recursos de información o metadatos; y por último el *Usuario* hace uso del recurso de información durante su existencia. El usuario puede ser un humano, una organización, un agente de software o una aplicación.

3. Acciones.

Creadores, *Proveedores de Servicio* y *Usuarios* ejecutan acciones que aplican tanto a recursos de información como a los metadatos. Un creador puede por ejemplo, crear, capturar, preparar, digitalizar, editar, ensamblar, actualizar, re-usar o descartar un recurso de información; un proveedor de servicio puede mantener, validar, asegurar la calidad, vender, etc (no puedes poner etc aquí.. debe ser completa la lista). un recurso de información o metadatos (entre otras acciones); un usuario puede buscar, seleccionar, usar, leer, reenviar, (entre otras)etc. (idem) un recurso de información o metadatos.

El enlace entre un registro de metadatos y el recurso descrito por él, puede tomar una de las dos formas siguientes:

1. Los elementos pueden estar contenidos en un registro separado del ítem o recurso, como en el caso del registro en el catálogo de las bibliotecas; ó
2. Los metadatos pueden estar incluidos en el recurso mismo.

Un ejemplo de metadatos que aparecen junto con el recurso mismo, son los datos del sistema "*Cataloging in Publication*" (*CIP*) para libros y publicaciones diversas ([CIP-LOC, 2002](#); [CIP-NLA, 2003](#)), y los datos de catalogación de la Biblioteca del Congreso de Los Estados Unidos que aparecen en el revés de la página del título de muchos libros publicados en los EUA y otros países. El objetivo es colocar datos que sirvan para la catalogación, dentro de la publicación misma, como se menciono anteriormente. Estos sistemas, entre otras cosas, facilitan el procesamiento y catalogación de los libros y las publicaciones por parte de las librerías, bibliotecas y vendedores de libros ([Bell, 2002](#)). En el capítulo 2 (sección 2.3.1) se mencionan con detalle las diferentes arquitecturas y maneras de modelar sistemas de información que hacen uso de metadatos

1.3.2.2 Algunos modelos o esquemas de metadatos

Se presenta un listado de los esquemas más usados actualmente, clasificados como esquemas de uso general y de descripción bibliográfica, y esquemas usados por diferentes comunidades de usuarios para

¹⁴ <http://ltsc.ieee.org/wg12/doc.html>

documentar colecciones de datos geoespaciales y no-geoespaciales. Se muestra la dirección Web principal de soporte y documentación para cada uno.

Esquemas de metadatos generales y de descripción bibliográfica:

- Dublin Core: <http://dublincore.org>
- LOM, Learning Objects Model: <http://ieee-lom>: <http://ltsc.ieee.org/wg12/>
- TEI, Text Encoding Initiative: <http://www.tei-c.org/>
- GILS, Global Information Locator Service: <http://www.gils.net/>
- EAD, Encoded Archival Description: <http://lcweb.loc.gov/ead/>
- MARC: <http://www.loc.gov/marc/>
- OAI, Open Archives Initiative: <http://www.openarchives.org/>
- Z39.50: <http://lcweb.loc.gov/z3950/>
- RDF, Resource Description Framework: <http://www.w3.org/RDF/>

Esquemas de metadatos para colecciones de datos de diferentes disciplinas:

- *Datos geo-espaciales.* FGCD, Federal Geographic Data Committee: <http://www.fgdc.gov/>
- *Datos biológicos.* FGDC-NBII, FGDC National Biologic Information Infrastructure: <http://www.fgdc.gov/nbii>
- *Ciencias Sociales.* DDI, Data Documentation Initiative: <http://www.icpsr.umich.edu/DDI/>
- *Datos ecológicos.* EML, Ecological Metadata Language: <http://knb.ecoinformatics.org/software/eml/>
- *Datos geo-espaciales, biológicos, etc.* GCMD-DIF, Global Change Master Directory, Directory Interchange Format: <http://gcmd.gsfc.nasa.gov/>

Se describen a continuación algunos de los modelos para metadatos listados anteriormente. Unos son de propósito general y otros tratan áreas específicas (educación, literatura, geografía, etc.). En la sección 1.3.3 se describirán en detalle varios de los modelos de metadatos para colecciones de datos científicos.

1.3.2.2.1 Dublin Core¹⁵

La iniciativa de metadatos *Dublin Core* (DCMI por sus siglas en inglés) trata de fomentar la adopción a gran escala de estándares de metadatos interoperables y de promover el desarrollo de vocabularios especializados de metadatos para la descripción de recursos.

Los elementos de metadatos *Dublin Core* son usados para complementar los métodos de búsqueda e indexación de metadatos, sin importar si el recurso correspondiente es un documento electrónico o un objeto físico real. El "*Dublin Core Element Set*" (DCMES) fue el primer estándar de metadatos creado por la DCMI, y provee un vocabulario semántico para describir un "núcleo básico" de propiedades de la información.

El objetivo original de DCMI fue definir un conjunto de elementos que pudiera ser usado por los autores para describir sus propios recursos Web (NISO, 2001, p.3). Dada la proliferación de recursos electrónicos y la incapacidad de los bibliotecarios de catalogar todos esos recursos, el objetivo fue definir unos pocos elementos y algunas reglas simples que pudieran ser aplicadas por "catalogadores no especializados" y que pudieran ser compartidos entre diferentes disciplinas o entre cualquier tipo de organización que necesitara organizar y clasificar información. Actualmente son 15 los elementos

¹⁵ <http://dublincore.org>

definidos. Se muestra un listado de estos elementos (entre paréntesis el nombre oficial del elemento en inglés)

Elementos de Dublin Core:

1. título (title),
2. asunto (subject),
3. descripción (description),
4. fuente (source),
5. idioma (language),
6. relación (relation),
7. cobertura (coverage),
8. creador (creator),
9. editor (publisher),
10. contribuidor (contributor),
11. derechos (rights),
12. fecha (date),
13. tipo (type),
14. formato (format),
15. identificador (identifier).

En la siguiente figura (Fig.2), se muestra un ejemplo de una página Web hipotética desde donde se accede a datos climáticos capturados por una estación meteorológica; y que posee metadatos de identificación de una colección de datos, incluidos en etiquetas “*META*” usando la sintaxis de los elementos de *Dublin Core*.

```

estacion_hechicera.html
<HTML>
<HEAD><TITLE>Estacion la Hechicera 2001<TITLE>
<META NAME="DC.Title"
  CONTENT="Datos Estación la Hechicera 2001">
<META NAME="DC.Description"
  CONTENT="Colección de datos climáticos capturados por la estacion meteorológica
  La Hechicera. Esta estación es parte de la Red de Estaciones del Edo. Mérida.">
<META NAME="DC.Creator.Name"
  CONTENT="Carlos Guada">
<META NAME="DC.Publisher"
  CONTENT="Centro de Calculo Científico, Universidad de Los Andes">
<META NAME="DC.Type"
  CONTENT="text">
<META NAME="DC.Date.Issued"
  CONTENT="2001">
<META NAME="DC.Subject.Geographic"
  CONTENT="La Hechicera, Mérida, Venezuela">
<META NAME="DC.Coverage.dateStart"
  CONTENT="01-2001">
<META NAME="DC.Coverage.dateEnd"
  CONTENT="12-2001">
<META NAME="DC.Format"
  CONTENT="text/txt">
<META NAME="DC.Identifier.URL"
  CONTENT="http://www.cecalc.ula.ve/datos/lahechicera/2001/">
<META NAME="DC.Rights.Use"
  CONTENT="Uso libre mencionando la fuente (DC.Identifier.URL)">
</HEAD>
<BODY>
<h1>Datos Estación Hechicera 2001</h1>
  <a href="/webclima/datos/lahechicera/2001/"> Ver archivo texto con datos año 2001 </a>
</BODY>
</HTML>

```

Fig.2. Ejemplo de metadatos de una colección de datos climáticos expresados en Dublin Core dentro de una página HTML.

Cualquiera puede usar los metadatos definidos por *Dublin Core* para describir recursos de un sistema de información. Las páginas Web son uno de los tipos de recursos más comunes que utilizan las descripciones de *Dublin Core*, usualmente dentro de las etiquetas <meta> de HTML como se ve en la figura anterior.

Dublin Core fue desarrollado para ser simple y conciso, y para describir documentos en el Web. Su uso se ha incrementado últimamente con otro tipo de materiales y aplicaciones más complejos. Los elementos de metadatos definidos por *Dublin Core* están siendo usados como la base de sistemas descriptivos por varios grupos de comunidades con un interés común, tales como:

- Organizaciones educacionales
- Bibliotecas
- Instituciones gubernamentales
- Sector de la Investigación científica
- Autores de páginas Web
- Corporaciones con sistemas de manejo de conocimiento

(ver proyectos basados en *Dublin Core*: <http://www.dublincore.org/projects/>)

1.3.2.2.2 IEEE-LOM¹⁶

Según (IEEE-LOM, 2002) el propósito del estándar *LOM (Learning Objects Metadata)* es el de facilitar la búsqueda, adquisición, y uso de objetos de aprendizaje” (*learning objects*), por parte de alumnos o instructores o por procesos automatizados de software. Este estándar también facilita compartir e intercambiar “objetos de aprendizaje”, permitiendo el desarrollo de catálogos e inventarios que toman en cuenta a su vez, la diversidad de contextos culturales y de lenguaje dentro de los cuales estos objetos y sus metadatos son usados.

La definición más aceptada de “objeto de aprendizaje” proviene del comité de IEEE que desarrolla el estándar *LOM: (IEEE Learning Technology Standards Committee)* “...cualquier entidad, digital o no digital, que pueda ser usada, re-usada o referenciada durante el aprendizaje apoyado con tecnología”.

1.3.2.2.3 LOC-MARC¹⁷

MARC es el acrónimo para *Machine-Readable Cataloging*. Define un formato de datos que surgió a partir de una iniciativa de la *Librería del Congreso de los Estados Unidos (LOC)* hace unos treinta años atrás (LOC-MARC, 2003). MARC provee mecanismos por el cual las computadoras intercambian, usan e interpretan información bibliográfica, y sus elementos de datos forman la base de muchos de los catálogos de las bibliotecas hoy en día.

1.3.2.2.4 Z39.50¹⁸

Z39.50 es un estándar que especifica un protocolo para recuperación de información basado en arquitecturas cliente/servidor (Z39.50, 2001). Especifica procedimientos y estructuras para que un

¹⁶ <http://ltsc.ieee.org/wg12/>

¹⁷ <http://www.loc.gov/marc/>

¹⁸ <http://www.loc.gov/z3950/agency/>

cliente busque en una base de datos provista por un servidor, recupere registros de base de datos identificados por una búsqueda y genere una lista de resultados. El estándar también soporta control de acceso, control de recursos, servicios extendidos, y una facilidad de ayuda. El protocolo maneja la comunicación entre aplicaciones que recuperan información y el cliente y servidor (que pueden estar en computadores diferentes); no maneja la interacción entre el cliente y el usuario final.

1.3.2.3 *Modularidad y extensibilidad de metadatos: los perfiles de aplicación de metadatos (Application Profiles)*

Los realizadores de proyectos que necesitan usar modelos o esquemas de metadatos, por lo general encuentran que ningún estándar de metadatos cumplirá enteramente con sus necesidades descriptivas.

En (Michener, 1997, p.6) se define un *perfil de aplicación* como “ensamblaje de elementos de metadatos seleccionados de uno o más esquemas de metadatos y combinados en un esquema compuesto”. Los *perfiles de aplicación* permiten la modularidad y extensibilidad de los metadatos a través del reuso de la semántica de otros esquemas de metadatos y a su vez los “re-empaquetan” para su uso en un dominio de aplicación particular.

Los estándares generales como *Dublin Core* a menudo deben ser usados junto con perfiles específicos del dominio de aplicación, por ejemplo *IEEE-LOM* para recursos educativos (ver sección 1.3.2.2); o el estándar *CSDGM-FGDC* debe ser usado junto con perfil de aplicación *NBII* para describir recursos relacionados con las ciencias biológicas (ver sección 1.1.1 para una descripción de este estándar). También existe la posibilidad de añadir nuevos elementos que ningún estándar proporciona.

Según (Lagoze, 1996) el *Marco de Referencia de Warwick (Warwick Framework)* proporcionó un modelo inicial para resolver la manera como los metadatos pueden ser agregados en “paquetes” con el objetivo de combinar diferentes conjuntos de elementos relativos a un recursos de información. Luego de este modelo inicial se hicieron varias otras propuestas que permitían extender las capacidades descriptivas de estándares generales como *Dublin Core*. En (UKOLN, 2001) y (Dekkers, 2001) se pueden encontrar explicaciones detalladas de los mecanismos usados para crear perfiles de aplicación, así como ejemplos descriptivos.

En el tema de los *perfiles de aplicación* están muy ligados conceptos y herramientas como los *espacios de nombres XML* y a *RDF (XML Namespaces y Resource Description Framework)*. Un *espacio de nombres (namespace)* es una colección formal de términos manejados de acuerdo a cierta política o algoritmo (Duval, 2002, p.1). *RDF* proporciona el mecanismo para combinar elementos individuales de una variedad de esquemas diferentes, permitiendo a los que están implementando un perfil o usando diferentes estándares, escoger los elementos que se adapten mejor a sus propósitos. En la sección 1.3.2.5 se darán definiciones de estos y otros términos relacionados.

1.3.2.4 *Ciclo de vida de los metadatos*

Se describe en esta sección un modelo de ciclo de vida para un recurso de información o metadatos propuesto por el *Centro Europeo de Estándares (CEN, 1999)*. Este modelo pretende ilustrar (en orden cronológico) las fases por las que una entidad u objeto de información pasa a lo largo de su existencia.

El modelo mostrado puede ser aplicado tanto a un recurso de información como a sus metadatos asociados. En la Fig.3 se ilustra el modelo.

El modelo consiste en fases por las que un recurso de información o metadatos pasa, desde su creación o adquisición hasta su fase final de existencia. Estas fases representan una secuencia en el tiempo en la que también pueden ocurrir diferentes iteraciones (lazos desde “Control de Calidad” por ejemplo). En (CEN, 1999) se describe detalladamente cada fase de este modelo.

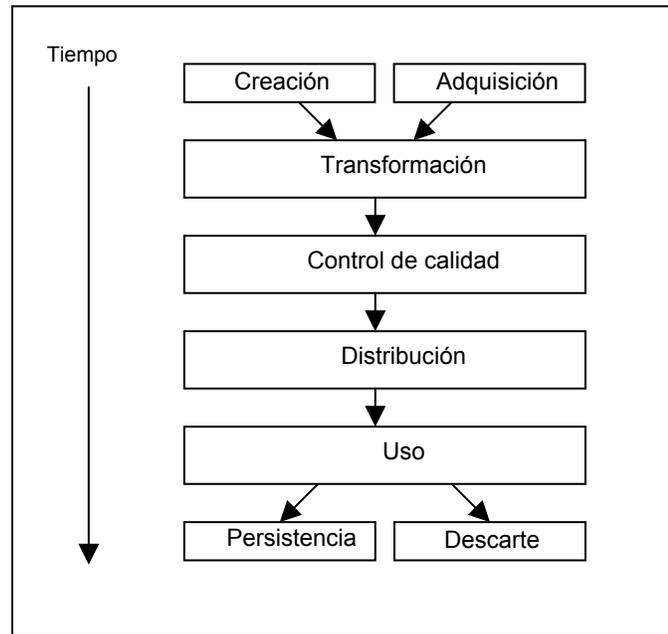


Fig.3. Ciclo de vida de un recurso de información o metadatos.

1.3.2.5 Codificación de metadatos

La semántica tiene que ver con el significado; la sintaxis trata de la forma. Se deben establecer acuerdos y estándares entre dos comunidades para que puedan compartir metadatos. Primeramente se deben establecer acuerdos en cuanto al uso de un modelo (esquema) de datos particular, y en cuanto al significado de los términos (elementos). Sin embargo hasta tanto no se haya alcanzado una convención compartida para identificar y codificar los nombres y valores de los elementos de información, no se podrá intercambiar metadatos de manera fácil y directa.

Es necesario destacar aquí la importancia de mantener la semántica y la sintaxis separadas tanto como sea posible. Hemos sido testigos en los últimos años de los rápidos cambios que esta experimentando el WWW; hemos observado varias versiones del lenguaje *HTML* y hemos observado el surgimiento y evolución de *XML* y sus derivados (*XML Schemas*, *XML Namespaces*, *XSLT*, *RDF Schemas*, etc.). Esta falta de estabilidad hace patente la necesidad de mantener la independencia entre la semántica de los elementos de metadatos y su representación sintáctica (Michener, 1997, p.7). Sin embargo, y debido al hecho que muchos de los recursos de información y metadatos “nacen” en formato digital, es de esperar que el proceso de creación y manejo de estos, incluya especificaciones acerca de la representación e intercambio de dichos recursos.

A continuación repasaremos brevemente algunos formatos de representación y codificación de recursos de información y metadatos.

HTML (Hypertext Markup Language)

La mayoría de los metadatos en el Web hoy en día son codificados usando *HTML* “incrustado” (*embedded*) dentro de los recursos Web. Esta aproximación tiene la ventaja de la simplicidad, no se necesitan sistemas adicionales para su producción, la infraestructura Web provee el sistema necesario en la forma de etiquetado *HTML* y protocolo *http* para su intercambio a través de redes de información; pero a su vez limita la riqueza estructural de los elementos de metadatos que pueden ser representados, y mezcla el formato de salida con los datos mismos.

*XML (Extensible Markup Language)*¹⁹

Según la organización *W3C*, *XML* es un formato basado en texto, simple y muy flexible derivado de *SGML*. El *XML* es un significativo avance en la manera como los datos son descritos e intercambiados por aplicaciones basadas en el Web, haciendo uso de formatos simples y flexibles basados en estándares (*XML-W3C, 2003*).

Iniciativas para metadatos como *Dublin Core* (ver sección 1.3.2.2) y *RDF* (del que se hablara en la siguiente sección) pueden usar a *XML* como lenguaje de implementación.

XML, como complemento de *HTML*, promete incrementar los beneficios derivados de la gran cantidad de información que se encuentra hoy en día en las redes de computadoras alrededor del mundo. Esto se debe a que *XML* provee un método uniforme para la descripción e intercambio de datos estructurados. La habilidad para describir datos estructurados en un formato abierto basado en texto y distribuir estos datos usando el protocolo *http* estándar, es muy importante por dos razones:

1. *XML* facilita declaraciones más precisas de los contenidos y permite obtener resultados de búsquedas más precisas a través de múltiples plataformas.
2. Una vez que los datos son localizados se pueden manipular y visualizar de múltiples maneras.

Las reglas que rigen su uso están siendo definidas por la organización *W3C (World Wide Web Consortium)*²⁰, asegurándose así que los datos estructurados serán uniformes e independientes de aplicaciones o compañías de software. Dentro del Consorcio *W3C* participan gran cantidad de compañías de desarrollo de software, lo que asegura por otro lado que el estándar sea soportado en el futuro por aplicaciones creadas por estas compañías.

El poder de *XML* se basa en que mantiene separados la interfaz de usuario de los datos estructurados, permitiendo la integración de datos provenientes de diversas fuentes. *XML* también facilita el intercambio de datos estructurados entre servidores. *XML* permite definir un conjunto ilimitado de etiquetas para datos (a diferencia de *HTML*, donde las etiquetas ya están definidas).

El marcado *XML*, aun siendo todavía un porcentaje pequeño en el Web, es el idioma que se esta usando de forma creciente para la codificación e intercambio de datos estructurados. *XML*, proporcionan las capacidades estructurales de las que *HTML* carece, haciendo fácil alcanzar los principios de modularidad y extensibilidad para metadatos que se mencionan en *Dual (2002, p.2)*.

XML Schemas y XML Namespaces

La noción de “*espacios de nombres*” (*namespaces*) es una parte fundamental de la infraestructura del Web (y en particular de *XML*) y permite la identificación única de elementos en el Web. Los “*espacios*

¹⁹ <http://www.w3.org/XML/>

²⁰ <http://www.w3.org/Consortium/>

de nombres” son dominios de nombres asociados con elementos que junto con el nombre del elemento produce un *URL* que identifica unívocamente el elemento.

Como se mencionó en la sección 1.3.2.5 las declaraciones de espacios de nombres junto con la infraestructura *XML*, permiten la construcción de conjuntos mezclados de metadatos dentro de un perfil de aplicación (Duval, 2002, p.7).

El término *XML Schema* se refiere en el contexto de *XML* a la definición de la estructura de una clase de documentos *XML*. Un *XML Schema* describe el ordenamiento e interrelaciones de elementos y atributos *XML* dentro de la clase de documentos a la cual el esquema se aplica. También provee definiciones de tipos de datos que se permiten en las instancias de documentos de una clase en particular. (Ahmed, 2001, p.27).

La especificación de *XML Schemas*, también define un lenguaje (esquemas) de representación que permite la especificación de *perfiles de aplicación (application profiles)* que incrementan las posibilidades de interoperabilidad. En la sección 1.3.2.3 también se habló brevemente de este tema.

*RDF (Resource Description Framework)*²¹

La iniciativa *RDF* consiste en una especificación, actualmente bajo desarrollo, elaborada dentro de las actividades relacionadas con metadatos de la organización W3C (Tauber, 2002). *RDF* es un mecanismo que provee interoperabilidad entre aplicaciones que intercambian información entendible por computadoras (*machine understandable*) usando los mecanismos de Internet y el WWW. Esta especificación hace énfasis en las facilidades que permitan el procesamiento automatizado de recursos Web.

RDF provee las siguientes características:

- Interoperabilidad de metadatos.
- Semántica para metadatos entendible por computadoras.
- Mayor precisión en ubicación de recursos (*resource discovery*) que las búsquedas de texto completo.

En el futuro *RDF* proveerá de capacidad de consultas uniformes y estándares, un lenguaje de procesamiento de reglas para toma de decisiones automatizadas que tengan que ver con recursos Web, y un lenguaje de recuperación de metadatos.

En términos generales, *RDF* provee una base para herramientas genéricas de autoría, manipulación y búsqueda de información entendible por computadoras en el Web, promoviendo por lo tanto la transformación del Web en un repositorio de información “procesable automáticamente” (Berners-Lee, 2003). La iniciativa *RDF* esta enmarcada a su vez dentro de lo que se ha llamado “*El Web Semántico*” (W3C-SW, 2003). Los promotores de esta idea lo definen como:

“...una extensión del Web actual en el cual la información posee un significado bien definido, permitiendo que las computadoras y la gente trabajen en cooperación” (Berners-Lee, 2001)

La base de funcionamiento del Web Semántico es que las computadoras deben tener acceso a colecciones estructuradas de información y conjuntos de reglas de inferencia que se pueden usar para conducir algún tipo de procesamiento racional de la información por parte de las computadoras.

²¹ <http://www.w3.org/RDF/>

1.3.3 Estándares para datos ambientales, biológicos, ecológicos y geo-espaciales

En un estudio piloto sobre el manejo de datos ambientales encargado por el gobierno de los Estados Unidos a un comité de especialistas, (US-CODATA, 1995, p.4) se destaca que dicho comité ha encontrado que las diferencias entre convenciones científicas usadas por diferentes disciplinas puede ser un severo impedimento para el uso y aprovechamiento de los datos, incrementando significativamente los costos de intercambio e integración entre diferentes colecciones de datos, y en algunos casos impide completamente este intercambio. Este estudio, entre las recomendaciones generadas, definió la siguiente: (Recomendación #4 del estudio) “*Los esfuerzos para establecer estándares de datos deben enfocarse en un conjunto clave de parámetros comunes cuya estandarización debería facilitar el intercambio y procesamiento de datos...*”. A continuación se describen algunos de los estándares para datos y metadatos científicos más importantes y usados hoy en día por diferentes comunidades de científicos.

1.3.3.1 FGDC²²

Uno de los conjuntos de elementos de metadatos más desarrollados y usados es el “*Estándar de Contenidos para Metadatos Geoespacial Digital*”²³ (CSDGM por sus siglas en inglés), desarrollado en 1995 por el “*Federal Geographic Data Commite*” de los Estados Unidos (FGDC) como parte de los esfuerzos de la “*National Information Infrastructure*”. Las colecciones de datos geoespaciales incluyen datos topográficos y demográficos, archivos generados o manejados por GIS (*Geographic Information Systems*), y archivos provenientes de aplicaciones de cartografía computarizada. Este estándar ha sido usado en una amplia variedad de áreas, que incluyen estudios de uso de suelo y aguas, biodiversidad, climatología y estudios de cambio global, percepción remota e imágenes de satélite. El estándar de contenidos de FGDC es de uso obligatorio para recursos de información y datos obtenidos con fondos del gobierno de los Estados Unidos, y también es usado internacionalmente.

1.3.3.2 NBII²⁴

El Perfil para datos biológicos (*Biological Metadata Profile*)²⁵ tiene como propósito proveer una extensión o perfil temático del FGDC-CSDGM para incrementar su utilidad al documentar datos y recursos de información provenientes de las ciencias biológicas. Esto incluye datos biológicos que son explícitamente geo-espaciales por naturaleza, así como datos que no son geoespaciales (como por ejemplo datos resultantes de investigación de laboratorio).

Este estándar soporta el uso e intercambio de datos biológicos entre usuarios a nivel internacional. También sirve como estándar de metadatos para la *National Biological Information Infrastructure* (NBII), que es un programa llevado a cabo por el gobierno de los Estados Unidos para incrementar la disponibilidad y acceso a datos e información de los recursos biológicos de la nación (es parte a su vez de la iniciativa NII: *National Information Infrastructure*²⁶). Este estándar es mencionado como uno de los principales ejemplos de un “*perfil de aplicación*” (ver sección 1.3.2.3) (UKOLN, 2001). Este perfil hereda todos los elementos de FGDC-CSDGM, pero incluye elementos adicionales que son capaces de documentar información acerca de taxonomía y nomenclatura de especies.

²² <http://www.fgdc.gov/>

²³ <http://www.fgdc.gov/metadata>

²⁴ <http://www.nbio.org/>

²⁵ http://www.fgdc.gov/standards/status/sub5_2.html

²⁶ <http://www.ibiblio.org/nii/>

1.3.3.3 Darwin Core (DwC)²⁷

Darwin Core (DwC) es un perfil que describe un mínimo conjunto de estándares para búsqueda y recuperación de colecciones de historia natural y bases de datos de observaciones de vida natural. *Darwin Core* es una de una serie de herramientas desarrolladas por "*The Species Analyst*", un proyecto de investigación que está desarrollando estándares y herramientas de software para dar acceso a las colecciones y bases de datos de historia natural existentes en el mundo. Este proyecto está basado en el "*University of Kansas Natural History Museum and Biodiversity Research Center*". Como se destaca en el sitio Web de DwC²⁸ "existe un factor común entre el contenido de casi todas las colecciones y bases de datos de observaciones que pueden ser explotadas para realizar búsquedas y recuperaciones ordenadas de estas colecciones de datos diversas. El DwC intenta proveer un conjunto de guías para aprovechar estas características comunes, sin importar el mecanismo usado para almacenar los contenidos de los registros".

A pesar de que este perfil fue originalmente diseñado para ser usado junto con el protocolo Z39.50 a través de servidores Z39.50, se puede también aplicar para definir contenidos y búsquedas *XML*, generadas por bases de datos usando el protocolo *HTTP*.

1.3.3.4 GCMD-DIF²⁹

Los metadatos usados por el "*Global Change Master Directory*" (*GCMD*) son un conjunto de atributos que ayudan a los usuarios a determinar si una colección de datos reúne sus requerimientos particulares para ser usados. El conjunto de atributos (campos) y su sintaxis asociada es conocida como el "*Directory Interchange Format*" (*DIF*). Este formato ha evolucionado en un periodo de 20 años y ayuda a la comunidad de usuarios en el descubrimiento y análisis de datos relacionados con las Ciencias de la Tierra (*GCMD, 2000*).

El formato del estándar *DIF* es compatible con el estándar *FGDC-CSDGM*. Esto permite que parte del contenido sea transformable (*mapped*) a *DIF* y viceversa. Para esto ocupa un programa llamado *DocMorph* que es capaz de convertir registros a formato *FGDC-CSDGM*. También a partir de *DIF* se puede ingresar y exportar metadatos de varios formatos. Actualmente, el proyecto *GCMD* está considerando realizar todos los "mapeos" o transformaciones a otros estándares, a través del uso de transformaciones *XSL* (eXtensible Stylesheet Language) (*GCMD, 2000*). En (*Day, 2002*) se encuentran referencias a sitios Web *sobre mapeos* o transformaciones (también se pueden llamar filtros) entre *DIF* y otros formatos y viceversa.

1.3.4 Estándar de metadatos usado en proyecto LTER

1.3.4.1 Estándar de Metadatos para datos no-geoespaciales

Los estudios ecológicos a menudo requieren la colección de muy diversos tipos de datos, incluyendo atributos que caracterizan y cuantifican el ambiente químico y físico, la fisiología de los organismos, la dinámica de población y de los ecosistemas, la composición de la comunidad, la estructura del paisaje, etc. Para este tipo de datos las coordenadas geográficas de los sitios de muestreo, a menudo no son relevantes. Sin embargo hasta ahora los mayores esfuerzos de estandarización han sido puestos en los datos geo-espaciales (*Michener, 1997, p.1*).

²⁷ http://tsadev.speciesanalyst.net/DarwinCore/darwin_core.asp

²⁸ http://tsadev.speciesanalyst.net/DarwinCore/darwin_core.asp

²⁹ <http://gcmd.gsfc.nasa.gov/>

En el reporte final del *Comité para el Futuro de los Datos Ecológicos a Largo Plazo (Future of Long-term Ecological Data - FLED)*³⁰ establecido por la *Ecological Society of America (ESA)* en 1993, se mencionaba la falta de estándares para datos ecológicos no-geoespaciales, y recomendaba el desarrollo de estándares de metadatos como parte de planes de desarrollo para un archivo de datos ecológicos disponible electrónicamente.

Michener (1997, p.2), a su vez argumenta que es poco probable que un solo estándar de metadatos, sin importar cuán comprensivo y completo pueda ser, sea capaz de abarcar todos los tipos de datos ecológicos, apoyándose en su complejidad. Como consecuencia de esto, y dentro de los esfuerzos de desarrollar la colaboración inter-sitios, Michener (y otros) propusieron en 1997, un conjunto genérico de descriptores de metadatos no-geoespaciales para las ciencias ecológicas (Michener, 1997).

Cada uno de los estándares mencionados en la sección 1.3.3 tuvo su origen y se desarrolló para cubrir diferentes necesidades. Por ejemplo, los estándares para el proyecto LTER siguen los lineamientos del FGDC, así como de otros organismos, pero adaptados a los requerimientos de los sitios individuales participantes en la red LTER. Siguiendo un enfoque comparativo, se les pidió a los sitios participantes que enviaran copia de los estándares que habían seguido (antes de su incorporación a la red LTER), se compararon y se formó una lista con los elementos comunes que luego formaron la base del estándar. Se agregaron elementos necesarios para el funcionamiento interconectado de los sitios. El estándar resultante se comparó con otros de uso común (en la sección 1.3.3 se hablará de varios de los estándares de mayor uso) y se trató de adaptar en lo posible a estos (Porter, 1997, pp.2,3).

La construcción de estándares para metadatos enfrenta numerosos problemas y dificultades. Por ejemplo surgen preguntas como la siguiente: ¿Cómo se puede determinar qué metadatos son absolutamente esenciales para el estándar?. Para contestar una pregunta como esta se puede hacer uso del proceso que siguen los científicos para adquirir y utilizar colecciones de datos ya existentes, lo que nos proporcionará una guía de cuales metadatos pueden ser requeridos. Por ejemplo, una vez que un científico reconoce la necesidad por un dato específico, debe buscar respuestas a varias preguntas, o seguir algunos pasos de manera ordenada:

1. ¿Qué colecciones de datos existen que cumplan con determinados objetivos?
2. ¿Porque esas colecciones de datos fueron recolectadas, y se adaptan a mis necesidades particulares?
3. ¿Esas colecciones de datos pueden ser obtenidos? Si es así ¿Cómo?
4. ¿Cómo están organizados y estructurados los datos?
5. ¿Existe información adicional que pueda facilitar que yo use e interprete los datos?

Estas cinco preguntas que se realizan (o deberían realizarse) los científicos durante la identificación, adquisición y utilización de los datos sirvió como base para crear cinco clases de descriptores que se usaron en la definición del estándar seguido por el proyecto LTER. (Michener, 1997, p.335). Resumiremos a continuación dichas clases de descriptores.

1.3.4.2 Descriptores Genéricos de Metadatos para las ciencias ecológicas (LTER)

Como se mencionó en la sección 1.3.4, Michener sugirió una lista de descriptores de metadatos para que pudiera ser usada como patrón y que podría servir como la base de definiciones de metadatos para sub-disciplinas o proyectos específicos (Michener, 1997)

Se delinearon en el trabajo citado anteriormente, cinco categorías (clases) de descriptores de metadatos:

³⁰ http://www.sdsc.edu/~ESA/FLED/VOL1_EXC.HTM

- I) Descriptores de la colección de datos
- II) Descriptores de la investigación que origino los datos
- III) Accesibilidad y condición de la colección de datos
- IV) Descriptores de la estructura de los datos
- V) Descriptores suplementarios

La Clase I incluye atributos básicos del conjunto de datos: título de la colección de datos, científicos asociados, resumen, y palabras clave. El propósito de los descriptores de la Clase I es el de alertar sobre la existencia de conjuntos de datos que caen dentro de dominios temporales, geográficos, espaciales y temáticos específicos. Se puede incluir en el resumen el contexto científico en el que fueron capturados los datos y los potenciales usos de la colección de datos, lo que ayudara a potenciales usuarios secundarios a identificar si los datos se adaptan a sus necesidades.

La Clase II incluye todos los metadatos relevantes que describe el proyecto o investigación que llevo a cabo el desarrollo de ese conjunto de datos. Se incluyen dos sub categorías; una relativa al proyecto global que usara los datos (descripción, propósitos del proyecto) que sirve para proveer un contexto más amplio; y otra relativa a la investigación y proceso de captura de datos (características del sitio de captura, diseño del muestreo, métodos de investigación, y personal del proyecto).

Los metadatos de la **Clase III** describen la condición de la colección de datos y de los metadatos asociados, así como información relacionada a la accesibilidad de los datos. Por ejemplo, puede haber restricciones relativas a derechos de autor, restricciones gubernamentales, o requerimientos relativos a mencionar y dar crédito a los entes generadores o patrocinadores de las investigaciones que originaron los datos.

Los metadatos de la **Clase IV** describen todos los atributos relacionados a la estructura de los archivos de datos. Todas las variables deben ser etiquetadas, definidas y caracterizadas en cuanto a su tipo de dato y formato. También se documenta cualquier anomalía de los datos (datos perdidos o faltantes por ejemplo).

La Clase V incluye todos los metadatos e información necesaria para facilitar el uso secundario, publicación de la colección de datos, o soporte para auditoria de los datos. Por ejemplo un científico puede necesitar revisar o asegurarse de la calidad de los datos, puede requerir revisar los cuadernos de campo originales, o los programas/algoritmos usados para procesar los datos.

Se muestra una tabla con estos descriptores adaptada de (Michener, 1997, p.335) y (Porter, 1997)

Clases	Descriptores
I. Descriptores de la colección de datos	A. Identidad de la Colección de Datos B. Código de identificación de la colección C. Descripción de la colección <ol style="list-style-type: none"> 1. creador(es) 2. resumen D. Palabras clave
II. Descriptores de la investigación que origino los datos	A. Descripción del proyecto <ol style="list-style-type: none"> 1. identidad 2. creador(es) 3. periodo del estudio 4. objetivos 5. resumen 6. fuentes de financiamiento B. Descripción de subproyecto <ol style="list-style-type: none"> 1. descripción del sitio

	<ol style="list-style-type: none"> 2. diseño del muestreo y el experimento 3. métodos de investigación 4. personal del proyecto
III. Estado y accesibilidad de la colección de datos	<p>A. Condición</p> <ol style="list-style-type: none"> 1. última actualización 2. última fecha de archivo 3. condición de los metadatos 4. verificación de entrada <p>B. Accesibilidad</p> <ol style="list-style-type: none"> 1. localización y medios de almacenamiento 2. persona(s) contacto 3. restricciones de copyright 4. restricciones propietarias 5. costos
IV. Descriptores de la estructura de los datos	<p>A. Archivo de la colección de datos</p> <ol style="list-style-type: none"> 1. identidad 2. tamaño 3. formato y modo de almacenamiento 4. información de encabezados 5. atributos alfanuméricos 6. caracteres y campos especiales 7. procedimientos de autenticación <p>B. Información de la variable</p> <ol style="list-style-type: none"> 1. identidad de la variable 2. definición de la variable 3. unidades de medida 4. tipo de dato 5. formato de datos <p>C. Anomalías en los datos</p>
V. Descriptores suplementarios	<p>A. Captura de datos</p> <ol style="list-style-type: none"> 1. formularios de datos usados 2. localización de formularios completados 3. procedimientos de verificación <p>B. Procedimientos de QA/QC</p> <p>C. Materiales relacionados</p> <p>D. Programas y algoritmos de procesamiento</p> <p>E. Archivo</p> <p>F. Publicaciones</p> <p>G. Historia del uso de la colección de datos</p>

Tabla 1. Descriptores de metadatos de Michener (1997)

1.3.5 Ecological Metadata Language: EML.

En 1997, usando como guía el artículo de Michener (1997) que definía descriptores de metadatos relevantes para las ciencias ecológicas (explicados en la sección anterior); investigadores del "National Center for Ecological Analysis and Synthesis" (NCEAS)³¹ comenzaron la implementación de la primera versión del "Ecological Metadata Language" (EML), que tenía como objetivo principal, definir un estándar y formato común que pudiera ser generado fácilmente a partir de metadatos existentes y facilitar la creación de colecciones de datos correctamente descritas y documentadas. Esta primera versión fue revisado varias veces hasta llegar a la versión 1.4.1. Esta versión representaba con exactitud las clases de metadatos definidas por Michener en su artículo de 1997.

A partir de esta versión inicial se hizo evidente que se necesitaba una revisión mayor, que incrementara su usabilidad y flexibilidad para la comunidad de científicos ecológicos. El proyecto

³¹ <http://www.nceas.ucsb.edu/>

"*Knowledge Network of Biocomplexity*" (KNB)³² comenzó los esfuerzos para revisar la especificación EML; y como resultado de estas revisiones fue liberada en el año 2001 la versión 2.0 Beta, para que fuera revisada, probada y considerada por la comunidad. En Diciembre del año 2002 fue liberada la versión 2.0 definitiva del estándar EML, que está siendo adoptado por todos los miembros del proyecto LTER (Michener, 2002, p.3).

1.3.5.1 Principios de diseño de EML

El desarrollo de EML ha seguido varios principios guía de diseño (McCartney, 2001, p.2)

- 1) Debe estar codificado en un formato procesable automáticamente ("*machine parseable*"), con fuerte soporte por parte de la comunidad científica y académica, y debe ser desarrollado teniendo en mente la independencia a software o plataformas específicas.
- 2) El extenso trabajo previo en estándares de metadatos dentro y fuera de la ecología debería ser usado como base, para mejorar la compatibilidad y reducir la redundancia.
- 3) El estándar deberá servir para integrar, en lugar de dictar soluciones para un sitio individual para crear, almacenar y manejar metadatos.

En la especificación formal de EML³³ se resaltan las siguientes características generales de EML versión 2.0.0:

Modularidad: diseñado como una colección de módulos básicos extensibles en lugar de un solo estándar grande y complejo. Esto quiere decir que se pueden añadir nuevos módulos para describir un recurso o tipo de datos muy específico.

Estructura detallada: en EML la información dividida en elementos altamente estructurados para facilitar procesamiento automático cuando se requiera.

Compatibilidad: EML adopta gran parte de su sintaxis de otros estándares de metadatos. Durante su desarrollo se tomaron en cuenta los siguientes estándares: *Dublin Core Metadata Initiative*, el "*Content Standard for Digital Geospatial Metadata*" de FGDC (CSDGM), y el perfil biológico de este estándar (FGDC-NBII), el estándar de información geográfico de ISO, "*International Standards Organization's Geographic Information Standard*" (ISO 19115), el "*ISO 8601 Date and Time Standard*", el *OpenGIS Consortium's Geography Markup Language (GML)*; *Scientific, Technical, and Medical Markup Language (STMML)*, y el *Extensible Scientific Interchange Language (XSIL)*.

Definición estricta de tipos de datos (Strong Typing): la estructura de EML está modelada en un lenguaje XML conocido como "*XML Schema*", que define las reglas que gobiernan la sintaxis de EML (cardinalidad, anidamiento de elementos, etc.). Esto permite la validación detallada tanto de la estructura como de los contenidos de un elemento. En las primeras versiones de EML se usaban DTDs³⁴ para validar, pero en la versión 2.0 definitiva solo se usan "*XML Schemas*".

1.3.5.2 Diseño y estructura de EML (2.0.0)

³² <http://knb.ecoinformatics.org>

³³ <http://knb.ecoinformatics.org/software/eml/eml-2.0.0/index.html>

³⁴ DTD: *Document Type Definition*. Documento que define la estructura de un archivo XML

EML, de una manera modular y extensible sirve para documentar datos ecológicos a través de una serie de tipos de documentos (esquemas) *XML*. Cada módulo describe una parte lógica de los metadatos que deberían ser incluidas con todas las colecciones de datos ecológicos.

Entre la características de diseño de *EML*, podemos mencionar:

1. Es *modular*, con esquemas separados que definen conjuntos de descriptores que los relacionan a categorías específicas de información.
2. Usa *tipos complejos de Esquemas XML* para lograr una aproximación de diseño orientada por objetos (OxO), en la cual clases abstractas son definidas y luego extendidas para crear variaciones específicas. Usando esta aproximación, *EML* define varios tipos de recursos de información, incluyendo "*dataset*" y "*literature*" que heredan un conjunto común de elementos que corresponden a los elementos básicos de identificación que se encuentran en muchos estándares de metadatos.
3. Es *extensible* por la capacidad de enlace entre múltiples módulos dentro de una colección de datos. Características derivadas de *XML* y de *RDF* permiten definir relaciones "sujeto-objeto" entre documentos de metadatos sin modificar los esquemas de módulos existentes.
4. Los módulos *EML* están organizados para *separar la descripción del contenido "lógico" de la información, de su instancia "física"*. Esta característica permite que los usuarios separen la información de la colección de datos, de los detalles del formato físico en que están codificadas o almacenadas, y simplifica el mantenimiento de los metadatos a medida que los formatos de los medios de almacenamiento o su ubicación cambian a través del tiempo. Un paquete de datos *EML* consistirá de un documento como *eml-dataset* o *eml-literature*, más cualquier otro modulo asociado, y opcionalmente objetos de datos (archivos de datos).

En el gráfico siguiente (Fig.4) se muestra la estructura modular de una de las clases (módulos) principales de *EML*: *EML-dataset* (McCartney, 2001, pp2).

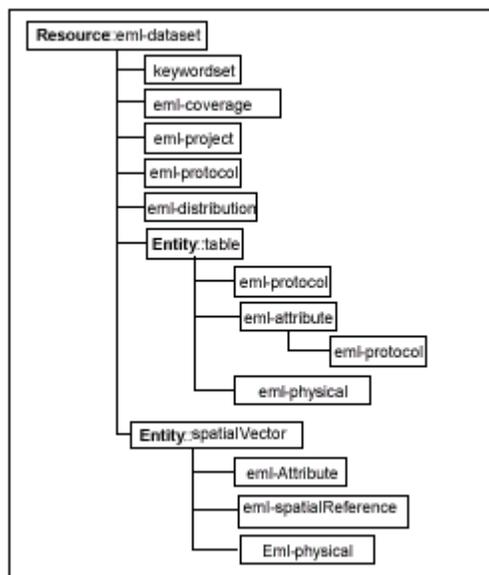


Fig.4. Estructura de la clase *eml-dataset*

1.3.5.3 Creación de Metadatos EML

Se están desarrollando actualmente varias herramientas de manejo de *EML*, destacándose el *Morpho*³⁵, que permite a los usuarios, insertar, editar, consultar, y recuperar documentos *EML*. Los documentos *EML* generados, también pueden ser enviados y almacenados en un servidor centralizado (*Metacat*)³⁶, que usa aplicaciones que también están siendo desarrolladas por el proyecto *KNB*. En el capítulo 2 de este trabajo y como parte de las recomendaciones en cuanto a métodos de creación, edición y distribución de datos y metadatos, se describirán en detalle las características de algunas de estas herramientas.

1.3.5.4 Implantación de EML en LTER

Actualmente el proyecto LTER, que es uno de los desarrolladores y promotores de este estándar, está adoptando su uso generalizado entre los sitios miembros.

En un reporte de Agosto del 2001³⁷, el grupo de trabajo encargado de definir las políticas y procedimientos relacionados con el manejo de los metadatos dentro de LTER, estableció que dos de las principales tareas de cada sitio en un futuro cercano, iban a ser las siguientes:

1. La reestructuración de los contenidos existentes a *EML*
2. El desarrollo de nuevos contenidos “compatibles” con *EML*

En esa época los encargados del tema dentro de *LTER*, pensaron que podían documentar una colección de datos a diferentes niveles, cada uno incorporando más capacidad y compatibilidad con *EML*. Esto quería decir que se podía comenzar a documentar una colección de datos con algunos de los elementos de *EML*, para luego ir añadiendo el resto, lo que iría agregando nuevas características y capacidades a las colecciones de datos. En ese mismo reporte se identificaron las diferentes situaciones

³⁵ <http://knb.ecoinformatics.org/software/morpho>

³⁶ <http://knb.ecoinformatics.org/software/metacat/>

³⁷ http://www.vcrlter.virginia.edu/nis/im2001/LTER_Metadata_Effort.htm

en que se encontraban los metadatos ya existentes y se dividieron en tres grupos (*no procesables, procesables y metadatos almacenados en bases de datos relacionales*), y se definieron estrategias generales para atacar cada caso, incluyendo recomendaciones de herramientas a usarse.

En Enero y Junio del 2002 se llevaron a cabo dos talleres (los primeros de varios otros planificados) que tenían como objetivo que los participantes migraran los metadatos de sus sitios a los estándares establecidos (*EML 2.0* principalmente). Se formaron grupos que aprendieron en cada caso a: realizar conectores de datos *EML* para bases de datos relacionales, a traducir metadatos en texto a *EML*, o crear nuevos metadatos *EML* desde cero. Después del taller, todos los sitios participantes estuvieron de acuerdo en migrar sus metadatos lo antes posible al nuevo estándar. En una reunión del comité de gerentes de información de LTER en Abril del 2002, mencionaban que esperaban que luego del segundo taller de metadatos en el verano del 2002, todos los sitios participantes fueran capaces de proveer metadatos formateados con *EML* para finales del 2002.

En diciembre del 2002 se liberó la versión definitiva del estándar *EML 2.0.0*, y se espera que todos los sitios pertenecientes a LTER lo adopten en un tiempo breve.

Para finalizar este capítulo diremos que la implementación y uso de un estándar como *EML* por parte de cualquier red de científicos, facilitaría grandemente, el uso compartido y la síntesis de datos entre sitios recolectores, en la medida que se desarrollen herramientas para compartir, integrar, y analizar datos y metadatos que hayan sido documentados usando como base y herramienta a este estándar.

2 *Propuestas para el manejo de datos ambientales y ecológicos.*

2.1 **Objetivos de la investigación y definición del problema.**

Según (UNDP-UNSO, 2001). "*uno de los problemas más serios en la creación, integración, uso y manejo de grandes bases de datos para investigación interdisciplinaria es la falta de metadatos adecuados*". El objetivo general de este trabajo es definir recomendaciones generales y una serie de métodos de trabajo que sirvan para resolver el problema del manejo de datos y metadatos relacionados con las ciencias ecológicas. En particular se proponen aquí estrategias y métodos prácticos que guíen la creación y uso de metadatos estructurados, basados en el trabajo de *W. K. Michener (Michener, 1997)* y la Red *LTER* de EUA (*LTER, 2003*) así como el uso de los estándares y herramientas más adecuados que permitan modelar, almacenar, recuperar, reusar e intercambiar los recursos y objetos de información descritos por dichos metadatos.

Este trabajo propone estrategias, métodos y herramientas que sirvan como punto de partida a desarrollos de sistemas de información y aplicaciones relacionadas con la ecología, biología y medio ambiente.

Se muestran como ejemplos de aplicación, prototipos y herramientas que aplican y aprovechan los estándares y tecnologías investigadas en este trabajo. En particular se muestra la aplicación de la investigación en el manejo de los datos provenientes de algunas estaciones meteorológicas que formaran en el futuro la *Red de Estaciones del Edo. Mérida*, así como las posibles recomendaciones para el manejo de datos y metadatos de la *Eco-Red Venezuela*.

En las siguientes secciones de este documento, y como resultado de la investigación llevada a cabo durante este trabajo, se proponen una serie de recomendaciones divididas en dos grupos:

- Recomendaciones para "Gerentes de Información" (sección 2.2)
- Recomendaciones para desarrolladores de sistemas de información (sección 2.3)

También se aclaran conceptos que deben comprender los gerentes de información y demás actores involucrados en el problema.

2.2 **Conceptos básicos y recomendaciones generales para "Gerentes de Información"**

Como parte de los resultados de esta investigación se explican a continuación conceptos generales y se proponen algunas recomendaciones que deben ser tomadas en cuenta por los gerentes de información involucrados en proyectos que tengan que ver con sistemas de información y manejo de datos ecológicos y científicos.

Se definen en las siguientes secciones, recomendaciones y conceptos divididos en cuatro temas:

- Definición de roles y acciones en el manejo de datos y metadatos (2.2.1)
- Definición, selección y uso de estándares y herramientas (2.2.2)
- Prueba, validación, control y aseguramiento de la calidad de los datos (2.2.3)

³⁸ <http://www.fgdc.gov/clearinghouse/clearinghouse.html>

³⁹ <http://www.gsdi.org/>

⁴⁰ Ver lista completa de herramientas XML en: <http://www.garshol.priv.no/download/xmltools/>

- Educación y entrenamiento (2.2.4)

2.2.1 Definición de roles y acciones en el manejo de datos y metadatos

Esta sección comienza con la definición de “gerente de información” en el contexto o dominio de los sistemas de información y redes de investigación que manejan datos ecológicos y ambientales o datos científicos en general para luego precisar los diferentes roles y acciones involucradas en el manejo de información..

2.2.1.1 Gerentes de información

Un “gerente de información” es la persona encargada de la captura, organización, distribución, conservación, etc. de recursos de información y colecciones de datos. En una organización que posee un nivel central de coordinación de una red de científicos y sitios o sub-organizaciones locales, y maneja los datos a nivel centralizado, deben existir dos tipos de gerentes de información; pueden existir, aparte de los científicos y usuarios de los datos, un *gerente de información central* y los encargados de la información producida en cada sitio o estación de captura de datos llamados *gerentes de información locales*. En la siguiente figura se esquematiza lo anterior (Fig. 5)

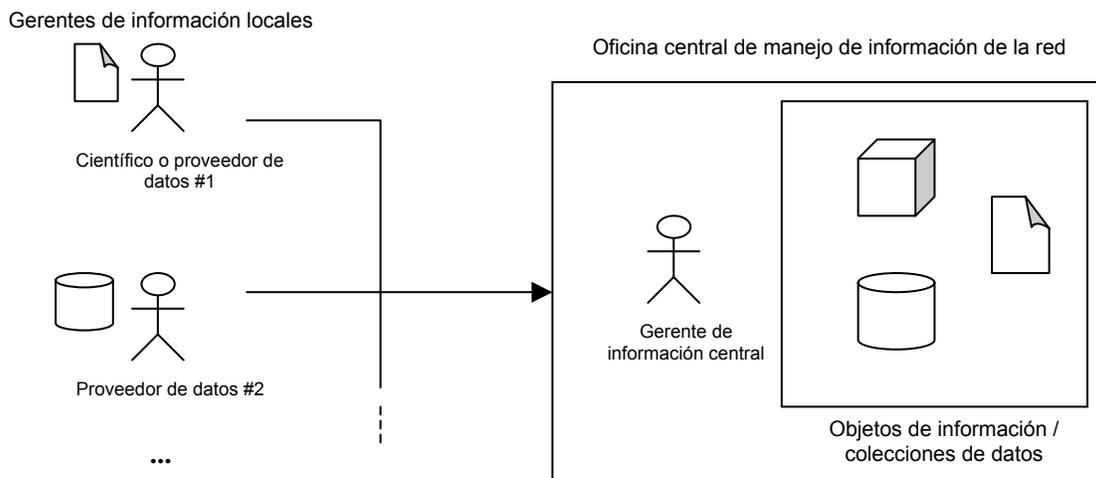


Fig.5. Roles en el manejo de metadatos.

El papel de *gerente de información local* lo puede ejercer un científico o una persona especializada en el manejo de información. A nivel centralizado, debe existir un especialista que ejecutara todas las labores de manejo de información necesarias (*gerente de información central*). Más adelante en esta sección se definen los roles y acciones de cada actor involucrado.

Los *gerentes de información* deben tener claro y entender los procesos involucrados en el manejo de datos y metadatos y los actores involucrados en dichos procesos. Según el modelo explicado en 1.3.2.1 estos procesos están definidos como “acciones” y a los “actores” se le deben asignar “roles”. Según Tomasic y Simon, (Tomasic, 1997) es conveniente ver el manejo de datos desde el punto de vista

de los procesos involucrados y los actores que toman parte en ellos, lo que coincide con el modelo de metadatos mencionado anteriormente.

En la siguiente figura (Fig. 6), se observa un modelo simplificado del ciclo de vida de los metadatos descrito en 1.3.2.4, que debe ser tenido en cuenta por los gerentes de información que están involucrados dentro de estos procesos ya que tendrán que participar, ellos mismos como actores, y manejar a los demás actores asignando roles o controlando acciones.

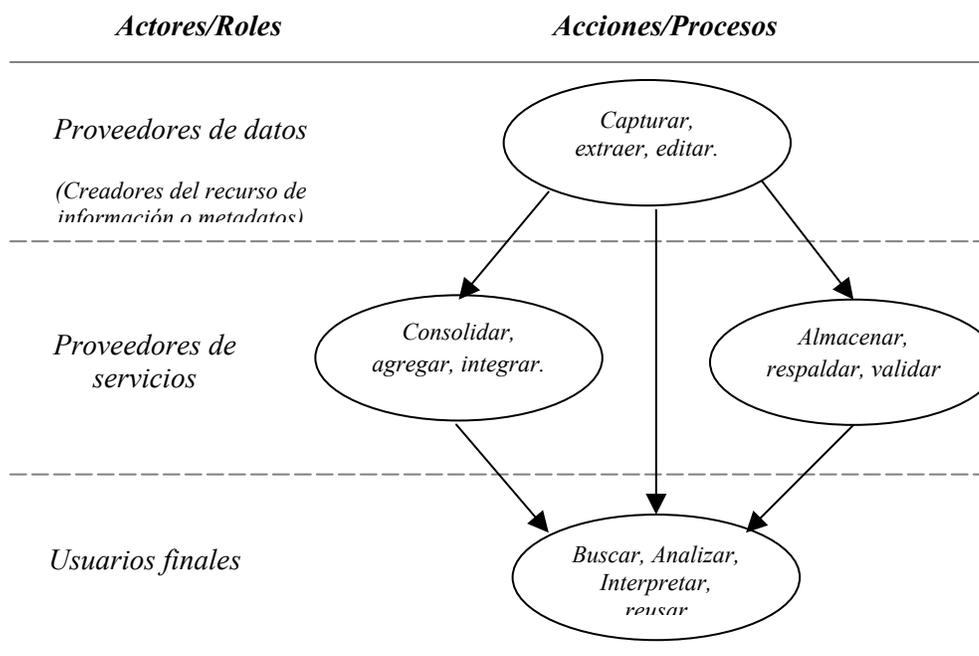


Fig. 6. Modelo simplificado del ciclo de vida de los datos y metadatos

Como se observa en la figura, en el modelo aparecen cuatro acciones principales (*capturar-extraer, consolidar-agregar, almacenar y Buscar-analizar-interpretar-reusar*) que se aplican a los recurso de información o metadatos, y tres tipos de actores (proveedor de datos, proveedor de servicios y usuario final), asociados a cada una de estas acciones. Los *gerentes de información* asumen a su vez algunos de los roles y funciones de los proveedores de datos y principalmente las funciones de los proveedores de servicios. Se describen a continuación cada uno de estos componentes en detalle.

2.2.1.2 Actores

Se describen los tres actores o roles que están involucrados en el modelo de manejo de metadatos descrito antes: el proveedor de datos, el proveedor de servicios y el usuario final. En este modelo los gerentes de información locales asumen el rol de proveedores de datos y los gerentes de información a nivel central asumen el papel de proveedores de servicios.

1. Proveedor de datos.

Genera, crea o captura el recurso de información (datos o metadatos). Por lo general es un científico que trabaja en un proyecto de investigación que necesita capturar y analizar datos. También puede ser personal técnico que esta encargado de los instrumentos de captura de los datos. En el caso de instrumentos automáticos de captura de datos, por ejemplo (estaciones meteorológicas automáticas, sensores remotos, etc.), estos también pueden ser considerados como proveedores de

datos. Es de resaltar que las personas más adecuadas para documentar los datos (generar los metadatos) pertenecen a este grupo.

2. Proveedor de servicios.

Ofrece a los usuarios servicios relacionados con el uso de los recursos de información o metadatos. Cuando se trata de distribución de información en formato digital, el proveedor de servicios realiza variadas tareas que tienen que ver con ayudar al usuario a identificar los contenidos y fuentes de información relevantes, servicios de seguridad y autenticación de los datos, servicios de catalogación, etc. (en la parte de “acciones” se listan más servicios).

3. Usuario final.

Usa (o re-usa) los metadatos y datos. Puede ser el mismo científico que los capturo, o puede ser un usuario secundario que necesita que los datos estén bien documentados (posean metadatos) para que pueda ubicar los datos y luego usarlos en su investigación.

2.2.1.3 Acciones

Los actores ejecutan acciones que se aplican tanto a los recursos de información como a los metadatos. A continuación una lista de acciones clasificadas de acuerdo al actor involucrado.

Acciones ejecutadas por los proveedores de datos (se refieren a las acciones aplicadas a los recursos de información o metadatos) :

- Crear el recurso o metadatos
- Escribir “
- Capturar “
- Preparar “
- Editar (combinar, ensamblar, actualizar, desensamblar, descartar, re-usar) “
- Almacenar “

Acciones ejecutadas por los proveedores de servicios (se refieren también a los recursos de información o metadatos):

- Mantener el recurso de información o metadatos
- Almacenar “
- Respalda “
- Integrar “
- Preservar “
- Archivar “
- Descartar “
- Validar “
- Asegurar la calidad “
- Distribuir “
- Asegurar o resguardar acceso “
- Contabilizar “
- Promocionar “
- Vender “

Acciones ejecutadas sobre los recursos de información por los usuarios finales:

- Especificar información requerida (definir consultas, definir estrategias de búsqueda, definir filtros)
- Localizar (buscar, localizar, ubicar, descubrir)
- Seleccionar (filtrar)
- Determinar derechos de uso (comprar, suscribirse)
- Establecer maneras de usar o acceder datos (adquirir herramientas y/o visualizadores, establecer lazos de comunicación)
- Acceder / recuperar la información
- Verificar (confirmar autenticidad, confirmar calidad, confirmar validez)
- Transformar
- Usar (leer, visualizar, re-usar, enviar)

2.2.1.4 Sistema de incentivos.

Uno de los elementos más críticos que influyen en el éxito o fracaso en la implantación y uso de sistemas de manejo de datos, es el desarrollo de un sistema de incentivos para los proveedores de datos y los gerentes de información. Lo anterior debe ir necesariamente acompañado del desarrollo de políticas de manejo y gerencia de la información, así como de código de buen uso y colaboración entre pares.

Porter (2001) por ejemplo, propone:

- Reconocimiento a los gerentes de información mediante asignación de co-autoría en trabajos e investigaciones que usen los datos.
- Publicación de colecciones de datos.
- Influencia de la publicación de datos en la obtención de fondos para la investigación (o publicación como condición de financiamiento).
- Cualquier premio o incentivo anterior debe tener como condición la calidad de las colecciones de datos.

En el informe [US-CODATA \(1995\)](#), se plantea muy claramente la necesidad de un sistema de incentivos y reconocimientos para los científicos y personas que contribuyan al trabajo interdisciplinario, ya que los sistemas vigentes usados en esa comunidad inhiben el trabajo colaborativo y la compartición de datos. Allí se proponen entre otras cosas, tratar la producción de colecciones de datos de manera análoga a publicaciones científicas, y recomiendan a las organizaciones involucradas re-evaluar las políticas de recompensa y trabajar colaborativamente con los científicos que están involucrados en el tema, para proveer fuertes incentivos y reducir las barreras para compartir datos. Adicionalmente recomiendan que los gerentes de información formen parte de los equipos de investigación desde el inicio de los proyectos y tengan igual condición que los investigadores principales (aunque en la práctica, muchas veces el propio investigador es el recolector de los datos o gerente de la información).

En cuanto a políticas de gerencia de la información, también según [Porter \(2001\)](#), se necesitan definir claramente:

Responsabilidades del proveedor de datos

- Aseguramiento de la calidad de los datos y productos de información
- Disponibilidad de los datos

Responsabilidades de proveedores de servicios, gerentes de información

- Mantenimiento de los sistemas informáticos

- Respaldo y disponibilidad de los datos (persistencia a largo plazo de los datos)

Responsabilidades del usuario final

- Atribución apropiada de autoría y asignación de créditos respectivos

2.2.2 Definición, selección y uso de estándares y herramientas

2.2.2.1 Estándares para metadatos

Todos los proyectos que involucren manejo de cantidades importantes de información deben definir y usar, desde un principio, convenciones científicas (estándares) que ayuden al manejo e intercambio de información. Como se mencionó en la sección 1.3.3, el uso de convenciones científicas comunes a varias disciplinas, es esencial para lograr un uso y aprovechamiento eficiente de las colecciones de datos.

Para el manejo de colecciones de datos científicos que tengan que ver con las ciencias ecológicas, se recomienda el uso de los descriptores de datos definidos por Michener (1997) (ver sección 1.3.4), que están siendo usados por el proyecto LTER y otras comunidades de científicos. Estos descriptores (clases) de metadatos se usaron para definir el *Ecological Modeling Language (EML)* que se recomienda como lenguaje estándar de representación de metadatos para colecciones de datos no-geoespaciales (ver sección 1.1.1).

El uso de estándares de representación de colecciones de datos asegurará en cierta medida, la interoperabilidad entre los diferentes sitios u organizaciones que los adopten. Por ejemplo todos los sitios que usen *EML* como lenguaje de representación de los metadatos de sus colecciones, podrán incorporarse a facilidades de búsquedas y consultas de dichas colecciones de datos gestionadas centralmente. También se podrán intercambiar o combinar los datos y metadatos con otros datos pertenecientes a otras redes que hayan adoptado las mismas convenciones y estándares.

Se recomienda revisar los siguientes recursos de información que contienen información valiosa relacionada con el tema de los estándares y manejo de información para ciencias ecológicas y ambientales:

Artículo: “*Metadata Descriptors for non-geospatial data*”. Michener, W. K., et al. 1997. . *Ecological Applications*. 7(1). 1997. pp 330-342.

Artículo importante donde se describen las clases de descriptores de metadatos usados como base para los estándares recomendados en este trabajo. Todos los trabajos posteriores de manejo de datos dentro de LTER tienen como base este. Se tratan temas desde la importancia de la preservación de los datos científicos, los problemas y retos que esto representa, y en particular se propone un sistema de descriptores para datos ecológicos, que forma la base del estándar *EML*.

Web LTER-USA: <http://www.lternet.edu>

Sitio Web donde se encontrara información detallada de todas las actividades y proyectos de esta red, así como artículos valiosos sobre el manejo de datos por parte de los miembros de esta red.

Web Knowledge Network for Biocomplexity: <http://knb.ecoinformatics.org>

Sitio Web principal del proyecto de desarrollo de *EML*, y sitio central de búsqueda de los datos del

proyecto LTER y de todas las organizaciones que usen las herramientas y estándares originados en esta red de científicos ecológicos.

Informe: “*Finding Forest in the Trees. The Challenge of Combining Diverse Environmental Data*”. Committee for a Pilot Study on Database Interfaces. U.S. National Committee for CODATA. National Academy Press. Washington, D.C. 1995.

Disponible a texto completo en el Web:

<http://www.nap.edu/openbook/0309050820/html/R1.html>

En este informe se realizó un estudio de casos exitosos en el manejo de datos ambientales (uno de ellos es el proyecto LTER), y basados en ellos, se definieron 18 recomendaciones que se deben tomar en cuenta en los proyectos que tengan que ver con la recolección, integración y análisis de datos ambientales.

Ponencia: “*The Metadata Landscape: Conventions for Semantics, Syntax, and Structure in the Internet Commons*”. Stuart L. Weibel, Senior Research Scientist, OCLC.

Disponible en el Web http://www.nfais.org/publications/metadiversity_preprints6.htm

Esta ponencia presentada en una conferencia de manejo de información sobre biodiversidad⁴¹, da una visión general sobre las convenciones y su estructura dentro del marco del descubrimiento de recursos de información en Internet y en un ambiente de información globalizado. Se definen entre otras cosas conceptos básicos sobre metadatos y su interoperabilidad entre comunidades con intereses diferentes.

2.2.2.2 Herramientas para metadatos

Existen diferentes herramientas que pueden ayudar en la creación y al manejo de colecciones de datos. Aquí nos concentraremos en las herramientas informáticas que fueron creadas especialmente para manejo de colecciones de datos científicos.

Los proyectos de estandarización del manejo de datos dentro de diferentes comunidades han creado o desarrollado a lo largo del tiempo, algunas herramientas para facilitar el trabajo de las comunidades involucradas (ver por ejemplo sitio web sobre herramientas para metadatos de FGDC, “*Metadata Tools*”⁴²). En el caso del proyecto LTER y el Proyecto KNB (*Knowledge Network of Biocomplexity*), se han desarrollado y se están desarrollando varias herramientas importantes para la creación, validación, almacenamiento, consulta y análisis de colecciones de datos. Dos de las principales herramientas que están siendo desarrolladas y usadas dentro de las tareas involucradas en el manejo de colecciones de datos en KNB y LTER son *Morpho*⁴³, y *Metacat*⁴⁴.

Morpho es una herramienta de creación edición, validación y documentación de colecciones de datos. Básicamente es una interfaz que es capaz de crear archivos *XML* de metadatos codificados según el estándar *EML*. *Metacat* es servidor o repositorio de metadatos y datos, que se puede conectar directamente desde *Morpho* o desde cualquier aplicación que implemente métodos de intercambio, que están basados en estándares de común uso en Internet. Ambas herramientas son de libre uso por cualquier persona, y se pueden obtener en Internet de forma gratuita. La comunidad LTER está desarrollando otras herramientas que permitirán realizar otras labores como descubrimiento de colecciones de datos en Internet, análisis y síntesis automático de datos, etc. Aunque las herramientas mencionadas juegan y jugarán un papel importante en el futuro de las redes de información asociadas a la ecología, biología y medio ambiente; el desarrollo de herramientas conformes a estándares como

⁴¹ <http://www.nfais.org/MetaDiversity1.asp>

⁴² *FGDC Metadata Tools*: <http://www.fgdc.gov/metadata/metatool.html>

⁴³ <http://knb.ecoinformatics.org/software/morpho>

⁴⁴ <http://knb.ecoinformatics.org/software/metacat>

EML es lo que permitirá una verdadera comunicación e intercambio de datos entre organizaciones con intereses comunes.

En la sección 2.2.2 se describen algunas de las herramientas de uso general para creación de metadatos estructurados, y se describen algunos detalles y funcionalidades de las herramientas que están siendo usadas específicamente por los miembros de la Red LTER para manejar colecciones de datos conformes con el estándar *EML*, entre ellas *Morpho* y *Metacat*. También más adelante, se mencionan los posibles usos de herramientas desarrolladas localmente, incluyendo las que se deriven del prototipo de sistema presentado en este trabajo de tesis (ver sección 0).

2.2.3 Prueba, validación, control y aseguramiento de la calidad (QA/QC) de los datos

Una de las acciones incluidas dentro del ciclo de vida de los metadatos y recursos de información es el control de calidad de dichos recursos. Según [Michener \(2001, p.2\)](#) los términos *QA/QC* se refieren a las estrategias diseñadas para prevenir la introducción de errores, o contaminación de los datos dentro de las colecciones. Se deben aplicar mecanismos de control de calidad (*QC*) durante el proceso de adquisición de datos que ayuden a identificar errores en la captura de los datos o el mal funcionamiento de los instrumentos de captura. Mecanismos de aseguramiento de la calidad (*QA*) son aplicados después que los datos son introducidos en un sistema de información para identificar errores de validez y consistencia de dichos datos.

La aplicación de mediciones de control de calidad durante la adquisición de datos puede reducir la contaminación de los datos. Por ejemplo, simplemente revisando dos veces los formularios de entrada de datos a medida que se van llenando, para confirmar que todos los campos han sido introducidos y que los códigos y medidas fueron introducidos correctamente, puede reducir los errores en gran medida. Obligar o forzar el uso de estándares para formatos, códigos, mediciones, campos de información, etc. ayuda a que los datos sean consistentes. Es altamente recomendable la aplicación de filtros y validaciones automáticas de rangos de entrada cuando los datos están siendo recolectados en un sistema de información computarizado.

Mediciones de aseguramiento de la calidad incluyen la revisión o búsqueda de patrones ilógicos entre los datos (consistencia) y crear patrones y resúmenes estadísticos. También se deben aplicar métodos y pruebas estadísticas para detectar valores extremos poco usuales. Estos valores pueden o no representar errores o contaminación de los datos, y deben buscarse explicaciones para estos ([Michener, 2001, p.3](#)). Se pueden comparar los resúmenes estadísticos actuales con datos de años anteriores para determinar tendencias centrales o variabilidad dentro de los datos que cambien notablemente. También científicos calificados pueden validar los datos haciendo revisiones en ellos, lo que incrementará la confianza en dichos datos.

Existen trabajos dentro de la Red LTER que apuntan a la búsqueda de métodos y herramientas para efectuar las labores de aseguramiento de la calidad de forma automática ([Nottrott, 2001](#)), donde se aprovechan las características de los datos estructurados usando *XML* y *EML* para un sistema de validación automático de colecciones de datos.

Algunas de las herramientas de manejo de colecciones de datos que se mencionan en la sección 2.3.3 ayudan en las tareas de control y aseguramiento de la calidad. Por ejemplo el programa *Morpho*, puede realizar validaciones al momento de crear las colecciones de datos, no permitiendo que un dato se deje de introducir estando definido como obligatorio. También se pueden definir los rangos de valores que pueden tomar los valores de los datos que están siendo incluidos en la colección, para así hacer notar los que estén fuera del rango establecido.

2.2.4 Educación y entrenamiento

En el informe [CODATA \(1995\)](#) del gobierno de EUA, se plantea la necesidad crítica de educar a los científicos acerca de los principios de gerencia de recursos de información y datos. En la recomendación número ocho de dicho informe se dice que “*las universidades deberían incluir cursos en sus currículos, que provean a los científicos ambientales de un entendimiento profundo de los principios y métodos involucrados con el manejo y gerencia de datos*”. Se dice también que se debería resaltar como la práctica del manejo de datos puede proveer beneficios inmediatos y a largo plazo para los científicos, particularmente aquellos involucrados en investigación interdisciplinaria ([CODATA, 1995, p.6](#)).

La gerencia y manejo de datos y metadatos dentro de una organización o red de científicos es una tarea que presenta variados retos y dificultades, por lo que es recomendable proporcionar a los involucrados el entrenamiento necesario para que se familiaricen con los conceptos básicos, herramientas, métodos, etc. El encargado de organizar, preparar y proporcionar este entrenamiento sería, en el caso de existir, la dependencia central coordinadora del manejo de datos de la red de científicos (mencionada en sección 2.2.1).

El *FGDC* por ejemplo, proporciona un completo juego de documentación disponible en Internet⁴⁵, que permite a las personas usar los estándares de metadatos, y los métodos y herramientas desarrollados por el gobierno de los EUA para el manejo de datos geoespaciales. Esta organización también promueve iniciativas de educación sobre metadatos para lograr integrar el uso de metadatos dentro de la vida y desarrollo profesional de las organizaciones que trabajan con datos geo-espaciales ([WYGIS, 1999](#)).

Otro ejemplo de documentación destinada a usuarios del estándar *FGDC*, diseñada para una comunidad en particular (oceanógrafos y científicos ambientales), es el manual de metadatos *del Centro de Servicios Costeros de la NOAA*⁴⁶ ([NOAA-CSC, 2003](#)) donde se intentan explicar de una manera sencilla todos los conceptos básicos y las herramientas que los científicos costeros deben usar para documentar, validar y publicar sus colecciones de datos.

Otras agencias gubernamentales de los EUA que hacen uso del estándar *FGDC* también proporcionan documentación que ayude a los que deban trabajar en el manejo de recurso de información. Algunas agencias estatales y locales proporcionan guías y manuales que ayudan en el trabajo con metadatos en sus dependencias, todo esto enmarcado dentro de la iniciativa “*National Spatial Data Infrastructure*” (NSDI)⁴⁷.

El proyecto que está desarrollando y promoviendo el uso de *EML* como lenguaje de intercambio de datos ecológicos (*KNB, Knowledge Network of Biocomplexity*)⁴⁸, junto con la red LTER, poseen diferentes mecanismos para ayudar a los miembros de LTER interesados en adoptar *EML*. Como se menciona en la sección 1.3.5.3 de este informe, desde principios del año 2001 el Grupo de tareas sobre metadatos de LTER (*LTER Metadata Task Group*)⁴⁹, está llevando a cabo talleres donde se enseña a los miembros, las diferentes aspectos del manejo de colecciones de datos.

Estos talleres también se han efectuado en diferentes países relacionados con la Red LTER Internacional. A su vez el *Comité de Gerentes de Información* de LTER (*Information Managers*

⁴⁵ <http://www.fgdc.gov/metadata/education.html>, <http://www.fgdc.gov/metadata/toollist/instru1.html>

⁴⁶ NOAA: National Oceanic and Atmospheric Administration. <http://www.noaa.gov>

⁴⁷ <http://www.fgdc.gov/nsdi/nsdi.html>

⁴⁸ <http://knb.ecoinformatics.org>

⁴⁹ <http://caplter.asu.edu/data/metadata/>

Committee) mantiene una publicación periódica disponible en Internet donde regularmente publica artículos sobre el tema de los metadatos (LTER Databits)⁵⁰

Cualquier programa de entrenamiento relacionado con manejo de colecciones de datos debería incluir estrategias para:

- Comprender la importancia del uso de estándares.
- Identificación de colecciones de datos a ser documentadas.
- Conversión de metadatos desde diferentes formatos y sistemas a EML.
- Uso de herramientas de documentación y creación de colecciones de datos.
- Diseño, análisis y optimización de consultas a colecciones de datos.

En la sección 3.3.3 de este informe y como uno de los productos de este trabajo de tesis, se propone un taller de manejo de datos para los miembros de la Eco-Red Venezuela, para los participantes en el proyecto de la Red de Estaciones del Edo. Mérida, y para cualquier científico o persona involucrada en el manejo de datos; que cubre algunos de los aspectos mencionados anteriormente.

2.3 Propuestas para desarrolladores de sistemas de información

Las siguientes sub-secciones están destinadas a las personas involucradas en el desarrollo de sistemas de información para las ciencias ecológicas. Se tratan temas que tiene que ver con el modelado de este tipo de sistemas, temas relacionados con el modelado y codificación de metadatos estructurados (en especial el estándar *EML*) y por ultimo se describirán las herramientas que están involucradas en el manejo de colecciones de datos desde el punto de vista informático. La comprensión de estos tópicos ayudara grandemente en cualquier esfuerzo de desarrollo de los tipos de sistemas de información tratados en esta investigación.

2.3.1 Arquitecturas de Sistemas de Información para datos ambientales y ecológicos

Se describen a continuación las diferentes tipos de arquitecturas que se proponen como base de desarrollo para sistemas de información para las ciencias ecológicas y ambientales y que sirvieron de modelo para el prototipo desarrollado durante este trabajo de tesis (en la sección 0 se describen las características detalladas de este prototipo).

2.3.1.1 Arquitecturas básicas para datos y metadatos

La forma de los metadatos que son procesados por una aplicación particular, puede ser categorizada de acuerdo a dos criterios generales (Ahmed, 2001; CEN, 1999):. El primero de estos criterios tiene que ver con la relación entre los metadatos y el recurso que describe; los metadatos pueden estar “incrustados” (*embedded*) dentro del recurso, o pueden ser “externos” al recurso. El segundo criterio especifica la relación entre los metadatos y la aplicación que los procesa o maneja. Una aplicación puede solo requerir trabajar con una sola fuente “centralizada” de metadatos, o puede requerir trabajar con fuentes de metadatos múltiples y distribuidas.

En la siguiente figura (Fig. 7) se observan tres posibles formas básicas o arquitecturas para sistemas de metadatos: a) los metadatos (M) son en si mismos datos que pertenecen al recurso de información (D); en b) se observa a los metadatos (M) como un conjunto disjunto de los datos (D) que describen; y en c) se muestra un ejemplo de “meta-metadatos” : o metadatos a nivel de colecciones (Mc) que

⁵⁰ http://lternet-183.lternet.edu/doc_archive/newsletters/databits/

documentan o describen dos colecciones (M1 y M2) relacionadas cada una con recursos de información o datos (D). En el presente trabajo de investigación, esta última forma o modelo de metadatos tiene especial relevancia, ya que es el modelo base de las arquitecturas de metadatos propuestas para el sistema prototipo desarrollado. Existen otras variaciones de modelado que pueden ser consultadas en (CEN, 1999, pp.10-12) y (Ahmed, 2001).

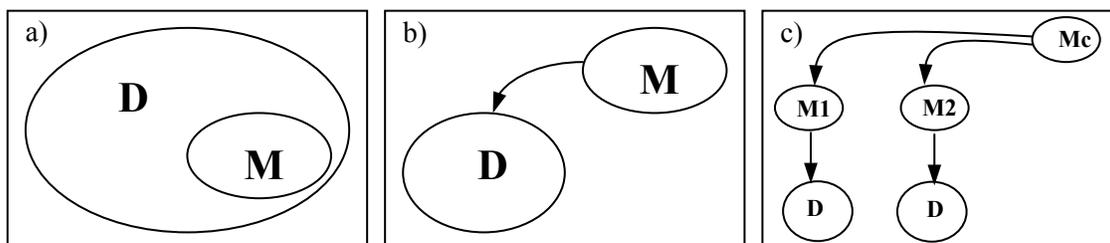


Fig. 7. Modelos básicos para sistemas de metadatos

En la siguiente sección describiremos un tipo de sistema de manejo de colecciones de datos difundido ampliamente entre diferentes comunidades científicas y gubernamentales: los “clearinghouse” de datos, cuya comprensión es fundamental para cualquier desarrollador de este tipo de sistemas de información.

2.3.1.2 Arquitecturas básicas de un “Clearinghouse” de datos

Un *clearinghouse de datos* según Nebert (1997) es “un sistema descentralizado de servidores disponibles en Internet que usa herramientas estándares de búsqueda y recolección de datos, a través del cual se pueden encontrar datos y metadatos agrupados según los intereses de una comunidad u organización científica, gubernamental, educativa, etc”. Cada servidor en un *clearinghouse* es un *miembro de la red*, a través del cual los usuarios obtienen servicios de información estandarizados y comunes a toda la red.

En el siguiente esquema (Fig. 8) se observa la aproximación básica de modelado para un sistema *clearinghouse* que permite el acceso a través de los metadatos, a colecciones de datos por parte un cliente solicitante de información.

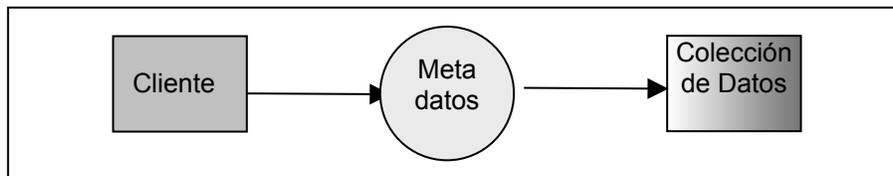


Fig. 8. Modelo básico de un sistema “Clearinghouse” de datos.

Algunas de las características básicas que comúnmente comparten este tipo de sistemas son las siguientes:

- Uso voluntario y por consenso de estándares internacionales.

- Se colocan a disposición de posibles usuarios, implementaciones de software libre para el manejo de este tipo de sistemas (para uso público o comercial).
- Promueven un vocabulario y mecanismos comunes para la búsqueda y ubicación (*discovery*) de datos en Internet.

En un *clearinghouse* de datos por lo general existen varios servidores de metadatos que son consultados a través de un cliente común (Fig. 9). Estos servidores de metadatos pueden dar acceso a las colecciones de datos, en forma de accesos a sistemas de bases de datos, sistemas de archivos, indicando la ubicación física de dichas colecciones o indicando los datos de los científicos y proyectos que poseen los datos.

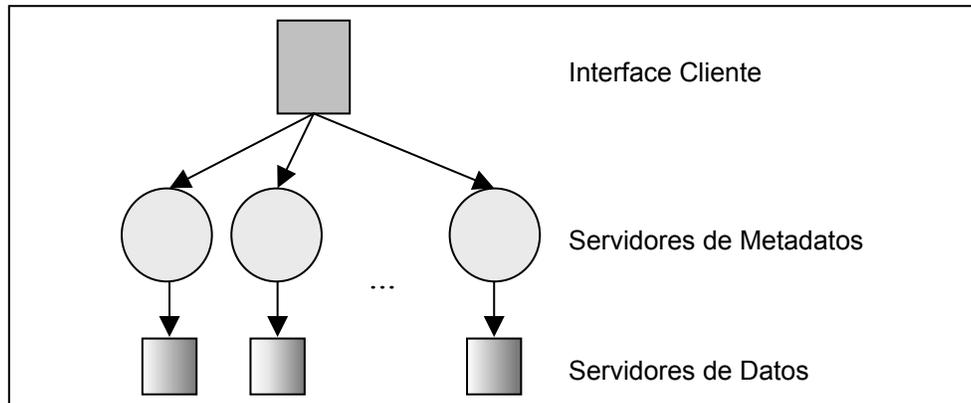


Fig.9 Modelo típico de un Clearinghouse de datos

También la base de datos que contiene los metadatos puede ser centralizada y permitir consultar o ubicar las colecciones de datos distribuidas en una o más bases de información, ubicaciones físicas, etc.

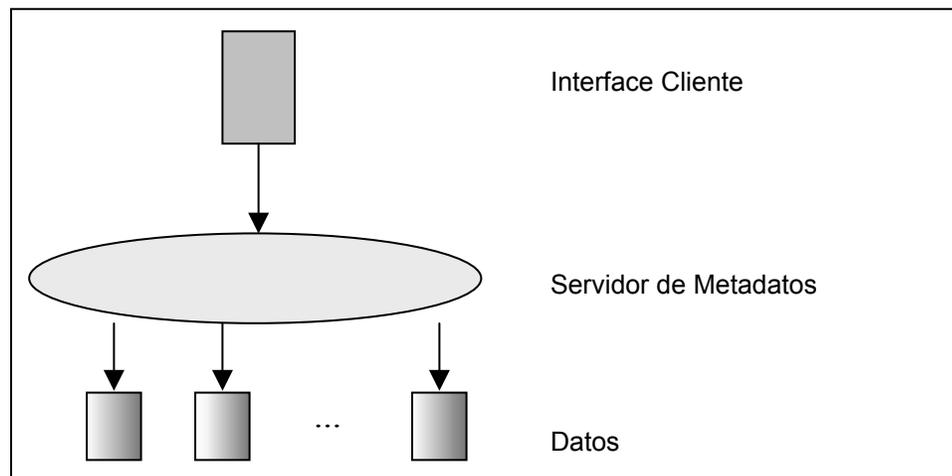


Fig.10. Clearinghouse centralizado de metadatos.

Es común (y recomendable) que exista un registro centralizado de *clearinghouses* que permita a los *clientes* ubicar a dichos servidores de metadatos. Este registro es mantenido por el grupo u organización que promueve el uso de dichos sistemas dentro de una comunidad con intereses compartidos. Este es el

caso por ejemplo del *clearinghouse de datos* del “Comité Federal de Datos Geográficos” (FGDC Clearinghouse)⁵¹, y de la “Infraestructura Global de Datos Geospaciales” de los EUA (*GSDI - The Global Spatial Data Infrastructure*)⁵² que permiten, a través de interfaces Web, consultar cualquier servidor de metadatos que haya seguido los estándares de dichas organizaciones. En las siguientes figuras se pueden observar un modelo de localización de fuentes de metadatos () y dos imágenes de la interfaz global de búsqueda de datos geospaciales de GSDI que usan este paradigma y arquitectura ().

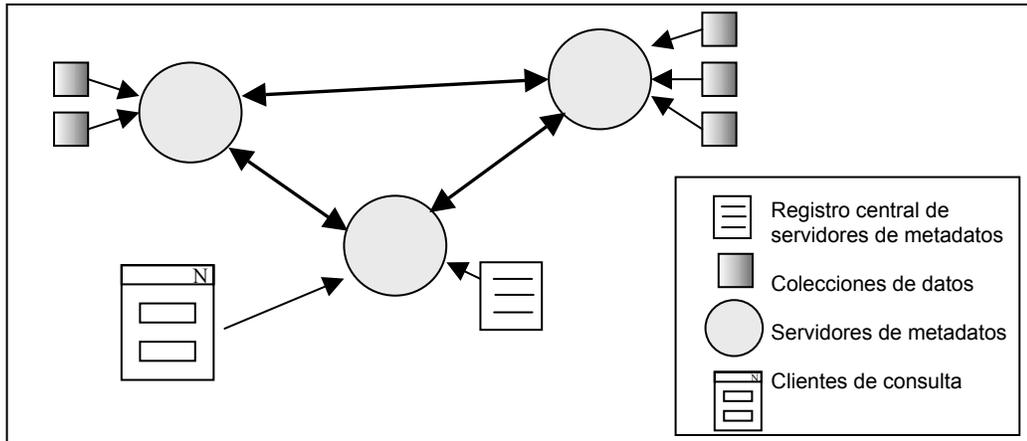


Fig.11. Localizando Fuentes de Metadatos: (red de servidores de metadatos)

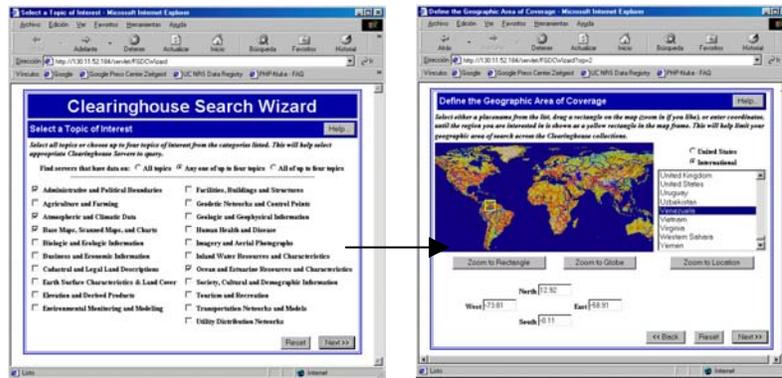


Fig.12. Interfaz global de búsqueda de datos geospaciales de GSDI.

El modelo de *clearinghouses* de datos, es seguido por diferentes comunidades como la de bibliotecólogos (revisar por ejemplo trabajo de la organización “Open Archives Initiative” (OAI)⁵³ relacionado con bibliotecas digitales y catalogación de publicaciones electrónicas), la comunidad de datos geo-espaciales, ya mencionada, y en particular ha influido en el trabajo relacionado con sistemas

⁵¹ <http://www.fgdc.gov/clearinghouse/clearinghouse.html>

⁵² <http://www.gsdi.org/>

⁵³ OAI: <http://www.openarchives.org/>

de almacenamiento y ubicación de colecciones de datos, llevado a cabo por la comunidad de científicos ecológicos y ambientales que trabajan en el proyecto LTER (mencionado en la sección 1.2.3.1).

2.3.2 Metadatos estructurados y estándares de intercambio de datos

Se discuten a continuación, como complemento de los conceptos definidos en 1.1.1, algunos aspectos del uso de metadatos estructurados y estándares de intercambio de información, que deben ser tomados en cuenta por los desarrolladores de sistemas de información ecológicos.

2.3.2.1 Codificación y formato de metadatos

Describir un recurso de información (como una colección de datos científicos) usando metadatos estructurados permite que estos sean comprendidos tanto por humanos como por “maquinas”, facilitando así la interoperabilidad entre sistemas; entendiendo este último termino como la habilidad de múltiples sistemas usando diferentes plataformas de hardware, estructuras de datos e interfaces, para intercambiar datos con una perdida mínima de contenidos y funcionalidad (NISO, 2001). Algunas de las funciones que se facilitan gracias al uso de metadatos estructurados son la ubicación (*discovery*), análisis y síntesis de datos. A continuación discutiremos algunas características del lenguaje XML como herramienta de representación de metadatos estructurados.

XML (del que se hablo en la sección 1.3.2.5) es un lenguaje de marcado basado en texto, diseñado para etiquetar el contenido o información en un documento con la finalidad de ser validado contra un esquema formal de contenido. Los esquemas son conjuntos de elementos de metadatos diseñados para describir un tipo particular de recursos de información. Los esquemas de metadatos por lo general especifican los nombres y significado (semántica) para los elementos (ver sección 1.3.2.3). Lo anterior hace que *XML* se adapte perfectamente a la representación de metadatos estructurados.

El uso creciente de *XML* en años recientes, tanto como lenguaje de representación de datos, y como lenguaje de intercambio a través de los protocolos de amplio uso en Internet hoy en día (*http*, *ftp*, etc.), ofrece diferentes ventajas, que lo convierten en fuerte candidato para ser considerado en cualquier proyecto de manejo e intercambio de datos e información de cualquier tipo.

El formato de presentación de los metadatos afecta seriamente la usabilidad de las colecciones de datos descritos por ellas (McCartney, 2000, pp2). La presentación de los metadatos en formatos impresos o texto *ASCII* por ejemplo, son adecuadas para la comprensión humana, pero no es adecuada para el procesamiento por parte de computadoras. Los documentos *XML* pueden ser fácilmente transformados en otros formatos a través de la especificación del lenguaje *eXtensible Stylesheet Language (XSL)*. En el siguiente gráfico () se destacan las posibilidades de separación entre datos y su formato de salida. Se muestra en a) el uso de *XML* junto con *XSL* para procesar múltiples archivos *XML* usando una sola transformación de formato, y en b) se esquematiza la generación de múltiples salidas o formatos a partir del mismo archivo de datos *XML*.

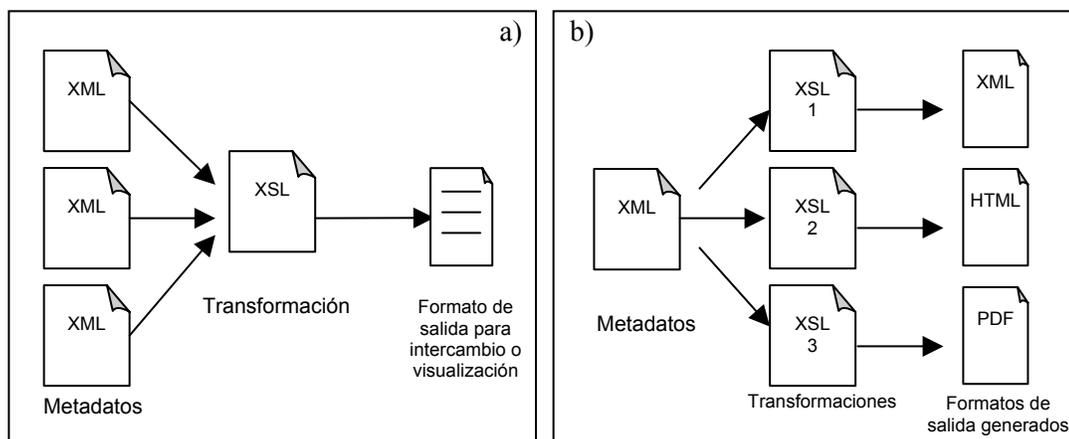


Fig.13. Esquema de transformaciones XML-XSL

La comunidad de científicos ambientales, y en especial los miembros de la red LTER, han venido trabajando en los últimos años para aprovechar las posibilidades que ofrecen los lenguajes derivados de XML y tecnologías relacionadas, y a partir del año 1999 recomiendan el uso de XML como lenguaje de representación e intercambio de metadatos estructurados entre sus miembros (McCartney, 2001).

En la sección 2.2.2 se describen algunas herramientas que han sido desarrolladas recientemente para ayudar en el manejo de datos y metadatos, y que pueden ser usadas por cualquier comunidad de científicos, como punto de partida para sus propios desarrollos.

2.3.2.2 EML como lenguaje de representación e intercambio de metadatos

En este trabajo se hace especial énfasis en el uso del lenguaje de representación de datos EML versión 2.0; desarrollado por miembros de la comunidad de científicos ecológicos, para representar las partes lógicas de las clases de descriptores genéricos para datos ecológicos definidos por Michener (1997) (explicadas en la sección 1.3.4.2). EML está codificado usando XML, y consiste en una serie de módulos que especifican el contenido de los metadatos (ver sección 1.1.1 para detalles de este lenguaje). Los creadores de este lenguaje hicieron una revisión de todos los modelos de contenido y estándares de otras comunidades como la de datos geo-espaciales (FGDC), la organización internacional de estándares (ISO), y las especificaciones del perfil de extensión del estándar de FGDC para colecciones de datos biológicos (perfil biológico de FGDC). Todas estas organizaciones y estándares fueron mencionados en la sección 1.3.3.

Un documento EML puede tener una vida larga como archivo y como formato, ya que es un archivo texto legible por humanos, que puede ser leído directamente por los científicos o administradores (gerentes) de datos, y también es de gran utilidad para aplicaciones de software ya que su formato estructurado es “procesable automáticamente” (*machine parseable*) (Jones, 2003). En el listado de la se observa la apariencia de un archivo EML. En el listado se incluyen solo el “título”, “creador” y “contacto” para la colección de datos. Este nivel de metadatos es útil para la ubicación y búsqueda simple de datos entre científicos ecológicos.

```

<?xml version="1.0"?>
<eml:eml
  packageId="eml.1.1" system="http://knb.ecoinformatics.org"
  xmlns:eml="eml://ecoinformatics.org/eml-2.0.0">

  <dataset>
    <title>
      Datos climaticos de la zona sur del lago entre 1999 y 2003
    </title>
    <creator id="23445" scope="document">
      <individualName>
        <givenName>
          Carlos
        </givenName>
        <surName>
          Gomes
        </surName>
      </individualName>
      <electronicMailAddress>
        carlosgomes@ula.ve
      </electronicMailAddress>
    </creator>
    <contact>
      <references>23445</references>
    </contact>
  </dataset>
</eml:eml>

```

Fig.14. Ejemplo de archivo EML (eml-dataset)

2.3.3 Tecnologías y herramientas para manejo de colecciones de datos

Se listan y describen brevemente algunas de las herramientas de propósito general para manejo de metadatos. Se muestran primero descripciones breves y direcciones en el Web de herramientas de utilidad para trabajar con el lenguaje *XML*. Casi todas son de uso gratuito, y obtenibles a través de Internet (solo una de ellas es una herramienta comercial). Luego se muestran herramientas desarrolladas específicamente para trabajar con archivos de metadatos según el estándar *EML*.

2.3.3.1 Herramientas relacionadas con XML⁵⁴

Editores XML:

Xeena.

Xeena es un editor XML desarrollado por IBM; fue escrito en Java, provee interfaz gráfica basada en iconos sensitivos al contexto que ayudan a insertar nuevos elementos y editar los existentes. También puede validar el documento contra un archivo *DTD*. La edición es realizada usando una vista estructurada (árbol), o directamente del código fuente *XML*. Se puede descargar y usar sin costo.

Web: <http://www.alphaworks.ibm.com/tech/xeena>

Amaya.

⁵⁴ Ver lista completa de herramientas XML en: <http://www.garshol.priv.no/download/xmltools/>

Amaya es un navegador y editor *XHTML* y *XML* desarrollado por W3C. Soporta *MathML* y *SVG*. También puede leer y salvar documentos *HTML* como *XHTML*. Su fuente es *Open Source* y de uso gratuito.

Web: <http://www.w3.org/Amaya/>

XML Notepad

XML Notepad es un programa desarrollado por Microsoft para editar *XML* de manera simple. Permite la edición rápida de datos y archivos basados en *XML*. Permite insertar datos dentro de prototipos *XML*. Su interfaz es muy simple y sencilla de usar. Representa gráficamente la estructura de árbol de los datos *XML*. Es de uso gratuito.

Web: <http://www.webattack.com/get/XMLnotepad.shtml>

XML Edit Pro

Este programa es un editor *XML* simple e intuitivo. Tiene características de edición de múltiples documentos, arrastrar y soltar entre nodos de documentos, así como copiar y pegar entre archivos *XML*. También integra el Navegador *Internet Explorer* y puede editar archivos vía *FTP*. Es gratuito.

Web: <http://www.daveswebsite.com/>

XMLSpy

XMLSpy es un completo editor *XML*. Puede trabajar con archivos *XML* de cualquier tipo, genera y edita archivos de *Schemas XML* (es la única herramienta mencionada aquí con esa capacidad) y *DTDs* a partir de muestras *XML* o desde cero; integra archivos *XML* con *XSL*, etc. es una herramienta comercial, y se puede bajar de Internet un demo completo por 30 días.

Web: <http://www.altova.com>

Herramientas para XSL

XMLSpy

Ya mencionado antes, este programa puede editar archivos *XSL*, y recibe y genera consultas *Xpath* a ser insertadas en archivos *XSL*.

Web: www.altova.com

XSpLit

Genera *DTDs* y archivos de formato *XSLT* desde un archivo *HTML* de muestra. Luego se puede combinar el archivo *XSL* con documentos *XML* que sigan la estructura definida en el *DTD*, y producir páginas *HTML* con el mismo diseño de la muestra original, pero con contenido sacados del archivo *XML*.

Web: <http://www.percussion.com/Download/xsplit/xsplit.htm>

XSL Tester

Visualizador de salidas *XSL*. Puede ejecutar funciones de edición *XML* y *XSL*. Se pueden ver las salidas como texto o procesadas como el explorador. Es distribuido gratuitamente.

Web: <http://www.vbXML.com/XSLtester/>

2.3.3.2 Herramientas relacionadas con EML

Se listan a continuación las herramientas que están siendo usadas y desarrolladas por el proyecto *LTER*, y *KNB*, que tienen relación con el manejo de colecciones de datos y *EML*.

Morpho

Morpho es una herramienta de administración de metadatos *EML*. Tiene funciones de edición, validación de estructura de archivos, y actúa como cliente para manejo de datos en la base de datos *Metacat*. Es de libre uso y distribución. Esta siendo usada actualmente en *LTER*.

Web: <http://knb.ecoinformatics.org>

Metacat

Metacat es un sistema de administración para archivos *XML*. Fue diseñado para ser usado como servidor de almacenamiento centralizado de colecciones de datos *EML*, que puede conectarse con otros servidores que usen los mismos mecanismos de comunicación e intercambio. Internamente puede hacer uso de cualquier herramienta de Base de Datos relacional. También es software abierto. Ya esta siendo usado por diferentes comunidades de científicos ecológicos, especialmente la Red *LTER* de EUA.

Web: <http://knb.ecoinformatics.org>

Xanthoria

Xanthoria es un sistema de consulta basado en *Xpath* para bases de datos distribuidas. Provee accesos mediante “*web services*” a fuentes de datos remotas, y usa *XSL* para traducir los contenidos a esquemas estándares. Esta actualmente en desarrollo y prueba.

Web: <http://ces.asu.edu/bdi/subjects/xanthoria>

Xylographa

Xylographa es una herramienta de creación de metadatos que permitirá generar archivos de datos *EML* con diferentes interfaces, y también generara metadatos a partir de los datos y metadatos ya existentes. Esta actualmente en desarrollo por parte de miembros de *KNB* y *LTER*.

Web: <http://ces.asu.edu/bdi/subjects/xylographa>

En la siguiente tabla se listan de forma muy resumida algunas de las herramientas mencionadas anteriormente, ordenadas según la tarea o acción en la que están involucradas:

Tarea o acción	Herramienta
Captura de metadatos, creación de metadatos estructurados y creación de colecciones de datos	Editores de texto ASCII, <i>XMLSpy</i> , <i>Morpho</i> , interfaces Web desarrolladas en este trabajo de tesis, y <i>Xylographa</i> en el futuro
Almacenamiento y manejo de colecciones de datos	Sistema de archivos, <i>Morpho</i> , <i>Metacat</i> , <i>Xindice</i>
Consulta de colecciones de datos	<i>Morpho</i> , <i>Metacat</i> , en el futuro herramientas como <i>Xanthoria</i>
Intercambio y descubrimiento de colecciones de datos	Sistema e Interfaces desarrolladas en este trabajo, <i>Metacat</i> .
Análisis, Validación y control de calidad	<i>XMLSpy</i> , <i>Morpho</i> , SAS (aplicación estadística)

Tabla 2. Herramientas para manejo de colecciones de datos *EML*

2.4 Sistemas prototipo para manejo de datos y metadatos

2.4.1 Descripción de prototipos desarrollados

Para validar los conocimientos adquiridos, para probar las diferentes técnicas relacionadas con tecnologías involucradas en los sistemas de información estudiados, y para adquirir experiencia en el manejo de colecciones de datos, a lo largo de este trabajo se realizaron varios prototipos de manejo de colecciones de datos. Estos prototipos están siendo validados contra los requerimientos de la Eco-Red Venezuela y los de la Red de Estaciones Bioclimáticas del Edo. Mérida. Los dos primeros prototipos han estado siendo usados y probados principalmente por los miembros de la Red de Estaciones del Edo. Mérida, y el último será usado por ambas organizaciones. En un futuro cercano la experiencia ganada servirá para implementar métodos y sistemas de información definitivos para estos dos proyectos y cualquier proyecto de manejo de datos ecológicos y ambientales.

Hasta ahora se han realizado tres prototipos enmarcados dentro de este proyecto de tesis. Describiremos brevemente a continuación las características de los dos primeros prototipos. A partir de la sección 0 se describirá formalmente la metodología usada y se discutirá con más detalles el tercer prototipo desarrollado, que incorpora la mayor cantidad de funcionalidad en cuanto al trabajo con metadatos y al manejo de colecciones de datos.

Descripción prototipos 1 y 2:

Prototipo etapa 1:

- Sitio Web prototipo con acceso a servicios de consulta de datos de estaciones
- Prototipos de formularios Web de envío de datos
- Acceso a datos sin procesar de estaciones meteorológicas
- Acceso a datos a través de mapas sensitivos
- Ejemplos de productos de información: animaciones, climadiagramas, datos diarios sin procesar, consulta de últimos datos capturados, etc.
- URL: <http://www.cecalc.ula.ve/webclima/PrototipoPaginasClima/>

Prototipo etapa 2:

- Acceso a archivos de metadatos en formato texto en servidor Web
- Organización, almacenamiento y publicación de datos y metadatos en sistema de archivos en servidor Web
- Envío regular de datos y metadatos a servidor Web desde cuatro estaciones meteorológicas en Mérida. Datos están siendo usados por miembros de la Red de Estaciones de Mérida.
- URL: <http://www.cecalc.ula.ve/webclima/datos/>



Fig.1519. Muestra de interfaz de prototipos 1 y 2

A continuación se muestran los detalles del tercer prototipo desarrollado, incluyendo la descripción de la metodología y herramientas utilizadas para su elaboración, se enumeran los requerimientos y se define la arquitectura del sistema. Se describen también sus diferentes componentes y se proporcionan algunas muestras de la interfaz de dicho prototipo. A partir de ahora, las referencias a “el prototipo” se referirán al tercer prototipo mencionado en 2.4.5.

2.4.2 Metodología y herramientas utilizadas.

2.4.2.1 Metodologías

Para el desarrollo de este trabajo, se usó la metodología de prototipos. Según [Somerville \(2000\)](#), esta metodología ayuda a los futuros usuarios y desarrolladores a entender los requerimientos para el sistema, y proporciona a los usuarios finales un sistema en funcionamiento. El uso de prototipos también ayuda, entre otras cosas, a que los usuarios tengan una idea concreta acerca de las posibles capacidades del sistema en desarrollo, y puedan intervenir en etapas iniciales de dicho desarrollo ([Somerville, 2000](#)) y, por ejemplo, pueden tener una idea temprana de las características de la interfaz de usuario del sistema.

En ([Baker, 2000](#)) se destaca que, partiendo de un marco de desarrollo modular de sistemas de información ecológicos, los prototipos de los módulos del sistema, pueden ser implementados y probados por cualquier sitio miembro de una red, con posibilidades de ser migrados o usados por cualquier otro miembro o sitio (o toda la red de sitios). Un prototipo potencia avances en el desarrollo del sistema de información, poniendo a disposición un producto para ser revisado y probado en un corto plazo, y enfoca la discusión en aspectos funcionales concretos. Los prototipos pueden ser desarrollados por uno o varios grupos, y debería ser evaluado por pequeños grupos antes de ser discutido en un rango mayor de sitios. Los participantes en el desarrollo y prueba deberían estar de acuerdo en que posiblemente algunas decisiones de diseño serán cambiadas y cualquier desacuerdo en cuanto a las características y funcionalidades de un sistema en prueba pueden ser modificadas para adaptarse a las sugerencias que vayan surgiendo entre los participantes. En la siguiente ilustración se muestra gráficamente el modelo de prototipos.

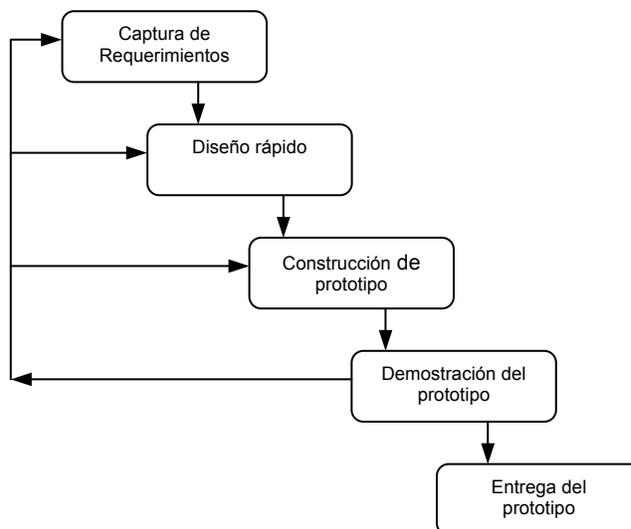


Fig. 16. Esquema del modelo de prototipos

2.4.2.2 Lenguajes de programación, modelado y representación de datos

A continuación se proporcionan algunas características de los lenguajes de programación, representación y transformación de datos usados en los prototipos desarrollados en este trabajo.

Lenguaje PHP⁵⁵

Las técnicas relacionadas con el desarrollo de prototipos se relacionan regularmente con el uso de lenguajes de alto nivel y lenguajes interpretados. Según Farrell (2002), las estructuras y sentencias de alto nivel y la ejecución inmediata de estos lenguajes interpretados, los hace ideales para análisis personalizado de datos, donde la misma persona que tiene que realizar el análisis o transformación de los datos puede escribir el programa o aplicación que realice la tarea. *PHP* (versión 4.x) es un lenguaje interpretado de alto nivel independiente de la plataforma de sistema operativo (se puede instalar en la mayoría de las plataformas de Hardware/S.O.), y permite un desarrollo rápido de programas y aplicaciones para el Web, lo que influyó en su escogencia para el desarrollo de los componentes principales del último sistema prototipo desarrollado en este trabajo de tesis.

Lenguaje de codificación de metadatos: XML

Se usó *XML* como lenguaje de codificación de los metadatos que conforman las colecciones de datos. En el capítulo 1 se discutieron las diversas características y ventajas del uso de *XML* como lenguaje de representación e intercambio de recursos de información.

Lenguaje de modelado de metadatos: EML

Como estándar de modelado de las diferentes clases de metadatos se usó *EML 2.0.0*, según está definido en <http://knb.ecoinformatics.org/software/eml>. (ver sección 1.1.1)

Lenguajes y herramienta de formato de metadatos: XSL y Parser XSL

Para realizar las conversiones de formato y salidas se utilizó el lenguaje *XSL* (ver 1.3.2.5), ya que permite mantener la separación entre los datos y el formato de salida y visualización finales entre otros beneficios. Usando *XSL* también es sencillo agregar formatos de salida adicionales a los metadatos, o crear filtros que codifiquen los metadatos con las especificaciones de otros estándares diferentes a *EML* (en inglés, para esto se usa el término *metadata crosswalks*).

Para interpretar los archivos *XSL* se usó el *parser MSXSL 4.0*⁵⁶ que implementa la versión 1.0 del estándar *XSL* definido por *W3C*⁵⁷.

2.4.3 Definición de requerimientos para sistema de manejo de colecciones de datos.

En la sección 2.2.1 se definieron los roles, actores y acciones involucrados dentro de un sistema de manejo de metadatos. Basándonos en dichas definiciones, se especifican a continuación los requerimientos de un sistema de manejo de datos y metadatos para las ciencias ecológicas. Para mostrar los requerimientos y funcionalidades básicas del sistema, usaremos diagramas de casos de uso según la especificación *UML* (OMG, 2003). Los diagramas de casos de uso sirven para representar el carácter funcional de los objetos. Se elaboran a partir de las especificaciones iniciales de los usuarios, permitiendo modelar las expectativas que estos tienen con respecto al sistema a desarrollar (López, 1998).

⁵⁵ <http://www.php.net>

⁵⁶ <http://msdn.microsoft.com/library/default.asp?url=/downloads/list/xmlgeneral.asp>

⁵⁷ <http://www.w3.org/TR/xslt>

Primero se enumeran textualmente las características deseadas de funcionamiento del sistema, para luego representar gráficamente dichas operaciones usando diagramas *UML* de casos de uso.

2.4.3.1 Características y descripción general para el sistema

Se debe diseñar un sistema que permita documentar los datos que están siendo capturados o generados por los científicos u organizaciones pertenecientes a una red de colaboración científica. Se debe hacer uso de los estándares y tecnologías de información y comunicación adecuados para facilitar las tareas de recopilación de datos y documentación de los mismos a nivel centralizado. Así mismo se deben proporcionar facilidades de consulta e intercambio a través de Internet (interfaces Web principalmente) de las colecciones de datos o los metadatos. A continuación una representación textual de los procesos principales requeridos.

2.4.3.2 Proceso general de creación y consulta de colecciones de datos

Se lista la secuencia de tareas y acciones pertenecientes a las dos funciones principales del sistema a ser desarrollado: creación y consulta de colecciones de datos.

Creación de colecciones de datos:

- Proveedor de datos captura o genera datos.
- Proveedor de datos accede a sitio Web de creación de colecciones de datos .
- Proveedor de datos introduce metadatos (descripción de datos) en formularios de captura Web.
- Sistema genera colección de datos en sistema central de almacenamiento y consulta.
- Proveedor de servicios efectúa labores de mantenimiento del sistema.

Consulta de colecciones de datos:

- Usuario final (humano o aplicación) accede a servidor (sitio Web) de consulta de colecciones.
- Usuario final (humano o aplicación) selecciona criterios de consulta y envía solicitud a través de interfaz Web o accede a recurso de información directamente usando el *URL* del recurso.
- Sistema central valida y recibe consultas y solicitudes de colecciones de datos.
- Usuario final obtiene metadatos de colecciones de datos en diferentes formatos (*HTML*, *TXT*, *XML*).

2.4.3.3 Diagramas de casos de uso

A continuación algunos diagramas de casos de uso (OMG, 2003; Martin, 1998) que representan gráficamente los procesos anteriormente descritos.

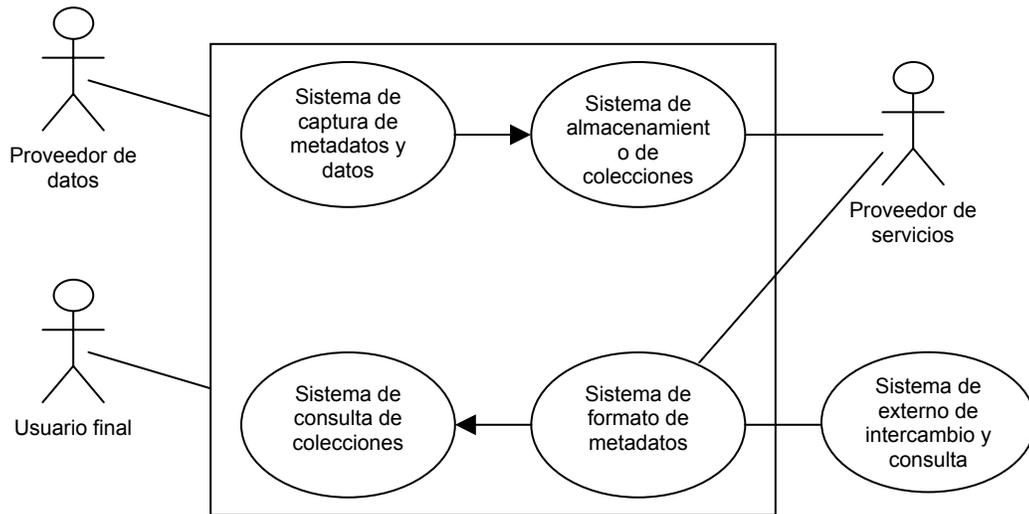


Fig.17. Diagramas de casos de uso para delimitar el sistema completo

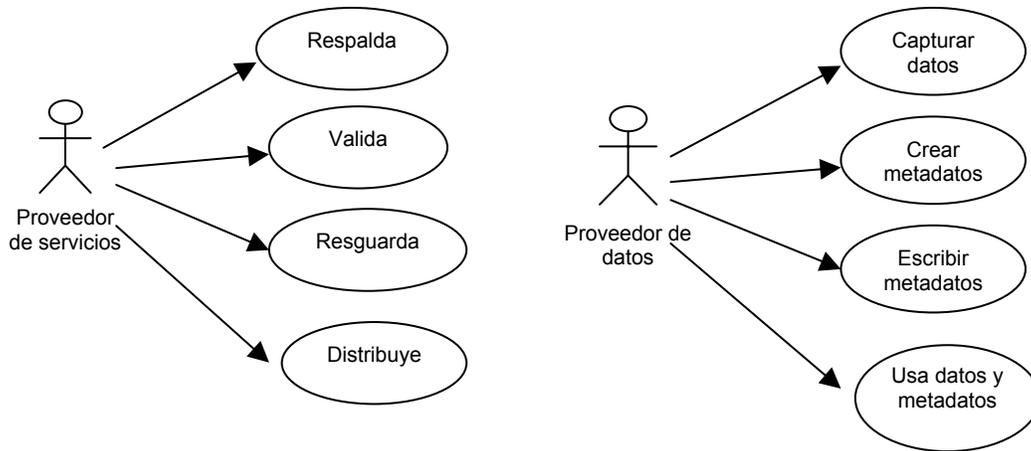


Fig.18. Casos de uso para Proveedor de servicios y Proveedor de datos

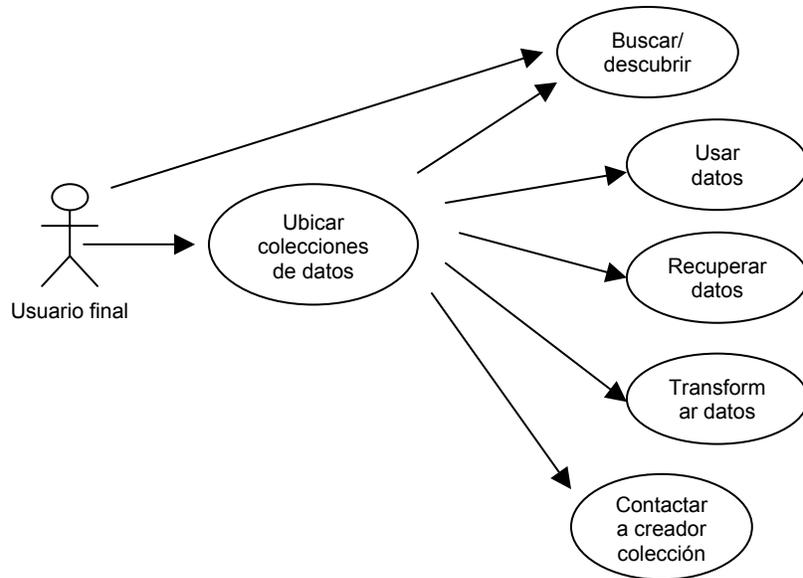


Fig.19. Casos de uso para usuario final

2.4.4 Arquitectura del prototipo de manejo de colecciones de datos

El prototipo se basa en una arquitectura cliente-servidor, compuesta de diferentes partes y subsistemas. Se uso como arquitectura base el modelo definido en 2.3.1. En la siguiente figura (Fig. 20) se observa el diagrama de estructura del sistema propuesto.

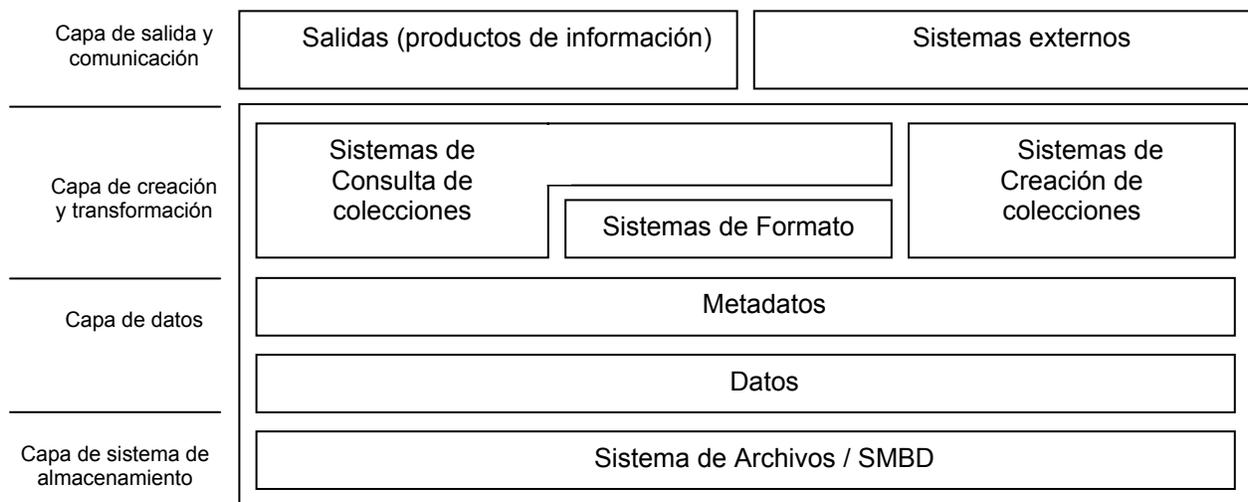


Fig.20 Diagrama de estructura del sistema

De la estructura presentada en la figura anterior se destacan las siguientes características:

- Separación lógica entre datos y metadatos.
- Separación entre capas de almacenamiento, datos y generación de salidas (formato).
- Generación de diferentes productos de información (salidas).
- Posibilidad de comunicación con sistemas externos.

2.4.5 Descripción del sistema prototipo desarrollado.

2.4.5.1 Partes del sistema prototipo

De acuerdo con los requerimientos definidos en 2.4.3, el prototipo de sistema de manejo de colecciones de datos está formado por cinco subsistemas y partes principales, a saber:

1. Cliente de captura de metadatos de colecciones.
2. Sistema de creación de colecciones.
3. Sistema de almacenamiento de colecciones y datos.
4. Sistema de consulta de colecciones.
5. Sistema de formatos de salida o productos de información.

Es requerida y considerada la posibilidad de comunicación con otros sistemas (como *Metacat*) vía protocolo *http*. También se contempló la posibilidad de que los archivos de datos estén almacenados tanto en el propio servidor como en cualquier ubicación en Internet (fuera del servidor de metadatos).

En el siguiente gráfico (Fig. 21) se pueden observar de forma esquemática todos los componentes de este sistema.

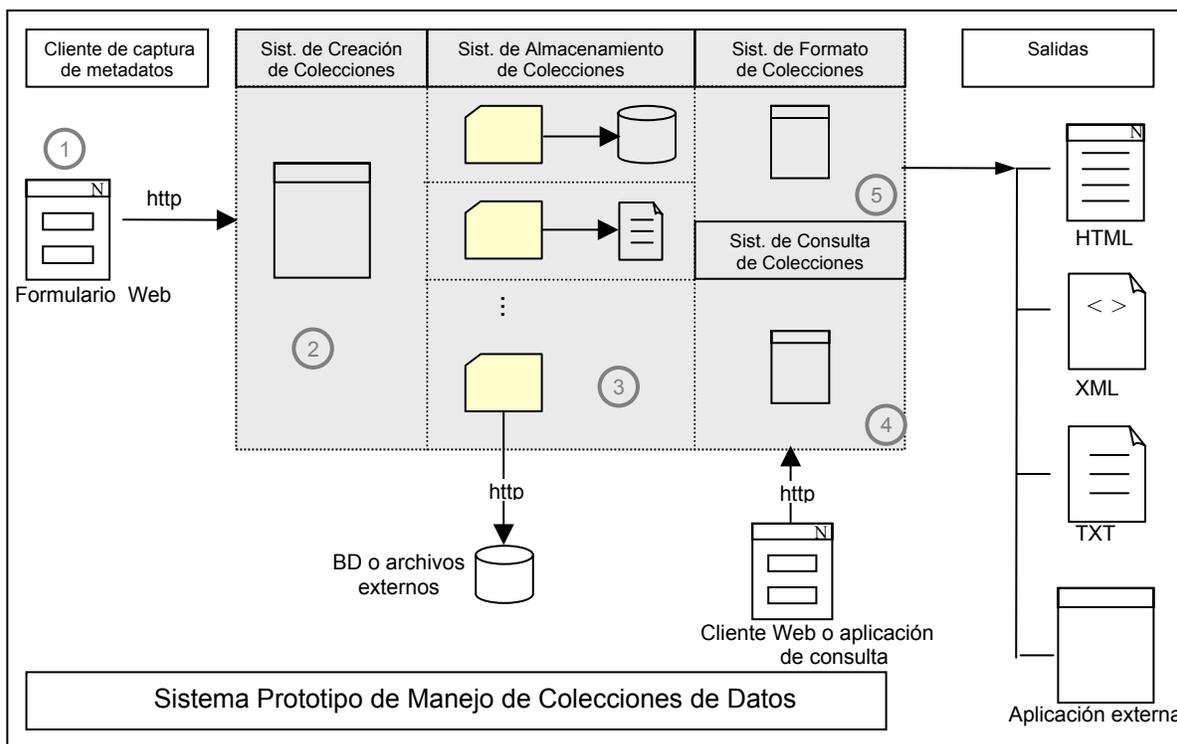


Fig.21. Componentes del sistema prototipo de manejo de colecciones de datos

A continuación la descripción detallada de cada una de las partes.

1. Cliente de captura de metadatos de colecciones

Es un cliente Web que es activado a través de Internet usando una dirección *URL* de acceso (Por ejemplo: `http://www.xxx.xx/sistema_metadatos/crea_colección.html`). Consiste en un formulario *HTML* que sirve de entrada a los datos básicos de una colección de datos. Los datos recolectados con esta planilla pertenecen a la clase *eml-dataset* correspondiente a la versión 2.0b6 de este estándar. Esta planilla está realizada usando código *HTML* y código *JavaScript* para validación y llenado automático de algunos campos. El formulario está conectado con un programa *CGI*⁵⁸ programado usando el lenguaje *PHP*, que realiza las operaciones del siguiente componente (sistema de creación de colecciones). Esta planilla es similar en funcionalidad a la usada por algunos miembros del proyecto LTER para la creación de colecciones de datos que serán parte de esta red (CNRS, 2003; OBFS) En la sección 2.4.6 (Fig. 26) se muestra la apariencia de la interfaz de captura de metadatos.

2. Sistema de creación de colecciones

Recibe la entrada del formulario cliente *HTML* mediante variables de programa y estructuras de datos correspondientes a cada parte del formulario. Este componente está programado usando el lenguaje *PHP*. Su función es la de crear archivos *EML-XML* que pertenecen a la clase *eml-dataset*, y almacenarlos de manera organizada dentro del sistema de archivos (*filesystem*) del servidor de metadatos donde está instalado el sistema. Por cada colección se genera un archivo *XML* de la clase *eml-dataset*, que contiene los metadatos de identificación de la colección. Se genera automáticamente un identificador de la colección usando números consecutivos, el código de tres letras asignado a la estación y la fecha de creación de la colección (p.ej. `par_20_180503.xml`)

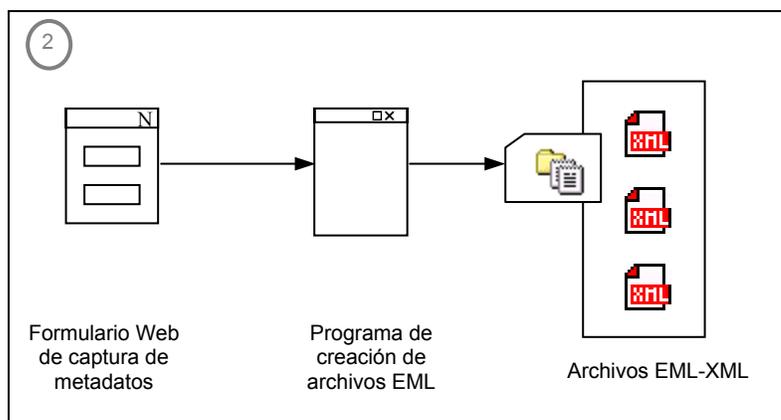


Fig.22. Sistema de creación de colecciones

3. Sistema de almacenamiento de colecciones

⁵⁸ CGI: Common Gateway Interface

Consiste en una serie de archivos *XML* (colecciones de datos) creados por el sistema anterior (componente #2) y organizados dentro del sistema de archivos del servidor. Cada colección de datos posee una “carpeta” o directorio en el sistema de archivos, dentro de la cual están los archivos de metadatos. Las colecciones de datos están compuestas de archivos *EML-XML*, que contienen los metadatos de la colección, y opcionalmente pueden contener los datos (dentro de archivos texto *ASCII* por ejemplo) o accesos a los archivos y bases de datos descritos por dichos metadatos. En la se observa el esquema de este componente.

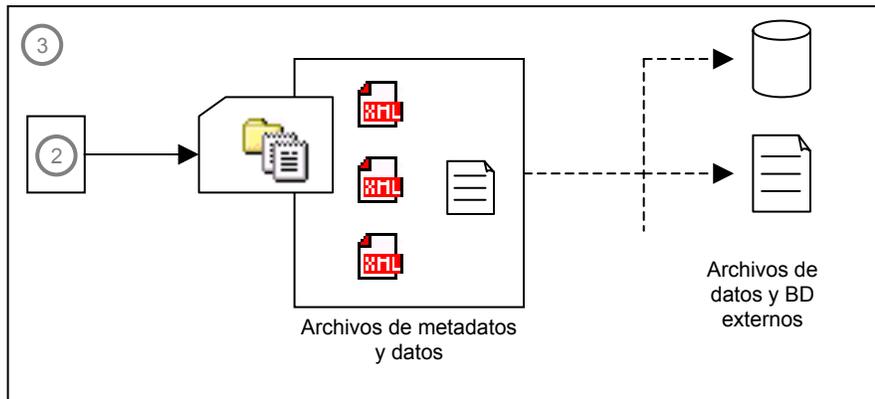


Fig.23. Sistema de almacenamiento de colecciones

4. Sistema de consulta de colecciones

En el prototipo desarrollado, la consulta consiste en el acceso al sistema de archivos por parte de programas *CGI-PHP* que cargan los contenidos de dichos sistemas de archivos y generan páginas *HTML* dinámicas con los listados de colecciones disponibles. A partir del listado se puede acceder a los archivos de metadatos individuales de cada colección. El sistema de formato se encarga de generar las diferentes salidas. También se puede acceder a los archivos de metadatos directamente usando su *URL* correspondiente dentro del servidor.

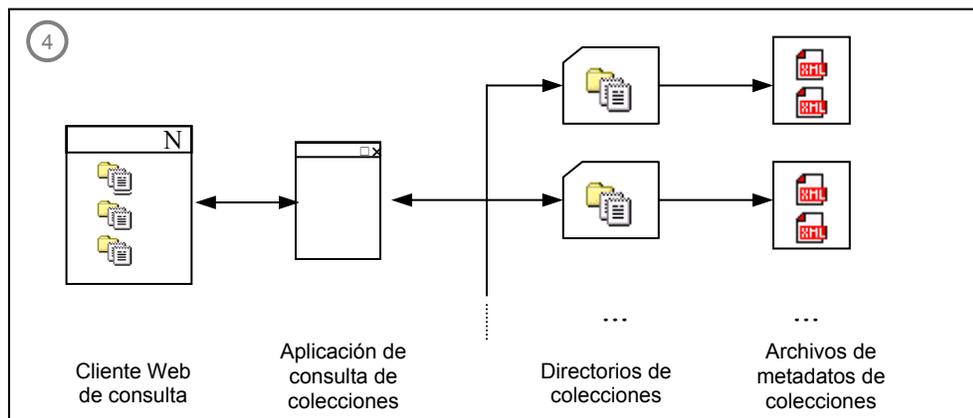


Fig.24. Sistema de consulta de colecciones

5. Sistema de formatos de colecciones

Su función es la de generar diferentes formatos de salida o productos de información. La salida de los metadatos se realiza a través de uno o más archivos *XSL* que dan formato a los metadatos en el archivo *EML-XML*.

Las salidas son generadas por transformaciones *XSL* dinámicas que se procesan al momento de activarse la consulta por parte de un cliente Web o aplicación externa. En este prototipo se incluyeron transformaciones *XSL* que generan HTML y salidas de texto ASCII a partir de los archivos *XML* de metadatos de cada colección. También se da la alternativa de consultar directamente los archivos *XML*. Esto último puede ser útil para conectar el sistema con programas o aplicaciones externas o programas de recolección (*harvesting*) de metadatos.

En la siguiente figura (Fig. 25) se observa un esquema de esta parte del sistema. Se observa en la figura que el sistema está compuesto principalmente de diferentes transformaciones *XSL* (*XSL 1*, *XSL 2*, etc.) que transforman los archivos *XML* que contienen los metadatos en diferentes productos de información o formatos de salida. Se pueden añadir fácilmente tipos o formatos de salida, simplemente añadiendo otras transformaciones de salida *XSL*, debido a que el formato se realiza de forma independiente a los otros módulos de procesamiento. Esto también contribuye a mantener la independencia entre datos y formato de salida.

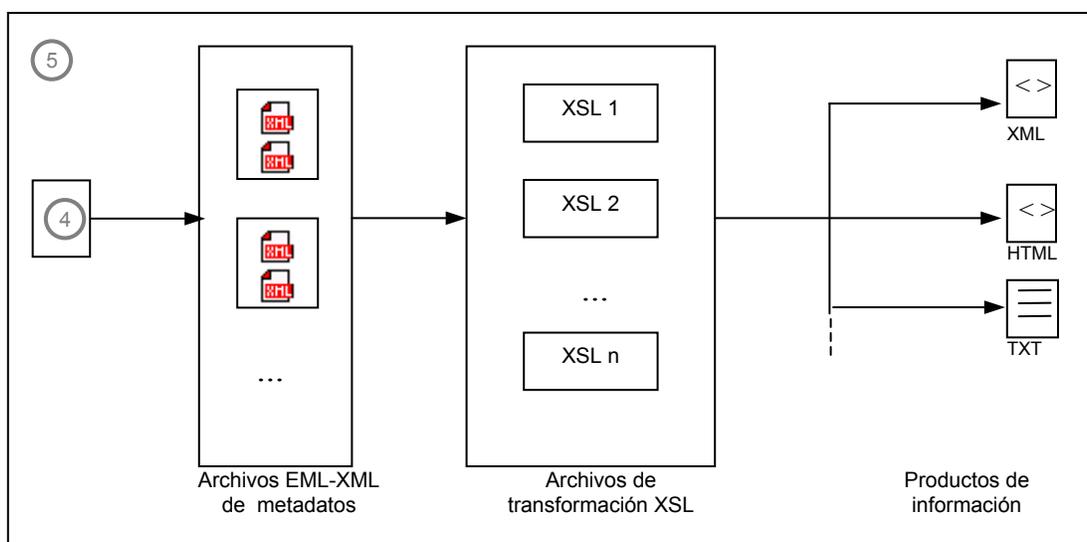


Fig. 25. Sistema de formato de colecciones

2.4.6 Características e interfaz del prototipo

2.4.6.1 Interfaz de sistema de creación de colecciones de datos:

Comenzaremos describiendo la interfaz del cliente de captura de metadatos que corresponde a la parte número 1 del sistema según la descripción proporcionada en 0. Para ingresar los datos de la colección de datos se diseñó un formulario *HTML*, usando diferentes tipos de elementos de entrada (cajas de texto, listas de selección, etc). El usuario al llenar estos campos está creando la documentación básica de una colección de datos.

Se incluyeron algunas funcionalidades y validaciones programadas en *JavaScript* que facilitan la tarea de llenado del formulario. Por ejemplo algunos campos se pueden llenar automáticamente con valores ya llenados, en el caso de que se tengan que colocar los mismos valores en otra parte del formulario. También están validados los campos obligatorios para que no se permita el envío del formulario sin haberlos completado.

En este prototipo se creó una interfaz para las estaciones pertenecientes a la Eco-Red Venezuela, por lo que se puede seleccionar una de las diez estaciones pertenecientes a la Eco-Red, como parte del proceso de documentación. En la siguiente figura (Fig. 26) se muestra parte del formulario de captura de metadatos.

Formulario de Creación de Nuevas Colecciones de Datos

Use este formulario para crear la descripción de una NUEVA colección de datos.
Campos con (*) son obligatorios.
Si tiene problemas con el llenado de este formulario, comuníquese con soporte@ecored.org.ve

1. INFORMACION BASICA

* Nombre y Apellido de la persona que está llenando este formulario Nombre Apellido

* Nombre del sitio de ECO-RED Venezuela Seleccione su sitio aquí Código

Título de la Colección de Datos

2. CREADOR DE LA COLECCION DE DATOS (DATA SET ORIGINATOR)

Función (rol) del creador de la colección (Originator's rol)

* Nombre

* Apellido

Nombre de la Organización (Organization Name)

Fig.26. Interfaz Web de captura de metadatos

2.4.6.2 Interfaz de consulta de colecciones

En este prototipo las consultas al repositorio donde están almacenadas las colecciones de datos pueden hacerse de dos maneras:

1. Navegando por los enlaces hipertexto de la interfaz web de las colecciones.
2. Usando directamente el URL del archivo de metadatos de la colección de datos correspondiente a la estación buscada, junto con la transformación de salida escogida.

En la siguiente figura (Fig. 27) se muestran dos pantallas de la interfaz web de las colecciones. Una muestra el listado de las estaciones que poseen colecciones de datos en el sistema y la otra muestra la lista de colecciones existentes para una estación cualquiera.

A esta parte del sistema se accedería con una dirección URL como la siguiente (esta dirección es hipotética):

http://www.eco-red.org.ve/consulta_colecciones/

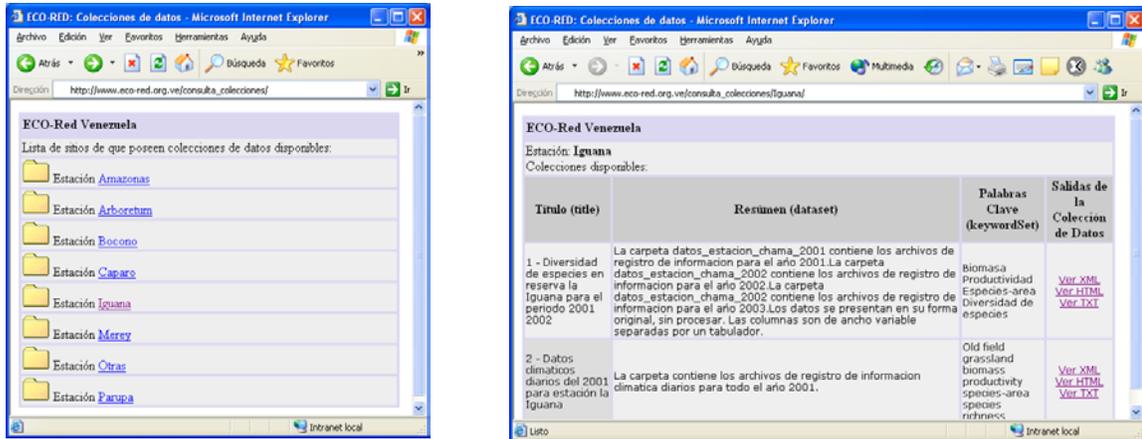


Fig.27. Interfaz Web de consulta de colecciones de datos

La segunda manera de consultar una colección es usando directamente el URL del archivo de metadatos de la colección. Por ejemplo con el siguiente URL colocado en la barra de direcciones de un navegador Web:

http://www.ecored.org.ve/datasets_xml/Parupa/par_17_180503.xml

obtendríamos el archivo XML con los metadatos de la colección correspondiente. Si se usa un navegador con capacidad de mostrar archivos XML (*Internet Explorer* por ejemplo) veríamos algo parecido a lo siguiente (Fig. 28):

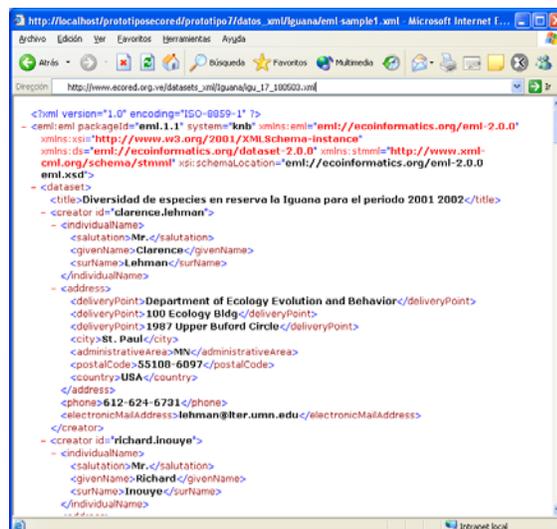


Fig.28. Salida de consulta de archivo XML de colección de datos

Es de hacer notar que esta consulta puede ser realizada internamente por otra aplicación o sistema. Por ejemplo un sistema central de recolección de datos puede “llevarse” o recolectar estos datos e incluirlos en un catálogo centralizado. También una aplicación de búsqueda inteligente o minería de datos puede acceder directamente a los datos almacenados en archivos *XML*, entre muchas otras posibilidades. Es de destacar también la posibilidad de transferir estos datos al sistema *Metacat*⁵⁹, haciendo uso también del mecanismo de consulta antes explicado, combinado con las posibilidades de inserción de datos de dicho sistema.

Existe también un mecanismo de consulta que permite generar diferentes formatos de salida (aparte de la salida *XML*). Por omisión se incluye dentro del sistema prototipo una transformación *XSL* que genera una salida *HTML*, de los metadatos de la colección a la que se accedería por ejemplo, con un *URL* como el siguiente:

```
http://www.ecored.org.ve/consulta_coleccion/parser.php?URL=/Iguana/emt-sample2.xml&XSL=salidahtml_1.xsl
```

Esta consulta generaría una salida *HTML* en el navegador Web como la siguiente (Fig. 29):

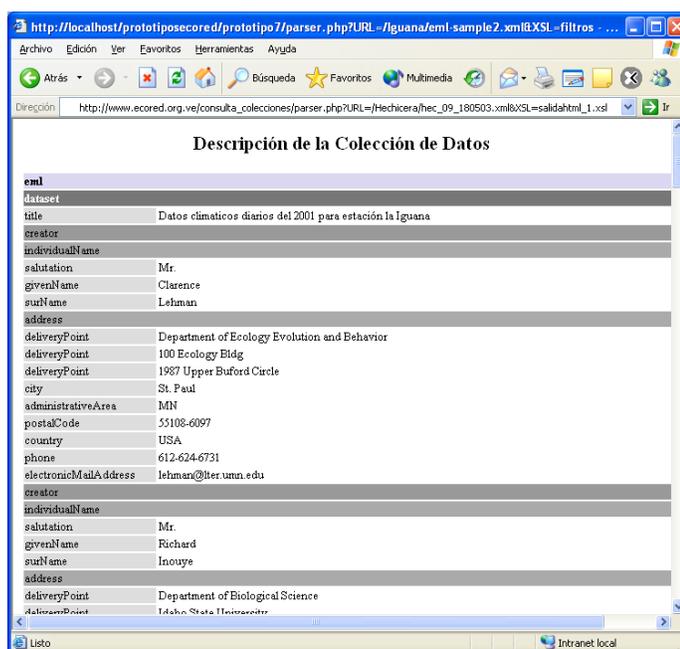


Fig.29. Salida *HTML* de consulta de colección de datos

También se incluyó una transformación *XSL* que genera una salida en formato texto *ASCII*. En la siguiente figura (Fig.30) se visualiza dicha salida en un navegador Web, activada desde la pantalla de consulta de colecciones de una estación en particular.

⁵⁹ <http://knb.ecoinformatics.org/software/metacat/>

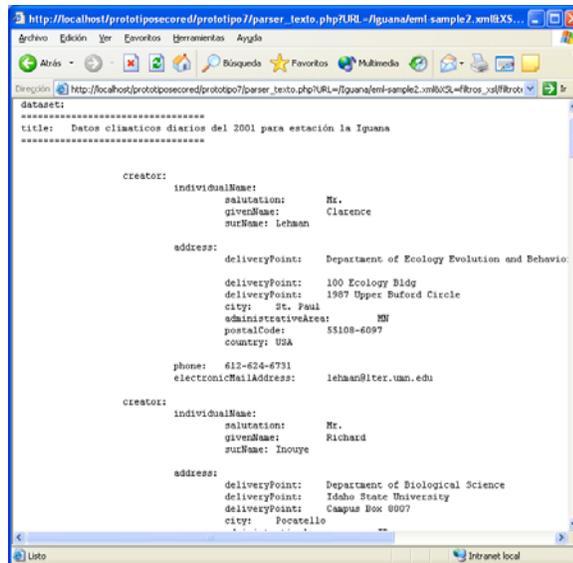


Fig.30. Salida en formato texto a partir de interfaz de consulta de colecciones

Con solo crear transformaciones *XSL* de formato adicionales e incorporarlas al sistema, se pueden generar diferentes productos de información a partir de los metadatos almacenados.

Las variables que permite este programa de consulta son:

- **URL:** Dirección *url* en el servidor donde se encuentra el archivo de metadatos *EML*.
- **XSL:** nombre del archivo de transformación que se desea aplicar.

Las transformaciones *XSL* de formato que generan las diferentes salidas están almacenadas separadas del resto de los archivos del programa prototipo (en un directorio particular), pudiéndose añadir fácilmente archivos de transformaciones *XSL* que pueden ser activados dándole como valor a la variable “*XSL*” el nombre del archivo *XSL* deseado. Vemos a continuación ejemplos de valores que pueden tomar las variables “*URL*” y “*XSL*”:

```
URL = /Hechicera/hec_09_180503.xml
XSL = salidahtml_1.xsl
```

Se pueden crear filtros para realizar transformaciones de los metadatos a otros estándares de representación diferentes a *EML* (*metadata crosswalks*), e invocarlos de la manera explicada.

En el próximo capítulo se discutirán diferentes aspectos de la aplicación de los conocimientos, técnicas y metodologías aprendidas y desarrolladas durante esta investigación a dos redes de trabajo que manejan diferentes tipos de datos ambientales, y ecológicos, la Eco-Red Venezuela y la Red de Estaciones Bio-climáticas del Edo. Mérida.

3 *Aplicación a Casos de Estudio: Eco-Red Venezuela y Red de Estaciones del Edo. Mérida*

En las secciones siguientes se discutirán aspectos relacionados con las posibles alternativas de aplicación de las investigaciones y labores realizadas para este trabajo de tesis, en dos casos de estudio locales: la Eco-Red Venezuela, y la Red de Estaciones del Edo. Mérida. En las secciones 1.2.3.3 y 1.2.3.4 se describieron estos dos proyectos. Este capítulo muestra en parte los resultados de la investigación realizada.

A continuación se discuten las diferencias o similitudes entre la red LTER de EUA con los casos de estudio y realidades locales, se enumeran las actividades realizadas hasta ahora dentro de estos proyectos seguido de una propuesta de plan de manejo de colecciones de datos para la Eco-Red y la Red de Estaciones de Mérida.

3.1 **Similitudes y diferencias con LTER-EUA y los casos de estudio. Aspectos del manejo de datos en Venezuela.**

Durante este trabajo se han estudiado los métodos y herramientas que se usan en la red LTER de los Estados Unidos (e Internacional⁶⁰), y se ha considerado, el uso de los mismos estándares de representación e intercambio de metadatos como parte de la solución a los aspectos relacionados con el manejo de colecciones de datos. Sin embargo antes de proponer métodos de trabajo específico para ser aplicados en las organizaciones y proyectos que se usaran como caso de estudio (Eco-Red Venezuela y Red de Estaciones del Edo. Mérida), se ha considerado útil hacer una comparación de las condiciones y características de un proyecto como LTER de los EUA, con las condiciones y características de proyectos locales y nacionales, similares en objetivos y características básicas.

3.1.1 **Comparación de características de LTER con casos de estudio de este trabajo**

Se presenta a continuación una tabla comparativa con las características de LTER EUA y las características y condiciones de proyectos similares en Venezuela (Tabla 3) (los casos de estudio para este proyecto). Los datos para esta tabla provienen en el caso de LTER, de la documentación y sitios Web usados en este trabajo de tesis, y en los casos de estudio locales, proviene de la recopilación de información y comunicaciones personales de los participantes en estos.

LTER EUA	Eco-Red Venezuela y Red de estaciones de Mérida
Número de sitios miembro de LTER: 25 sitios principales y cerca de 130 estaciones asociadas.	Miembros ECO-RED Vzla.: 10, pero potencialmente existen muchos candidatos a pertenecer a esta red. Sitios en Red de Estaciones de Mérida actuales: 4 estaciones de captura de datos y 4 otras dependencias que aportan datos esporádicamente.
Cada sitio LTER esta compuesto de varias estaciones de investigación y captura de datos.	Miembro = estación. En general, los miembros de ECO-RED y Red de Estaciones poseen una o pocas estaciones de investigación asociadas.
Muchos miembros de LTER poseen sistemas de información (infraestructura informática, SMBD, personal especializado en informática.)	Pocos miembros usan sistemas de información o manejadores de bases de datos. Solo hacen uso de herramientas informáticas básicas (procesadores de palabras, hojas de cálculo, etc)

⁶⁰ <http://www.ilternet.edu>

Mayoría de sitios LTER poseen conexión permanente a Internet (como servidores de información)	Mayoría carece de conexión permanente en el sitio, pero personal tiene acceso (por su cuenta) a e-mail, Web, etc.
Tienen necesidad grande de coordinar actividades inter-sitios.	Igual pero a menor escala. Aunque cada sitio no está totalmente establecido o estable como productor de datos científicos
Consideran prioritario las labores de síntesis e integración de la información de toda la red	Prioridad en manejar las colecciones de datos a nivel básico: preservación y documentación de las colecciones de datos
Reservas en cuanto a la publicación de datos recolectados antes de generar publicación de resultados de la persona que los recolecto. En marcha programa de educación en toda la red	Igual reservas en cuanto a publicación. Necesidad de concientizar y educar con respecto a compartir documentación sobre colecciones de datos (y datos)
Datos heterogéneos entre sitios. Dificultad de compartir y encontrar datos	Igual. No hay mucha información sobre la manera como recolectan los datos en cada sitio
Existe oficina coordinadora administrativa central	Igual.
Existen grupos de desarrollo e investigación temáticos para toda la red (financiados por entes gubernamentales nacionales de ciencia)	No existen grupos ni organización de investigación y desarrollo, pero se tiene acceso a investigadores de instituciones cercanas. Recientemente se constituyó la Eco-Red como Laboratorio Nacional, lo que le da acceso a financiamiento por parte del gobierno
Manejo de información guiado por políticas a largo plazo (CIMF: <i>Common Information Management Framework</i>)	Manejo de información guiado por necesidades inmediatas.
Grupos de trabajo temático se reúnen frecuentemente	Reuniones generales de miembros y coordinadores de organizaciones y proyectos
Están intentando estandarizar formatos de metadatos a ser usados en toda la red	No existen metadatos ni documentación de las colecciones de datos de los miembros

Tabla 3. Comparación de características de LTER con casos de estudio

A lo largo de este trabajo de tesis se fue acumulando información sobre la manera como los científicos y organizaciones trabajan con datos científicos en proyectos locales y algunos a nivel nacional en Venezuela. También se tiene información sobre los participantes en la red LTER Internacional (ILTER, 2000). En (Torréns, 2001) se hace un diagnóstico de la situación en algunas dependencias y proyectos en Venezuela, relacionados con el manejo de datos científicos y ambientales. Se resume a continuación nuestra opinión sobre la situación en cuanto a manejo de datos en proyectos locales (y dependencias gubernamentales o de investigación nacionales), reforzando también el cuadro de diferencias entre proyectos locales y LTER de EUA mostrado anteriormente.

3.1.2 Algunos aspectos de manejo de datos en proyectos locales:

A continuación algunos aspectos que caracterizan a los proyectos que tienen que ver con manejo de información, en el ámbito de venezolano.

- No existe comunicación sistemática entre dependencias gubernamentales o universitarias involucradas en el manejo de datos. Los proyectos donde están involucrados datos científicos, son realizados de forma aislada e individual incluso dentro de una misma institución como la Universidad.
- Inexistencia casi total de documentación de los datos (metadatos) en proyectos o dependencias que están capturando o han capturado datos en el pasado.
- Casi nunca se usan estándares de ningún tipo para capturar, almacenar o intercambiar datos.
- Las dependencias u organismos que están dispuestas a compartir datos usan cada una formatos diferentes para almacenamiento de los datos (datos en papel, hojas de cálculo y archivos provenientes de procesadores de palabras son los formatos de archivo e intercambio más comunes). Pocas usan bases de datos relacionales (ninguna organización estudiada posee interfaz Web pública a datos o metadatos)

- Datos poseen errores e inconsistencias debidas a la manera de transcripción o captura, o a fallas en los instrumentos de capturas. Ausencia de controles regulares o permanentes de calidad de los datos. Control de calidad manual en algunos pocos casos.
- Existe mucha “desconfianza” y “celo” que frena el intercambio y compartición de datos entre científicos. Algunos argumentan poseer derechos por X cantidad de años antes de compartir los datos. Ignorancia de la diferencia entre compartir y distribuir *datos* y compartir y distribuir *metadatos*. Poca cantidad de investigación inter-sitios.
- Datos poseen restricciones de acceso o poseen un alto costo de adquisición cuando se trata de escalas espacio-temporales extensas.
- Datos usados en un proyecto o investigación se desechan luego de finalizada la investigación que los originó (solo se mantiene la síntesis en formato de publicación, artículos, libros, etc.).
- Ausencia de conciencia en cuanto al uso que otro científico pueda darle a los datos, o ausencia de estudios a largo plazo que ameriten inversión y costos que implica documentar y mantener los datos por largos periodos de tiempo.
- Falta de personal capacitado en varias áreas relacionadas. Ausencia de grupos de personas especializadas en diferentes tareas de manejo de información en cada organización o proyecto.

3.2 Actividades realizadas en proyectos Eco-Red Venezuela y Red de Estaciones de Mérida.

Se mencionan aquí de forma resumida varias de las labores llevadas a cabo durante la realización de este trabajo de tesis. Estas labores estuvieron enmarcadas al inicio, en el proyecto de la *Red de Estaciones del Edo. Mérida*; y luego las labores y aplicación del trabajo se amplió a la *Eco-Red Venezuela*.

Lista resumen de tareas realizadas durante este trabajo de tesis:

- Estudio de tecnologías de captura de datos (estaciones meteorológicas para proyecto de la Red de Estaciones del Edo. Mérida).
- Adquisición y puesta en funcionamiento de dos estaciones meteorológicas automáticas (nuevas).
- Recuperación de dos estaciones meteorológicas automáticas.
- Comienzo de captura de datos de prueba provenientes de cuatro estaciones automáticas. Pruebas de envío de datos usando tecnologías inalámbricas.
- Creación de sitio Web con documentación del proyecto de red de estaciones climáticas de Mérida: <http://www.cecalc.ula.ve/webclima/> (Prototipo 1).
- Creación de sitio Web prototipo para proyecto de red de estaciones climáticas de Mérida: <http://www.cecalc.ula.ve/webclima/PrototipoPaginasClima/> (Prototipo 2).
- Organización de los datos capturados por las estaciones.
- Búsqueda y recopilación de datos de otras fuentes (MARNR, ULA, etc.).
- Creación de archivos de metadatos de todas las estaciones.
- Mantenimiento y actualización permanente de archivos de datos y metadatos.
- Prueba de herramientas de edición de metadatos *Morpho*.
- Presentaciones sobre manejo de datos ante encargados Eco-Red y Red de Estaciones de Mérida.
- Desarrollo de prototipos de manejo de colecciones de datos (se realizaron tres prototipos principales).
- Organización de talleres de manejo de datos.

3.3 Plan general para manejo de colecciones de datos para Eco-Red Venezuela y Red de Estaciones del Edo. Mérida

Se proponen seguidamente recomendaciones a dos niveles: primero a nivel de las oficinas coordinadoras de los proyectos que están siendo objeto de estudio y segundo se plantea un proceso de manejo de colecciones de datos que puede ser aplicado dentro de estos proyectos.

3.3.1 Tareas organizacionales

Se plantean a continuación una serie de recomendaciones y tareas que deben ser consideradas a nivel de la oficina coordinadora de las organizaciones y redes de trabajo e investigación, en este caso la Oficina coordinadora de la Eco-Red Venezuela y la coordinación del proyecto de la Red de Estaciones del Edo. Mérida.

Tareas a nivel de coordinación de la red:

- Organizar grupo de trabajo central de manejo de colecciones de datos (gerentes de información centrales). Se recomienda una o dos personas a nivel central, encargadas directamente de las labores de planificación, y diferentes labores de mantenimiento de colecciones de datos.
- Realizar encuestas (o actualizar la ya aplicadas) en sitios participantes, para recabar información sobre recursos humanos disponibles, equipos informáticos, colecciones de datos, etc.⁶¹ Dame un diseño de encuesta..
- Identificar y organizar recursos humanos en cada sitio donde se maneje información (gerentes de información locales). Debe existir por lo menos una persona que asuma el rol de gerente local de información en cada sitio de la red.
- Discutir y definir por parte de las oficinas de coordinación central de los proyectos, la estrategia general de manejo de colecciones de datos para la red, estándares y apolíticas, junto con gerentes locales de información (usando como base este trabajo, el trabajo de la red LTER y experiencias particulares de cada miembro). Entre otras cosas se debe comenzar con incentivar procesos de documentación de colecciones de datos dentro de actividades de cada estación o miembro.
- Buscar sitios voluntarios para participar en pruebas de sistemas de información, procedimientos, etc., y establecer mecanismo de validación y críticas de dichos sistemas.
- Definir y dar a conocer la política oficial de manejo de datos en la red o proyecto. En algunos sitios se le da mucha importancia a esta tarea y es una de las primeras que se realizan. Aquí se define la accesibilidad y políticas de distribución de los datos entre otras cosas⁶².
- Documentar todos los procesos de manejo de colecciones de datos y colocar dicha documentación a disposición de los miembros de la red⁶³.
- Definir y aplicar estrategias y métodos de preservación y respaldo de la información (a largo plazo).
- Organizar talleres de formación sobre manejo de colecciones de datos científicos (ver anexo 1) para gerentes locales de información.
- Se deberían coordinar actividades de entrenamiento y asesoría a nivel de gerentes centrales de información con participación de expertos de otras redes LTER.
- Evaluar funcionamiento de sistemas y métodos de trabajo a nivel de toda la red.
- Corregir o adaptar procesos, métodos y herramientas de acuerdo a necesidades particulares.

⁶¹ Ver diferentes encuestas y experiencias llevadas a cabo por la estación LTER-Luquillo de Puerto Rico en: <http://luq.lternet.edu/datamng/Reports/DMatLUQdocuments.htm>

⁶² Ver ejemplo de política de manejo de información de sitio LTER en: <http://luq.lternet.edu/datamng/imdocs/dmpolicy.htm>

⁶³ Ver por ejemplo guías para manejo de información en: <http://luq.lternet.edu/datamng/imdocs/division.html>

- Gerentes de información centrales deben informar a miembros de cualquier cambio en los métodos o herramientas de trabajo. En general debe existir mecanismo de comunicación permanente entre gerentes locales de información y oficina central o gerentes de información centrales.

Comentarios adicionales:

Actualmente tanto la Eco-Red Venezuela y la red de estaciones del Edo. Mérida poseen los recursos humanos mínimos para efectuar labores de gerencia de la información a nivel central. La Eco-Red Venezuela posee profesionales dedicados al manejo de la información de dicha Red, y la coordinación del proyecto de la Red de Estaciones del Edo. Mérida posee personal técnico y de coordinación mínimo, que en este momento efectúa labores de documentación de las colecciones de datos que provienen de las estaciones de captura de datos, realizando algunas de las tareas que tiene que ver con la administración de las colecciones de datos, generación de documentación, preparación de manuales de entrenamiento, control de calidad de los datos, etc.

3.3.2 Creación de colecciones de datos: procesos recomendados y escenarios posibles.

Se presenta a continuación una propuesta general de procesos para la creación de colecciones de datos en las organizaciones que están participando como casos de estudio, desde el punto de vista de las tareas de manejo de información. Se discuten también los posibles escenarios en cuanto a manejo de datos que se pueden presentar.

3.3.2.1 Proceso general propuesto para creación de colecciones de datos

Se muestra el proceso en forma de tabla de tres columnas (Tabla 4) donde la primera columna indica al actor involucrado en la tarea (*quien* la realiza), la segunda especifica que tarea va a efectuar el actor correspondiente (*que* tarea) y la tercera indica la herramienta o técnica que se puede usar para efectuar dicha tarea (*como* se debe realizar la tarea). La definición de los actores puede revisarse en la sección 2.2.1 de este trabajo.

Para definir a los actores se usaran las siguientes abreviaturas:

GIL: Gerente de información local GIC: Gerente de información central CUF: Científico o usuario final de las colecciones de datos

Actor (Quien)	Tarea (Que)	Herramienta (Como)
GIL, GIC	GIL recibe entrenamiento teórico y técnico de oficina coordinadora de red (de gerente central de información-GIC)	Taller de manejo de colecciones de datos
GIL	GIL evalúa las necesidades de documentación de sus colecciones de datos, y selecciona las más adecuadas. Comprueba que existen metadatos mínimos exigidos.	Documentación de estándar de metadatos
GIL	GIL revisa requisitos o condiciones de publicación de datos y metadatos de su institución (derechos de autor y publicación, normas de referenciar, citar y usar las colecciones de datos, etc.)	Normas de sitio o institución que origina los datos
GIL	GIL crea metadatos para colecciones de datos del sitio usando	Morpho,, Interfaces Web

	herramientas locales o interfaces Web y sistemas de creación de colecciones.	o patrones (Excel, Texto, XML, etc)
GIL	GIL realiza primer control de calidad a los datos o metadatos.	Morpho, editor XML
GIL	GIL envía metadatos (y datos opcionalmente) a servidor central de metadatos.	Morpho, interfaz Web, FTP.
GIC	Gerente de Información Central (GIC) revisa, agrega y valida metadatos enviados. Asigna identificadores a colecciones.	Morpho, aplicaciones de validación EML
GIC	GIC publica colecciones de datos en servidor de la red	Sistema de archivos de servidor, Servidor Web, Morpho, Metacat
GIL	GIL actualiza (localmente y en servidor central) metadatos en el caso de colecciones de datos que aun están siendo recopiladas (cada vez que surjan nuevos datos o cambie algún dato o información dentro de la colección)	Morpho, Interfaces Web
GIL	GIL contactara a investigadores que capturaron los datos en caso de correcciones, datos faltantes o errores en las colecciones de datos que están siendo documentadas.	N/A
GIC	GIC mantendrá, preservara, respaldara, etc. colecciones de datos en servidor central	Documentos y manuales de respaldo. Sistemas informáticos en servidor central.

Tabla 4. Proceso de creación de colecciones de datos

En el siguiente gráfico (Fig.31) se resumen los pasos anteriores, y se representa esquemáticamente el proceso básico de obtención de datos y creación de colecciones, destacando que algunas de las tareas no siguen un orden secuencial (adaptado de Luquillo, 2001).

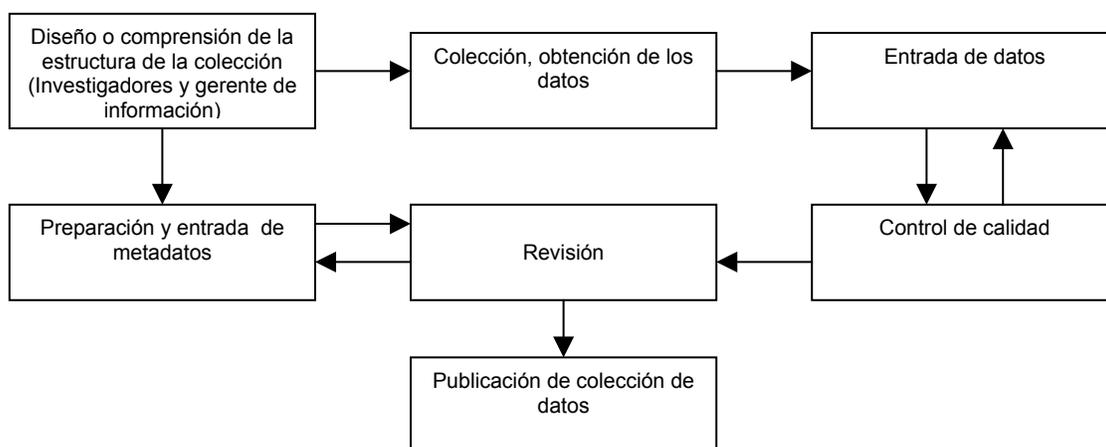


Fig.31. Esquema de entrada de datos y creación y publicación de colecciones

3.3.2.2 Escenarios posibles para el manejo de colecciones de datos

La oficina coordinadora de red hará evaluaciones y censos entre los miembros de la red, para averiguar condiciones generales actuales de colecciones de datos y uso de sistemas de información o aplicaciones locales de manejo de los mismos; así como posibles requerimientos que pueden surgir de las estaciones y redes participantes. Si existen herramientas en uso entre los miembros se evaluará estrategias de conversión, conexión y traducción a estándares y sistemas propuestos.

Se listan algunos escenarios que se pueden presentar en cada sitio o estación miembro de la Eco-Red y la Red de Estaciones del Edo. Mérida, y se comenta en cada caso el método y herramientas necesarias para tratar con dicho escenario. Los tres primeros escenarios suponen que gerentes de información locales tienen algún tipo de acceso a Internet en sus sitios o cerca de ellos. Estos escenarios fueron adaptados de (Databits, 2002)

Escenario 1: Sitio con datos y metadatos en archivos texto o en hojas de cálculo y desea compartirlos con demás miembros de red.

Asumiendo que el sitio quiere seguir manteniendo su sistema de formato y archivos como lo ha venido haciendo, pero quiere a su vez compartir sus datos y metadatos con otros miembros de la red o con cualquier científico interesado, se pueden tomar varias vías, entre ellas:

- a) Que los gerentes de información usen *Morpho*, o un editor *XML* para documentar las colecciones que estén interesadas en compartir, y enviarlas al servidor *Metacat* de la Red o al servidor de *KNB*. Desventaja: impráctico para grandes volúmenes de datos y metadatos.
- b) Usar interfaces Web de captura de metadatos como la desarrollada en este trabajo y enviar colecciones de datos al servidor central de la red. Desventajas: las mismas anteriores.

La opción I) tiene la ventaja de que al crear colecciones de datos estructuradas (*EML*) con *Morpho*, se solucionan problemas de falta de metadatos adecuados para las colecciones locales, ya que ayuda a documentar los datos a todos los niveles (sitio, colección, entidades de datos, atributos de los datos, etc.) y se está creando un repositorio local de información ordenado y con capacidades de consulta, que puede ser compartido casi instantáneamente con otros científicos.

Escenario 2: Sitio desea crear sistema de información con funcionalidades propias.

Existen diferentes funcionalidades administrativas particulares que un sitio puede querer incluir en un sistema de información que no tengan nada que ver con el manejo de metadatos estructurados en *EML*. Por ejemplo se puede querer un sistema de manejo y seguimiento de tareas, proyectos o informes internos, para lo que herramientas como *EML*, *Morpho*, interfaces Web o *Metacat* no están diseñadas (son herramientas especializadas), lo que obliga a crear sistemas separados de manejo administrativo y de manejo de datos.

Aun así están apareciendo muchos productos y opciones de diseño de sistemas de información que pueden incluir soporte para estructuras de datos *XML* (bases de datos *XML* nativas o funcionalidades *XML* en BD relacionales por ejemplo), lo que en un futuro ayudaría a interconectar sistemas actualmente disímiles.

Escenario 3: Sitio desea que sus científicos ingresen sus propios metadatos al sistema

Este caso se puede presentar cuando un sitio no tenga personal encargado del manejo de los datos (gerente local de información), o también cuando el científico quiere mantener actualizados los metadatos de sus colecciones a medida que un proyecto de investigación en desarrollo avance o no

quiere esperar por la disponibilidad de otro personal para que documente los datos. En este caso lo más adecuado es utilizar la herramienta *Morpho*, para ingresar y editar en cualquier momento las colecciones de datos por parte del propio científico, teniendo la posibilidad de manejar los datos localmente (edición consulta), y también puede compartirlos enviándolos al servidor central de la red.

Escenario 4: Sitio no desea administrar ningún sistema, solo documentar y compartir sus metadatos y datos.

Para este caso es suficiente el uso de una interfaz Web sencilla (como la desarrollada en este trabajo) para crear colecciones. Este caso abarca los sitios que tienen difícil acceso a herramientas informáticas y carecen de conexión permanente a Internet. Adicionalmente pueden usar *Morpho* y *Metacat* para editar y consultar colecciones de datos propias y de otros sitios.

Escenario 5: Existe voluntad de documentar y compartir datos pero solo se poseen mínimas facilidades de informática y ninguna posibilidad de conectividad a Internet en el sitio o cerca de él.

Aquí los datos se recopilarán de la manera tradicional (formatos en papel, hojas de cálculo, archivos texto, formularios patrón, etc.) y con alguna regularidad, formato y método acordados se harán llegar a los gerentes de información centrales, que se encargaran de complementar la información adecuadamente.

Es fundamental que se identifiquen otros escenarios y requerimientos que pueden presentarse. Una de las formas de hacerlo es a través de encuestas (formales o informales) que ayuden a conocer las situaciones particulares de cada sitio miembro de una red. Por otro lado en otras ocasiones, el mecanismo de las encuestas no ha provocado mucha participación, por lo que este debe apoyarse en otro mecanismo de recopilación de información o en un sistema de recompensas, como se mencionó en la sección 2.2.1.4. Otra de las formas de hacerlo es durante los talleres iniciales de manejo de datos y metadatos, una vez que los participantes entiendan los conceptos básicos involucrados

3.3.3 Entrenamiento: Talleres de Manejo de información para científicos y gerentes de información.

Como parte del entrenamiento necesario para que los científicos y gerentes de información, a nivel central y a nivel local, adquieran los conocimientos necesarios para el manejo y gerencia de colecciones de datos, se propone la realización de talleres relacionados con el tema del manejo de datos y metadatos. También se debe pedir asesoría directa de los miembros del equipo de manejo de información de la Red LTER de EUA (*Information Managers Committee*), relativa al uso de los últimos avances en tecnologías de información que se están desarrollando para uso de los miembros de LTER.

Como ejemplo se muestra a continuación información descriptiva del taller propuesto (Ver Tabla 5) y en el anexo 1 se muestra la tabla de contenido de un taller básico de manejo de información que será dictado a los miembros de la Eco-Red Venezuela y de la Red de Estaciones del Edo. Mérida o a los miembros de cualquier proyecto o red que involucre a un grupo de científicos que tengan el interés de preservar y compartir los datos que resultan de sus investigaciones y esfuerzos.

Ficha del Taller:

Nombre: Taller básico de manejo de datos y metadatos

Audiencia: Gerentes de información centrales, Gerentes de información locales, Científicos

Duración: 16 horas. 8 horas teóricas y 8 horas de actividades prácticas.

Instructores o dependencia encargada: Instructores de la Oficina Central de Manejo de Datos de proyecto Eco-Red Venezuela y Red de Estaciones del Edo. Mérida.

Objetivos:

Los participantes deberán aprender los conceptos básicos involucrados en el manejo de información científica, ecológica, ambiental, etc., y deberán conocer la importancia de documentar y preservar los datos. También deberán aprender a usar las herramientas que tienen a su disposición para realizar diferentes tareas relacionadas, así como a ubicar los recursos de información disponibles en Internet que sean de su interés, potenciando de esta manera el trabajo colaborativo e interdisciplinario.

Recursos técnicos necesarios:

Laboratorio de PCs con conexión a Internet (uno por participante), software Morpho, instalado en todas las máquinas, navegadores Web, editor XML (XMLSpy, XMLEDitPro o XMLNotepad), sistema de proyección (Videobeam)

Información y datos que deben estar disponibles al inicio del taller:

Muestras de datos y metadatos de los sitios participantes (en cualquier formato que sea legible en el laboratorio del taller)

Tabla 5. Información sobre el taller básico de manejo de datos y metadatos

Por último, también se recomienda la preparación de talleres más avanzados y especializados, donde se abarque en detalle el uso o de una herramienta de software o hardware en particular, o técnicas que tengan que ver con el uso de sistemas de información, control y aseguramiento de la calidad de los datos y metadatos, uso especializado de datos (en modelos de simulación por ejemplo), entre otras áreas.

4 Conclusiones y trabajo futuro

El uso de metadatos estructurados estandarizados representa un significativo avance dentro de las ciencias ecológicas y ambientales y la ecoinformática. Proveen la manera de catalogar la creciente base de archivos de datos y resuelven el problema de representación de metadatos usando estándares sintácticos comunes. También potencian el desarrollo de aplicaciones cada vez más versátiles, que permitirán a los usuarios compartir, encontrar y hacer uso de archivos interconectados de datos ecológicos.

En esta investigación se propone el uso de metadatos estructurados estandarizados para facilitar la colaboración entre científicos ecológicos, biólogos, o científicos de cualquier otra disciplina. Se usaron como casos de estudio iniciales dos proyectos donde están involucrados redes de científicos y manejo de datos, el proyecto de la *Red de Estaciones bio-climáticas del Edo. Mérida* y la *Eco-Red Venezuela*.

Este trabajo se concentró en estudiar los estándares y tecnologías relacionadas con la descripción de colecciones de datos para las ciencias ecológicas, y tenía el objetivo de proponer métodos y herramientas que sirvan para el desarrollo de sistemas de información dentro del dominio antes mencionado. Se tomaron en cuenta las actividades y experiencias de proyectos equivalentes en otras partes del mundo, y en especial el trabajo del grupo de manejo de información de la Red *LTER (Long Term Ecological Research Network)* de los EUA.

Como base estructural de diseño para la documentación de datos, se consideraron los descriptores para datos ecológicos (no-geoespaciales) propuestos por Michener (1997) y el lenguaje para metadatos *EML (Ecological Metadata Language)* desarrollado a partir de estos descriptores. Se comenzó implementando prototipos de sistemas Web de consulta de datos científicos que fueron utilizados y probados exitosamente; y por último se implementó un prototipo de un sistema de creación y consulta de colecciones de datos, que hace uso de los estándares y lenguajes estudiados (genera, procesa y consulta archivos en el formato estándar *EML*). Para el desarrollo de prototipos se hizo énfasis en el uso de lenguajes y técnicas de programación que están siendo usadas actualmente en muchas áreas y dominios de aplicación (*XML, XSL*, etc.). Estos prototipos fueron realizados para que se adapten a las necesidades de los casos de estudio, pero pueden servir de punto de partida para cualquier otra iniciativa similar.

La interoperabilidad de los contenidos de metadatos exportados al estándar *EML* debería constituir un valioso aporte para el descubrimiento y comparación de recursos de información como soporte a las ciencias ecológicas y ambientales. Los documentos *XML* que sigan un estándar para descriptores de tipos de documentos, como los de *EML*, pronto serán usados ampliamente como medio de intercambio de datos y metadatos para la investigación ecológica. Estas definiciones comunes de formato de datos permitirán un intercambio rápido y estandarizado de información y la interoperabilidad entre sitios en redes de investigación ecológica y sitios pertenecientes a otras áreas del conocimiento.

Es oportuno mencionar aquí que al usarse la infraestructura y estándares ya desarrollados y establecidos, los desarrolladores de sistemas de información toman ventaja de la experiencia de comunidades ya existentes, evitando re-inventar estándares y herramientas para cada nuevo dominio de desarrollo.

Los creadores de los datos deben ser animados a usar herramientas de creación de metadatos estructurados, para lo cual se hace necesario un adecuado entrenamiento y toma de conciencia entre estos. En esta investigación se propone también la realización de talleres de manejo de datos para

gerentes de información y científicos, que ayudaría en parte a que se haga uso efectivo de las herramientas y recomendaciones, y estas se conviertan en práctica diaria dentro de sus labores de investigación (para los científicos) o soporte a la investigación (para los gerentes de información). Para lograr esto hacen falta el establecimiento de políticas claras de manejo de la información dentro de las instituciones involucradas, acompañadas de códigos de buen uso y colaboración entre pares. Se planteó también el desarrollo de un sistema de incentivos para los proveedores de datos y gerentes de información.

Como resultado de la investigación surgieron diferentes propuestas y recomendaciones destinadas a desarrolladores de sistemas de información, especialmente a nivel de modelado de datos y arquitecturas de sistemas de información para manejo de colecciones de datos, lo que se espera constituya un valioso aporte para investigaciones y desarrollos futuros. Desde el punto de vista de las organizaciones y redes de colaboración científica, se comparó la situación de las redes y proyectos de manejo de datos de la Red LTER con las realidades locales, y se dieron recomendaciones que tienen que ver con el uso de herramientas, estándares y procesos para la documentación de datos. En este sentido diremos que actualmente en Venezuela existe un atraso de unos 15 a 20 años con respecto a proyectos como LTER de EUA. Sin embargo, gracias a tecnologías de fácil acceso y poco costo y a la adopción de convenciones científicas, es posible disminuir rápidamente esa brecha.

Entre los aportes de esta investigación, se puede mencionar el hecho de que los resultados y aprendizaje obtenidos se están probando y aplicando inmediatamente en la solución del problema de manejo de datos para los dos proyectos que originaron esta investigación. Aunque se está en las etapas iniciales de desarrollo de estos proyectos, se han logrado implementar métodos funcionales iniciales de manejo de datos, especialmente para la Red de Estaciones del Edo. Mérida, de cuyas estaciones miembro se reciben, catalogan y publican en Internet datos en forma continua⁶⁴. Se espera en un futuro cercano, probar con rigurosidad los métodos, herramientas y sistemas propuestos en este trabajo, en el ámbito de otros proyectos como la Eco-Red Venezuela.

Mientras EML provee una sintaxis consistente para representar colecciones de datos, no hace mucho para tratar con las diferencias semánticas entre colecciones de datos. Nuevos objetivos de investigación se están planteando actualmente para ofrecer soluciones basadas en ontologías que enlacen descripciones EML con conceptos que estén más cerca de los parámetros por los cuales se define el conocimiento ecológico (McCartney, 2001).

Este trabajo se puede complementar con otros estudios que abarquen en detalle aspectos tratados aquí superficialmente, como por ejemplo, uso de herramientas automáticas de control de calidad de los datos que hagan uso de metadatos estructurados; desarrollo de herramientas de minería de datos que se valgan también de los metadatos y de las posibilidades de interconexión entre colecciones y repositorios de datos, etc. Los objetivos que se deberían plantear para el futuro (y que ya se están planteando en otros lugares), tienen que ver con levantar una infraestructura que permita pasar de administración, gerencia o manejo de datos a la gerencia del conocimiento (Brunt, 2001).

Terminaremos diciendo que el uso adecuado de tecnologías, herramientas y convenciones científicas esta comenzando a facilitar la creación y uso compartido de colecciones de datos en un sinnúmero de campos del conocimiento, y la disponibilidad de los datos y metadatos ya no es (o pronto no lo será) un factor limitante para contestar preguntas fundamentales para la ciencia. Gracias a la ayuda de computadoras y redes de información los científicos tienen acceso cantidades masivas de datos, y podemos compartir y discutir dichos datos con colegas de todo el mundo. El reto actual es el de diseñar

⁶⁴ Ver: <http://www.cecalc.ula.ve/webclima/>

sistemas que sirvan para comprender datos tan complejos o interrelacionados que no podemos (ni tiene sentido) entender de forma aislada.

5 Bibliografía y referencias

- Ablan, M., Andressen, R. et al. (2000). *SIBiLa: Sistema de Información Bioclimática del Lago de Maracaibo*. Propuesta de proyecto de investigación ante el CDCHT-ULA. Octubre 2000.
- Ahmed, K., Ayers, D., et al. (2001). *Professional XML Meta Data*. Birmingham, UK: Wrox Press.
- Baker, K. S., Benson, B. J., Henshaw, D. L., Blodgett, D., Porter, J. H., Stafford, S. G. (2000). Evolution of a Multisite Network Information System: The LTER Information Management Paradigm. *BioScience*. Vol. 50, No11. November 2000. pp 963-978.
- Bayard, C. J., et al (2001). *Tempalte-driven End-user Ecological Database Design* LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://lternet-183.lternet.edu/doc_archive/
- Berners-Lee, T., Hendler, J. y Lassila, O. (2001). *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*. Scientific American. May 2001. Recuperado en Junio 2003, de: <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>
- Berners-Lee, T. (2003). *Frequently Asked Questions about RDF*. W3C organization. Recuperado en Junio 2003, de: <http://www.w3.org/RDF/FAQ>
- Brunt, J. W., et. al. (2001). *The Future Of Informatics In Long Term Ecological Research*. LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://lternet-183.lternet.edu/doc_archive/
- Baker, K. S., Brunt, J. W., Blankman, D. (2001). *Organizational Informatics: Site Description Directories for Research Networks*. LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://lternet-183.lternet.edu/doc_archive/
- CEN (1999). Comité Européen de Normalisation, (European Committee for Standardization) *Model for Metadata for Multimedia Information*. [Workshop Agreement]. 1999. Central Secretariat, Brussels. Belgium.
- CIP-LOC (2002). *Cataloging in Publication Program*. Library of Congress Cataloging in Publication Division. Recuperado en Junio 2003, de: <http://cip.loc.gov/cip/> y de <http://lcweb.loc.gov/catdir/>
- CIP-NLA (2003). *National Lybrary of Australia CIP System*. National Library of Australia, Canberra. Australia. 2003. Recuperado en Junio 2003, de: <http://www.nla.gov.au/services/CIP.html>
- CM-LSAL (2001). Carnegie Mellon Learning Systems Architecture Lab. *Metadata Tools*. Carnegie Mellon University. Recuperado en Junio de 2003, en: <http://www.lsal.cmu.edu/>
- CNRS (2003). California National Reserve System. *University of California*. Recuperado en Junio 2003 de: <http://nrs.ucop.edu/>
- Bell, P. J. (2002) *Cataloging in Publication*. Recuperado en Junio 2003, de: <http://members.aol.com/catpawpress/PubCip.html>

Databits (2003). *DataBits: An electronic newsletter for Information Managers*. Spring 2003 Issue
Recuperado en Junio de 2003, en:
<http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/03spring/>

Databits (2002). *DataBits: An electronic newsletter for Information Managers*. Spring 2002 Issue
Recuperado en Junio de 2003, en:
<http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/02spring/>

Day, M. (2002). *Metadata. Mapping between metadata formats*. UKOLN: The UK Office for Library and Information Networking, University of Bath, Bath, United Kingdom. Recuperado en Junio 2003, de:
<http://www.ukoln.ac.uk/metadata/interoperability/>

Dekkers, M. (2001). *Application Profiles, or how to Mix and Match Metadata Schemas*. Cultivate Interactive. January 2001. Recuperado en Junio 2003, de: <http://www.cultivate-int.org/issue3/schemas/>

DIMES (1997). *Data and Information management in the Ecological Sciences: Synopsis from a field station perspective*. DIMES workshop. August 8, 1997.

Duval, E., Hodgins, W., Sutton, S., Weibel, S. (2002). *Metadata Principles and Practicalities*. D-Lib Magazine. Vol. 8, Num. 4. Recuperado en Feb. 2003, de:
<http://www.dlib.org/dlib/april02/weibel/04weibel.html>.

Eco-Red Venezuela (2002). *Red Venezolana de Estaciones de Investigación Ecológica a Largo Plazo*. Recuperado en Junio 2003, de:
<http://cesimo.ing.ula.ve/LA-ECO-RED/ECORED/ecored.shtml>

Farrell, P. (November 2002). *Interpreted Languages*. School of Earth Sciences, Stanford University. Recuperado en Junio de 2003, de:
<http://pangea.stanford.edu/computerinfo/unix/programming/interpreted.shtml>

Federal Geographic Committee. *FGDC Metadata Tools*. (s.f.). Recuperado en Junio 2003, de:
<http://www.fgdc.gov/metadata/metatool.html>

GCMD (2000). Global Change Master Directory. *The State of the GCMD - CY 2000*. National Agency of Spatial Administration, NASA. Recuperado en Junio 2003, de:
<http://gcmd.gsfc.nasa.gov/Aboutus/state/state00.html>

Jones, M. B. (2003) *A brief overview of Ecological Metadata Language*. DataBits: An electronic newsletter for Information Managers. Spring 2003 Issue. Recuperado en Junio de 2003 de:
<http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/03spring/>

Henshaw, D. L., Spycher, G., Remillard, S. M. (2001). *Transition from Legacy Databank to an Integrated Ecological Information System*. LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://lternet-183.lternet.edu/doc_archive/

IEEE-LOM (2002). *Draft Standard for Learning Object Metadata*. Learning Technology Standards Committee of the IEEE. Institute of Electrical and Electronics Engineers, Inc. 2002. 3 Park Avenue. New York, NY, USA. Recuperado en Junio 2003, de:
http://ltsc.ieee.org/doc/wg12/LOM_1484_12_1_v1_Final_Draft.pdf

-
- ILTER (2000). *The International Long Term Ecological Research Network 2000, Perspectives*. U.S. LTER Network Office. University of New Mexico, Department of Biology, Albuquerque New Mexico, USA. 2000.
- KNB (2003). Knowledge Network of Biocomplexity. *EML Frequently Asked Questions*. Recuperado en Junio 2003, de: <http://knb.ecoinformatics.org/software/eml/eml-2.0.0/docs/eml-faq.html>
- Lagoze, C. (1996). *The Warwick Framework. A Container Architecture for Diverse Sets of Metadata*. Digital Library Research Group, Cornell University. D-Lib Magazine, July/August 1996. ISSN 1082-9873. Recuperado en Junio 2003, de: <http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>
- LOC-MARC (2003). *MARC Standards*. Network Development and MARC Standards Office, 2003. Library of Congress, USA. Recuperado en Junio 2003, de: <http://www.loc.gov/marc/>
- López, N., Migueis, J.(1998). *Integrar UML en los proyectos*. Ediciones Gestión 2000, Barcelona, 1998.
- LTER (2003). *Long Term Ecological Research Network*. Recuperado en Junio 2003, de: <http://www.lternet.edu>
- LTER-Luquillo (2001). *Luquillo Experimental Forest Information Management System*. Recuperado en Junio 2003, de: <http://luq.lternet.edu/datamng/index.html>
- Martin R. (1998). *UML Use Case Diagrams*. Engineering Notebook. C++ Report. Nov-Dec 98. Object Mentor Inc. Recuperado en Junio 2003, de: <http://www.objectmentor.com/resources/articles/usecases.pdf>
- McCartney, P. (2000). *Report of the Long-Term Ecological Research Metadata Committee Meeting*. February 2000. LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://intranet.lternet.edu/committees/information_management/
- McCartney, P. H. (2001). *Using XML-encoded Metadata as a Basis for Advanced Information Systems for Ecological Research*. LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://lternet-183.lternet.edu/doc_archive/
- Michener, W. K., Brunt, J. W., Vanderbilt, K. L. (2001). *Ecological Informatics: a Long Term Ecological Research Perspective*. LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://lternet-183.lternet.edu/doc_archive/
- Michener, W. K., et al (1997). *Nongeospatial metadata for the ecological sciences*. Ecological Applications. 7(1). 1997. pp 330-342.
- Michener, W. K (1999). *Ecological Metadata*. LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://lternet-183.lternet.edu/doc_archive/
- Naumann, F., Rolker, C. (1999). *Do Metadata Models meet IQ Requirements?*. Berlin-Brandenburg Graduate School in Distributed Information Systems. Berlín, Germany. Recuperado en Feb. 2003 de: <http://citeseer.nj.nec.com/naumann99do.html>
- Nebert, D. (1997). *The Geospatial Data Clearinghouse. Towards a Common Spatial Data Infrastructure*. Federal Geographic Data Committee, March 1997.

- NISO. (2001). National Information Standards Organization. *Metadata Made Simpler*. [Booklet]. Hodge, G.: Autor
- NOAA-CSC (2003). National Oceanic and Atmospheric Administration. *The Wonderful World of Coastal Metadata*. NOAA Coastal Services Center, 2234 South Hobson Avenue, Charleston, SC. USA. Recuperado en Junio de 2003, en: <http://www.csc.noaa.gov/metadata/>
- Nottrott R. W., Jones M. B. (2001) *Using XML-Structured Metadata to Automate Quality Assurance Processing for Ecological Data*. National Center for Ecological Analysis and Synthesis. University of California Santa Barbara. Santa Barbara, CA, Recuperado en Febrero de 2003, de: http://lternet-183.lternet.edu/doc_archive/
- OBFS (2003). Organization of Biological Field Stations. Recuperado en Junio 2003, de: <http://www.obfs.org/>
- OMG. (2003). Object Modelling Group. *UML Resource Center*. Recuperado en Junio 2003, de: <http://www.omg.com/uml>
- Porter, J., Henshaw, D., Stafford, S. (1997) *Research Metadata in Long-Term Ecological Research (LTER)*. Proceedings of Metadata 97. 1997. Recuperado en Feb. 2003, de: <http://computer.org/proceedings/meta97/papers/jporter/jporter.html>
- Porter, J. H., Ramsey, K. (2001a). *Integrating Ecological Data: Tools and Techniques*. LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://lternet-183.lternet.edu/doc_archive/
- Porter, J. H., Brunt, J. (2001b). *A Network System for Long-Term Ecological Research (LTER) Data and Information Management*. LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://lternet-183.lternet.edu/doc_archive/
- Porter, J.H.(2001) *Enabling New Research and Collaboration through Sharing Ecological Information*. Bonanza Creek LTER Meeting, Fairbanks, AK, Oct. 2001. Recuperado en Junio 2003, de: http://atlantic.evsc.virginia.edu/presentations/BNZ2001IM_files/frame.htm
- SDSC-LTER (2002). *San Diego Supercomputer Center and Long Term Ecological Research.Network Collaborative Research*. Recuperado en Junio 2003, de: <http://www.sdsc.edu/sdsc-lter/>
- Smith, D. J., Benson, B. J., Balsiger, D. F. (2001) *Designing Web Database Applications for Ecological Research*. LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://lternet-183.lternet.edu/doc_archive/
- Somerville, I., (2000). *Software Engineering*. Addison-Wesley Pub Co; 6th edition (August 11, 2000)
- Spedding V. (2000). *The drive for world-wide ecological databases*. Scientific Computing World, April May/2000:16-19.
- Stafford, S., Kaplan, N. E., Bennett, Christopher (2001). *Through the Looking Glass: Wath do we see, What have learned, What can we share?*. LTER NET Office, Albuquerque. NM, US. Recuperado en Feb. 2003, de: http://lternet-183.lternet.edu/doc_archive/
- Tauber, J. (2002) *Metadata, Archival, Genealogy*. SCHEMA.NET. Recuperado en Septiembre 2002, de: <http://www.schema.net/metadata/>

Tomasic, A., Simon E. (1997). *Improving Access to Environmental Data using Context Information*. INRIA Rocquencourt. SRAE - Ministère de l'Environnement. Recuperado en Marzo 2003, de: <http://www.first.gmd.de/applications/tomasic/ercim.html>

Torrens R. (2001). *Desarrollo de Sistemas de Información Bio-climática, Propuesta de Tesis de Grado*. Postgrado de Computación. Universidad de Los Andes. Mérida.

UNDP-UNSO Program. *Environmental Information Systems*. Recuperado en Noviembre de 2002, de: <http://www.undp.org/seed/unso/prog/eis/sld001.htm>

US-CODATA (1995). U.S. National Committee for CODATA. *Finding the Forest in the Trees. The Challenge of Combining Diverse Environmental Data* (Committee for a Pilot Study on Database Interfaces). Washington, DC: Washington Academy Press.

UKOLN (2001). *Metadata Watch Report #3*. Forum for Metadata Schema Implementers. UK Office for Library and Information Networking – UKOLN (2001). Recuperado en Junio 2003, de: <http://www.schemas-forum.org/metadata-watch/third/section2.html>

W3C-SW (2003). *Semantic Web. W3C Semantic Web Activity*. Recuperado en Junio 2003, de: <http://www.w3c.org/2001/sw/>

WYGISC (1999). *Development of Metadata Education Strategies. Wyoming Geographic Information Science Center (WyGISC)*. University of Wyoming. Recuperado en Junio 2003, de: <http://www.sdvc.uwyo.edu/metadata/education.html>

XML-W3C (2003). *Extensible Markup Language (XML)*. W3C XML Core Working Group. Recuperado en Junio 2003, de: <http://www.w3.org/XML/>

Z39.50 (2001). *The Z39.50 Document*. Z39.50 Network Development & MARC Standards Office, 2001. Library of Congress, USA. Recuperado en Junio 2003, de: <http://lcweb.loc.gov/z3950/agency/document.html>

6 Anexos

6.1 Tabla de contenidos detallada de taller básico de manejo de datos y metadatos⁶⁵

I. Base Teórica

Sección 1.

Metadatos. Que son? Porque son importantes?

- 1.1 Que significa la palabra “metadatos”?
- 1.2 Importancia de los metadatos
- 1.3 Formas de metadatos
- 1.4 Qué es una colección de datos?
- 1.5 Ubicando colecciones de datos en Internet
- 1.6 Quien crea o genera los metadatos?
- 1.7 Información y recursos adicionales sobre metadatos
- 1.8 Ejemplos de metadatos y de comunidades que usan metadatos.
- 1.9 Búsqueda de metadatos

Sección 2.

Estándares de contenido para Metadatos. Lenguaje de Metadatos para la Ecología (EML)

- 2.1 Estándares para metadatos. Para que estándares? Cuales estándares existen?
- 2.2 Lenguajes y formatos de representación de metadatos
- 2.3 Características básicas de EML. Recursos en el Web sobre EML.
- 2.4 Preguntas más frecuentes sobre metadatos, estándares y sobre EML.

Sección 3.

Donde comenzar? Tareas iniciales

- 3.1 Hacer un inventario de los datos
- 3.2 Categorizar y dar prioridades a las colecciones de datos
- 3.3 Seleccionar los procedimientos y herramientas a utilizar

Sección 4.

Selección de las herramientas más apropiadas

- 4.1 Decidiendo entre el uso de Bases de datos o sistemas de archivos
- 4.2 Herramientas de software y hardware necesarias
- 4.3 Herramientas de metadatos disponibles y como obtenerlas

Sección 5.

Tareas básicas para la creación de colecciones de datos

- 5.1 Escogiendo e instalando las herramientas
- 5.2 Usando una herramienta de creación de metadatos
- 5.3 Ejercicio de creación de metadatos
- 5.4 Validación y revisión de metadatos

Sección 6.

Compartiendo los metadatos. Otros aspectos sobre los metadatos.

- 6.1 Que voy a hacer con los metadatos? A quien le sirven los metadatos? (uso local vs uso global)

⁶⁵ Ver material completo de este taller en: <http://www.cecalc.ula.ve/webclima/tallermetadatos/>

6.2 Haciendo que sus metadatos se puedan intercambiar

6.3 Mantenimiento y gestión de colecciones de datos

II. Actividades Prácticas

Práctica Sección 1.

Búsqueda de datos y metadatos en Internet. Uso de *clearinghouses* de datos.

Práctica Sección 2.

Análisis del Lenguaje de Metadatos para la Ecología (EML).

Práctica Sección 3.

Tareas iniciales de organización de una colección de datos propia.

Práctica Sección 4.

Descripción y demostración de herramientas disponibles. *Morpho*, Interfaz Web, etc.

Práctica Sección 5.

Creación de una colección de datos de la ECO-RED Venezuela y la Red de Estaciones del Edo. Mérida usando interfaz Web y *Morpho*.

Práctica Sección 6.

Compartiendo los metadatos. Publicando datos y metadatos en Internet.

Indice

A

35, 37

Amaya.....51
arquitectura cliente-servidor.....59, 69

C

Centro Europeo de Estándares-CEN17
CGI.....61, 62, 69, 90
35, 45, 46, 47
35, 46, 47
Clearinghouse de datos45
Clearinghouse FGDC47
Cliente de captura de metadatos.....60
colecciones de datos53

D

Descriptores Genéricos de Metadatos28
Dublin Core18, 22

E

Eco-Red Venezuela.....71, 72, 73
35, 44, 53, 69, 77, 78, 87
Editores XML51
EML 2, 18, 30, 31, 32, 33, 35, 40, 41, 42, 43, 44,
50, 52, 56, 61, 62, 63, 69, 79, 86, 87
Ecological Metadata Language30
EML 2.031
eml-dataset32, 50, 61
eml-dataset32
35
eml-literature.....32
módulos de EML.....31, 32
Estándar de metadatos.....27
Estándares.....13
estructuras de metadatos.....17

F

FGDC ..10, 13, 18, 21, 25, 26, 27, 35, 41, 43, 50,
82, 90
formulario *HTML*63

G

35, 36
35, 44

13, 35

Global Change Master Directory GCMD. 10, 16,
27

H

herramientas.. 2, 7, 10, 12, 13, 16, 25, 26, 33, 34,
35, 36, 39, 40, 41, 42, 43, 44, 49, 51, 52, 53,
54, 69, 70, 71, 73, 74, 75, 77, 78, 81, 86
herramientas de búsqueda..... 45
herramientas de uso general..... 42
HTML..... 20, 23, 24, 35, 57, 61, 62, 63, 66, 69
HTTP 27, 35

I

interfaces Web 56

J

JavaScript..... 61, 64

K

KNB 52, 75
KNB (Knowledge Network of Biocomplexity). 41
35, 44
Knowledge Network of Biocomplexity-KNB . 30

L

ILTER 8, 9, 10, 13, 14, 27, 28, 30, 33, 35, 40, 41,
42, 44, 49, 52, 61, 69, 70, 72, 81, 82, 83, 84,
85, 90

M

manejo de datos y metadatos..... 36
Marco de Referencia de Warwick..... 22
Metacat 33, 35, 41, 52, 53
metadata crosswalks..... 68
metadatos
Lenguaje de codificación de metadatos 56
metadatos 2, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18,
19, 20, 21, 22, 23, 24, 25, 26, 27, 29, 30, 31,
33, 34, 35, 36, 37, 38, 40, 43, 44, 45, 46, 47,
48, 49, 50, 52, 53, 56, 59, 61, 62, 69, 70, 71,
74, 76, 77, 78, 86, 87
ciclo de vida de los metadatos 42
clases de descriptores de metadatos..... 28
estándares de metadatos..... 27, 32
formato de metadatos..... 56
interoperabilidad de metadatos 41
mantenimiento de metadatos 32

<i>metadata crosswalks</i>	56	Red de Estaciones Bio-climáticas del Edo. Mérida.....	12
metadatos estructurados	42	Red de Estaciones del Edo. Mérida	20, 35, 44, 53, 69, 71, 72, 73, 75, 77, 78, 87
meta-metadatos.....	45	Roles en el manejo de metadatos.....	17
re-uso de metadatos	38	S	
sistemas de metadatos	45	servidor Web	54, 69
Metodologías	54	T	
Michener.....	9, 13, 14, 27, 28, 30, 35, 40, 50, 83	talleres.....	77
minería de datos	79	talleres de manejo de información	76
<i>Morpho</i>	33, 41, 43, 52, 53, 71, 74, 75, 76, 77	U	
<i>MSXSL</i>	56	UML	56, 69, 83, 84
N		X	
National Center for Ecological Analysis and Síntesis-NCEAS	30	<i>Xanthoria</i>	52
O		<i>Xeena</i>	51
<i>Open Archives Initiative</i>	48	XML ...	22, 23, 24, 27, 31, 32, 35, 43, 48, 49, 50, 51, 52, 56, 57, 61, 62, 63, 66, 69, 81, 84, 90
P		XML Edit Pro	51
perfiles de aplicación.....	22	XML Notepad.....	51
PHP	55, 61, 62, 69	XML Schemas	51
prototipos.....	53	XMLSpy	53
modelo de prototipos	55	XMLSpy	51
Q		Xpath.....	52
QA/QC.....	30, 35, 42	XPath	52
R		XSL	27, 35, 49, 51, 52, 56, 63, 67, 68, 69
recolección de metadatos.....	63		