

**PLATAFORMA PARA LA INTEGRACIÓN DE MÚLTIPLES FUENTES
DE DATOS DESCRIPTIVOS Y ESPACIALES EN UN SISTEMA DE
INFORMACIÓN AMBIENTAL**



Tesis de Maestría

Ing. Juan Carlos Corrales Muñoz

Director

Dr. Ing. Álvaro Rendón Gallón

Universidad del Cauca

Maestría en Ingeniería, Área Telemática

Facultad de Ingeniería Electrónica y Telecomunicaciones

Departamento de Telemática

Línea de investigación en Ingeniería de Sistemas Telemáticos

Aplicaciones soportadas en Internet

Popayán, Noviembre de 2004

CONTENIDO

	Pág.
1. INTRODUCCIÓN	5
2. GESTIÓN E IMPACTO AMBIENTAL	8
2.1 La Gestión Ambiental	8
2.1.1 Los Actores y la participación en la Gestión Ambiental	10
2.2 Impacto Ambiental.....	11
2.2.1 Conceptos Generales de la Evaluación de Impacto Ambiental	11
2.3 Importancia de la Información para los Procesos de Gestión Ambiental.....	14
3. SISTEMAS DE INFORMACIÓN AMBIENTAL Y TECNOLOGÍAS QUE FACILITAN LA CONSTRUCCIÓN DE UN ENTORNO DE INTEGRACIÓN DE MÚLTIPLES FUENTES DE DATOS	16
3.1 Sistemas de Información para la Gestión Ambiental.....	16
3.1.1 Sistemas de Información Geográfica	17
3.1.2 Sistemas de Gestión de Bases de Datos.....	17
3.1.3 Intercambio de Datos e Información	17
3.2 Interoperabilidad de Sistemas de Información	19
3.2.1 Conceptos Fundamentales	20
3.2.2 Arquitecturas de Interoperabilidad	22
3.3 Mecanismos de Acceso a Datos Distribuidos	30
4. MODELO DE DATOS	35
4.1 Sistemas de Información Geográfica	35
4.1.1 Modelos Espaciales.....	35
4.2 Sistema de Gestión de Bases de Datos	41
4.2.1 Modelo Entidad – Relación	41
4.2.2 Modelo Relacional	46
5. ARQUITECTURA DE LA PLATAFORMA	51
5.1 Introducción.....	51
5.2 Arquitectura Conceptual	52
5.3 Funcionalidades	54
5.3.1 Módulo Wrapper	55
5.3.1.1 Árbol de Funciones.....	55
5.3.1.2 Descripción de Casos de Uso.....	55
5.3.2 Módulo Mediador.....	60
5.3.2.1 Árbol de funciones.....	60
5.3.2.2 Descripción de Casos de Uso.....	61
5.4 Arquitectura física.....	65
5.4.1 Descripción de las Capas de Aplicación, Lógica de Mediación y Software ...	66
5.4.1.1 Módulo Wrapper	66
5.4.1.2 Módulo Mediador	68
5.4.2 Diagrama de Implantación	70

6. PROCESO DE INTEGRACIÓN DE LOS SISTEMAS DE GESTIÓN DE INFORMACIÓN	73
6.1 Configurar Conexión SIG.....	73
6.2 Publicar Información.....	74
6.2.1 Navegación en los Metadatos.....	75
6.2.2 Publicación de Objetos.....	77
6.2.3 Publicación de Atributos	78
6.3 Descriptor Local	82
6.3.1 Actualizar Metadatos	82
6.3.2 Creación del Descriptor Local.....	83
6.4 Acceso al Mediador	83
6.5 Visualización de información	85
7. CONCLUSIONES Y RECOMENDACIONES	90
BIBLIOGRAFÍA	93

ÍNDICES DE TABLAS

	Pág.
Tabla 4.1 – Comparación entre Modelos	41
Tabla 5.1 – Descripción caso de uso consultar información	58
Tabla 5.2 – Descripción caso de uso resolver consulta	63

ÍNDICE DE FIGURAS

Figura 3.1 - Arquitectura del Esquema Global para Múltiples Fuentes de Datos.....	24
Figura 3.2 – Arquitectura para los Lenguajes Múltiples de Bases de Datos	25
Figura 3.3 – Esquema de Federación	26
Figura 3.4 – Clasificación de Esquemas Federados	27
Figura 3.5 – Arquitectura del Sistema Mediador	28
Figura 3.6 – Componentes del Sistema Mediador	29
Figura 3.7 – Arquitectura de los Servicios Web.....	31
Figura 3.8 - Pila de Protocolos de los Servicios Web	33
Figura 4.1 - Representación gráfica de los modelos Vector y Raster	39
Figura 4.2 – Representación del Modelo Entidad-Relación.....	42
Figura 4.3 – Ejemplo del Modelo Entidad-Relación.....	45
Figura 4.4 – Tabla Relacional	47
Figura 4.5 – Relaciones entre Tablas.....	49
Figura 5.1 - Sistema Integrado de Información Ambiental.....	51
Figura 5.2 – Esquema Conceptual de la Arquitectura	53
Figura 5.3 – Diagrama de Casos d+e Uso Módulo Wrapper	57
Figura 5.4 – Diagrama de clase C.U. Consultar Información.....	59
Figura 5.5 – Diagrama de Casos de Uso Módulo Mediador	62
Figura 5.6 – Diagrama de clase C.U. Resolver Consulta	65
Figura 5.7 – Diagrama capas de aplicación, lógica de mediación y software módulo wrapper.....	67
Figura 5.8 – Diagrama capas de aplicación, lógica de mediación y software módulo mediador.....	68
Figura 5.9 – Diagrama de Implantación	71
Figura 6.1 - Interfaz de Parámetros de Conexión.....	73
Figura 6.2 - Mensaje de resultado de la conexión a la Fuente de Datos	74
Figura 6.3 - Interfaz del Menú de Publicación	75
Figura 6.4 - Interfaz de Listado de Objetos	76
Figura 6.5 - Interfaz de Publicación de Objetos.....	77
Figura 6.6 - Mensaje de Publicación Exitosa.....	78
Figura 6.7 - Interfaz de listado de Atributos de un objeto no publicado	79
Figura 6.8 - Interfaz de Listado de Atributos de un Objeto publicado	80
Figura 6.9 - Interfaz de Publicación de Atributos.....	81
Figura 6.10 - Mensajes Obtenidos del proceso de Publicación	82
Figura 6.11 - Menú Descriptor Local.....	83
Figura 6.12 - Documento WSDL de Acceso al Módulo Mediador.....	84
Figura 6.13 - Metadatos de la Plataforma	85
Figura 6.14 – Listado de mapas.....	86
Figura 6.15 – Mapa de Sismos	87
Figura 6.16 – Mapa de Cuencas	88

1. INTRODUCCIÓN

El desarrollo, prevención y ordenamiento de los recursos naturales y de las poblaciones que con ellos interactúan es una tarea fundamental que los países en desarrollo deben afrontar mediante el diseño participativo de alternativas de gestión y planificación del territorio. Para lograr que esta tarea sirva a los fines de fomentar una prosperidad en los aspectos sociales, naturales, ambientales, urbanos y económicos, tales alternativas han de considerar el uso ecuánime y adecuado de las potencialidades físicas y naturales, concretando así una opción viable de desarrollo humano sostenible para las comunidades.

En el ámbito regional, la dependencia esencial que los asentamientos humanos tienen del agua y otros recursos naturales requeridos para su desarrollo, provenientes de los ecosistemas estratégicos protegidos, es cada día más ostensible y plantea serios conflictos para el bienestar social y el desarrollo económico de la región. Por ello, el manejo de información sobre las variables ambientales, incluyendo la disponibilidad de datos en tiempo real para determinar la evolución de los componentes del medio natural, es una necesidad y prioridad para un país como Colombia, donde procesos como la deforestación y la erosión de los suelos han generado problemas de gran importancia como inundaciones, crecientes, avalanchas, alteración del ciclo hidrológico y desastres naturales, causando en suma la degradación de nuestros ecosistemas y afectando la calidad de vida de nuestras comunidades.

Frente a esta realidad es perentorio desarrollar un programa de investigación que permita:

- el levantamiento de la línea de base ambiental de ecosistemas estratégicos protegidos,
- la adquisición de información ambiental en tiempo real, y su almacenamiento y procesamiento de forma rápida,
- el desarrollo de un sistema de alerta temprana sobre cambios ambientales de origen natural o inducidos,
- la descripción del comportamiento de los ecosistemas en cuanto a los flujos de nutrientes, y
- el desarrollo de modelos conceptuales que posibiliten la formulación de alternativas de gestión ambiental de los recursos naturales para ecosistemas estratégicos.

Este tipo de programa se concreta finalmente en planes para el desarrollo, prevención de riesgos y desastres, y ordenamiento de los recursos naturales y de las sociedades que interactúan en una determinada zona o región.

Pese a que se ha reconocido su importancia, estos programas no han podido ser llevados a la práctica con éxito debido a múltiples factores, entre los cuales se destacan:

- Dispersión y desarticulación de información y de esfuerzos institucionales.
- Vacío de información cartográfica geocientífica básica y temática (descriptiva, diagnóstica y prescriptiva) en niveles de escala que permitan propuestas de

ordenamiento ambiental (generación de alternativas sostenibles de uso del suelo, ecoturismo, investigación.)

- Diferentes percepciones y visiones sobre el territorio que generan incompatibilidad de la información biofísica y socio-económica.
- Falta de información biofísica, geomorfológica y socio-cultural que permita consolidar los planes de manejo ambiental.
- Carencia de información para la alimentación del sistema de información ambiental, que implica la recolección, almacenamiento y procesamiento de datos en tiempo real.
- Poca capacidad de interacción regional de las instituciones involucradas en el proceso de gestión ambiental de la zona.

En este contexto, el Grupo de Ingeniería Telemática y el Grupo de Estudios Ambientales han venido desarrollando, con el apoyo de la Vicerrectoría de Investigaciones, un proyecto de investigación denominado ARIADNA¹, con el propósito de estudiar estos problemas y brindarles soluciones soportadas en el uso de las tecnologías de la información y las comunicaciones. Este proyecto estuvo dentro de los finalistas del Concurso de Estocolmo (Stockholm Challenge) en el año 2002, en la categoría de medio ambiente², y luego de sus primeros resultados ha logrado interesar a entidades del sector como INGEOMINAS y la Unidad Especial de Parques Nacionales Naturales.

En la actualidad el proyecto ARIADNA (Sistema Integrado de información Ambiental) cuenta con financiación de la Vicerrectoría de Investigaciones de la Universidad del Cauca y COLCIENCIAS mediante Contrato 026/02 suscrito entre el Centro de Investigación de las Telecomunicaciones - CINTEL y la Universidad del Cauca. Éste proyecto tiene como objetivo general diseñar e implementar un prototipo de un Sistema Integrado de Información Ambiental, que permita el seguimiento de variables ambientales en un ecosistema estratégico protegido, la formulación de planes de desarrollo, prevención de riesgos y desastres, y el ordenamiento de los recursos naturales para dicho ecosistema.

En el marco del anterior proyecto se encuentra la Tesis de Maestría en Ingeniería, Área de Telemática titulada “PLATAFORMA PARA LA INTEGRACIÓN DE MÚLTIPLES FUENTES DE DATOS DESCRIPTIVOS Y ESPACIALES EN UN SISTEMA DE INFORMACIÓN AMBIENTAL”. Este trabajo se concentra en la generación de una plataforma de integración para un Sistema de Información Ambiental soportado en un Sistema de Información Geográfica, que facilite la concepción y elaboración de planes de manejo ambiental en una zona protegida³ con el fin de alcanzar el desarrollo, ordenamiento y gestión de los recursos naturales de dicha zona.

La plataforma de integración permitirá que las instituciones involucradas en los procesos de gestión ambiental de una zona en común puedan concentrar la información sobre el ecosistema estratégico analizado, articular dicha información a los procesos de ocupación del espacio, socializarlos y ponerlos a disposición en forma expedita para procesos de investigación y desarrollo regional, optimizando así el esfuerzo y la orientación de los procesos de investigación prioritarios para cada institución.

¹ Proyecto ARIADNA: Adquisición Remota de Información Ambiental para Diagnóstico y gestión de recursos Naturales. Códigos VRI 139 (2000) y 524 (2001).

² http://www.challenge.stockholm.se/finalists_index.html.

³ P1= N:740000, E:1085000, Alto Yarumal; P2= N:755000, E:1085000, Laguna de Sánchez; P3= N:755000, E:1065000, Patíco; P4= N:740000 E:1065000, Santa Rosa. Área: 300 Km².

De ésta manera se concreta una opción viable de desarrollo humano sostenible a través de la utilización de la Electrónica, Telecomunicaciones e Informática como elemento esencial en la generación de valor agregado a los recursos naturales y a las comunidades asentadas en la zona de influencia del proyecto.

En el siguiente capítulo de ésta monografía se realiza un análisis crítico y descriptivo del proceso actual de Gestión e Impacto Ambiental y de las herramientas para el buen uso de los recursos naturales. En el capítulo 3 se elabora un estado del arte sobre las tecnologías a utilizar en el desarrollo del trabajo. El capítulo 4 describe los modelos de datos elegidos (SIG-SGBD) para la implementación de la plataforma de integración. El capítulo 5 contiene una descripción detallada de arquitectura a utilizar para la plataforma. El capítulo 6 consigna el proceso de integración de los sistemas de gestión de información mediante la plataforma desarrollada en el presente proyecto. El último capítulo se exponen las conclusiones que se obtuvieron tras el desarrollo de toda la temática del trabajo y se darán también las recomendaciones sobre el uso de las diferentes tecnologías incorporadas en el proyecto.

2. GESTIÓN E IMPACTO AMBIENTAL

Este capítulo contiene un análisis crítico de las herramientas conceptuales que facilitan los procesos de Gestión Ambiental, tales como Impacto Ambiental y Evaluación de Impacto Ambiental. El análisis considera la información como insumo fundamental para dichas herramientas, las cuales tienen el objetivo de alcanzar el buen uso de los recursos naturales.

2.1 La Gestión Ambiental

Entendemos como gestión ambiental al conjunto de acciones emprendidas por la sociedad, o parte de ella, con el fin de proteger el medio ambiente. Sus propósitos están dirigidos a modificar una situación actual a otra deseada, de conformidad a la percepción que sobre ella tengan los actores involucrados. La gestión ambiental no solamente está referida al gobierno, sino que crecientemente depende de fuerzas sociales de muy diversa naturaleza, tal como lo evidencian diversos estudios sobre Latinoamérica y el Caribe [EAS92], [ESC99]. En su concepción más amplia, la gestión ambiental es un proceso permanente y de aproximaciones sucesivas en el cual diversos actores públicos y privados y de la sociedad civil desarrollan un conjunto de esfuerzos específicos con el propósito de preservar, restaurar, conservar y utilizar de manera sustentable el medio ambiente.

Por su injerencia amplia, la gestión ambiental puede ser abordada bajo diversas perspectivas y con diferentes escalas. Por ejemplo, se puede centrar en el ámbito rural o urbano, en una política específica (ej. contaminación del aire de un centro urbano, etc.), en una amenaza ambiental global (ej. impacto de emisiones sobre el calentamiento de la tierra, etc.), en el impacto ambiental de una actividad económica específica (ej. minería, energía, agricultura, etc.), o en la conservación y uso sostenible de un recurso estratégico (ej. bosques, aguas, etc.). La gestión ambiental, por lo tanto, puede ser abordada a distintos niveles de gobierno (central, departamental, municipal, etc.), o de grupos del sector privado en su concepción amplia, o en diversos ámbitos territoriales (global, regional, subregional, nivel metropolitano, ciudades, barrios, poblados, cuencas hidrográficas, etc.) [REW02].

La gestión ambiental parte de la necesidad de la sociedad de conservar y mejorar la "oferta y calidad ambiental", es decir, de los recursos que sirven para satisfacer las necesidades de los seres humanos, y que son fundamentales como soporte de la vida en la tierra. Ello conlleva el reto de detener y revertir el deterioro del medio ambiente con el fin de preservar y mejorar su calidad para las futuras generaciones. A partir de condicionantes ambientales y de tipo socioeconómico y político, y con base en la información disponible y las señales del entorno, los distintos actores públicos, privados y de la sociedad civil ponen en marcha políticas, planes, programas y proyectos para el cumplimiento de los objetivos planteados. En este proceso, los actores se desenvuelven en un determinado marco legal, y movilizan los recursos económicos, técnicos y humanos para la aplicación de diversos instrumentos. El grado de éxito o fracaso será determinado

por el nivel de compromiso que los actores tengan con los objetivos perseguidos, la solidez del proceso decisorio, el soporte económico, técnico y financiero disponible, y las condicionantes de orden político, económico y social particulares a cada país o entidad subnacional. El grado de éxito también se asocia con la participación ciudadana, la descentralización, y la integración-coordinación de la temática ambiental con otras políticas públicas, así como de las agencias ambientales con los otros actores del Estado, del sector privado y de la sociedad civil. La naturaleza del problema a resolver, la influencia internacional, y la oportunidad que surge en un momento para la aplicación de políticas e instrumentos, son factores que en gran medida influyen en los resultados del proceso [FIO95], [JWE97].

Al abordar la temática ambiental en países latinoamericanos y del Caribe se debe reconocer la diversidad de condiciones y problemas que actualmente posee la región. Los países que la componen tienen territorios, poblaciones, economías y ecosistemas de una gran diversidad, lo cual hace que los retos ambientales adquieran tonalidades y dimensiones diferentes entre los distintos países o subregiones. Algunos de los principales aspectos de deterioro ambiental de la región, refiriéndose a temas de relevancia para un grupo de países o para uno de ellos en particular, son: la pérdida de biodiversidad; la deforestación, pérdida de suelos y desertificación; el deterioro de las costas y el medio marino; los recursos hidrológicos y la contaminación de agua por fuentes industriales, agrícolas y domiciliarias; la creciente generación e inadecuada disposición de residuos sólidos; el deterioro en la calidad del aire en las grandes ciudades; y los desastres naturales y la vulnerabilidad ambiental de la población [REW02].

La problemática ambiental de los países está influida por condiciones de desarrollo económico y social, las cuales forman parte del conjunto de aspectos a ser considerados dentro de las actividades de gestión ambiental. Las relaciones entre crecimiento de la economía y la gestión ambiental, los efectos de las políticas macroeconómicas en el estado del medio ambiente, y las relaciones con las condiciones sociales, son aspectos que deberán tomarse en cuenta al abordar los temas críticos de gestión ambiental y analizar las políticas ambientales que la región ha adoptado [CAM93], [ARR96], [OCD97]. Esto incluye la dinámica de crecimiento poblacional, el ritmo de desarrollo económico, el patrón de distribución del ingreso, los patrones de producción y consumo, el acceso a empleo, acceso a servicios de salud y educación, y un sin número de factores que influyen en las variables ambientales en la región. También es fundamental prever los efectos que los actuales modelos económicos de la región y sus tendencias, marcados por el proceso de globalización, puedan tener sobre el estado del medio ambiente.

Las condiciones sociales de un país y sus formas de organización social son factores de gran incidencia en el resultado de las políticas ambientales. Son bien conocidas las relaciones tanto entre el crecimiento poblacional, la pobreza de amplios sectores de la población y el deterioro ambiental [WCE87], [CDM90]. Más recientemente, comienza a tenerse una comprensión más clara de las relaciones entre medio ambiente y conflicto social, como de sus consecuencias para la gestión ambiental de los países de la región [FNA01], [HDT99].

No obstante la manifiesta relevancia del contexto económico, político y social en el logro de objetivos de gestión ambiental, aún persisten dificultades en la integración de todos estos ámbitos. Las políticas económicas y sociales de los países no logran integrar aspectos de sustentabilidad ambiental ni tampoco se ha insertado de manera amplia la dimensión ambiental dentro de las políticas sectoriales específicas (por ejemplo:

industrial, agrícola, energética, etc.). Esto se puede atribuir a la poca voluntad política y capacidad de los países, pero también a que la ciudadanía, y con frecuencia las instancias decisorias, desconocen cuáles son los riesgos incurridos en materia ambiental en el momento de adoptar políticas de desarrollo a nivel sectorial. Pero debe subrayarse que éste no es un problema que sea endémico a las sociedades de América Latina y el Caribe; se presenta también con diversos matices en países desarrollados [ARR96], [JWE97].

2.1.1 Los Actores y la participación en la Gestión Ambiental

Los actores de la gestión ambiental son aquellos grupos de la sociedad que identifican la problemática del medio ambiente, y que formulan e implementan políticas dirigidas a su protección. Se consideran actores claves de la gestión ambiental a: las organizaciones gubernamentales, los grupos organizados de la sociedad civil, los medios de comunicación masiva, y las empresas del sector privado que se han organizado explícitamente en pro de la protección ambiental [REW02].

En los países en desarrollo, la gestión pública ambiental se ha originado más como una iniciativa de carácter gubernamental, que como una respuesta a las demandas de la ciudadanía, hecho que contrasta con el caso de los países desarrollados en donde las demandas ciudadanas por la protección del medio ambiente están en el origen de buena parte de la acción del Estado [DES98]. Esta es una situación que se explica principalmente por la baja prioridad que aún tiene el tema ambiental dentro de amplios sectores de la población, los que se interesan más por la satisfacción de sus necesidades básicas de alimentación, vivienda y servicios básicos.

La gestión ambiental estatal en los países de la región empieza por la constitución de un marco jurídico y de instituciones públicas cuya responsabilidad es diseñar e implementar planes y programas ambientales. Este fundamento legal e institucional formaliza la gestión ambiental dentro de los países, a través de la consagración constitucional, la creación de normatividad ambiental, la asignación de funciones ambientales a diversas agencias públicas, la creación de procedimientos ambientales en sectores públicos, la adhesión a tratados multilaterales, y la creación de sistemas de gestión ambiental a cargo de autoridades con responsabilidades ambientales explícitas [REW02].

Por otro lado, el desarrollo legal/institucional de la gestión ambiental produjo en sus orígenes sistemas altamente centralizados, con funciones casi exclusivamente asignadas al orden central en los países. Sin embargo, y aunque de manera no sistemática, a través de los años, y como respuesta a los graves problemas derivados de la alta centralización de funciones ambientales y a la baja efectividad a nivel local en cuanto al diseño de políticas, la región muestra adelantos de procesos de descentralización en los que las autoridades locales han asumido responsabilidades directas de gestión ambiental. En la actualidad está plenamente justificada la prioridad de descentralizar la gestión ambiental a regiones, provincias y municipios para fines ambientales [PNU98].

2.2 Impacto Ambiental

Durante mucho tiempo el concepto de impacto ambiental fue acuñado para los temas de contaminación y también estuvo centrado en lo urbano; luego el concepto se hizo extensible a especies animales, vegetales y a ecosistemas. Por ello se puede definir ampliamente el impacto ambiental como la alteración significativa de los sistemas naturales, provocado por acciones humanas. Por tanto, los impactos se expresan en las diversas actividades y se presentan tanto en ambientes naturales como en aquellos que resultan de la intervención y creación humana [ESP02]. En conclusión, el Impacto Ambiental es una alteración significativa del ambiente de carácter positiva o negativa. Pueden ser clasificados en:

- Directos, cuando involucran pérdida parcial o total de un recurso o deterioro de una variable ambiental (contaminar aguas, talar bosques, etc.).
- Indirectos, cuando inducen y/o generan otros riesgos sobre el ambiente (erosión antrópica, inundaciones, etc.).

Ahora bien, el proceso mediante el cual se realiza un control del impacto ambiental producido por determinado evento se denomina Evaluación de Impacto Ambiental (EIA), el cual está definido como un proceso singular e innovador cuya operatividad y validez como instrumento para la protección y defensa del medio ambiente está recomendado por diversos organismos internacionales. También es avalado por la experiencia acumulada en países desarrollados, que lo han incorporado a su ordenamiento jurídico desde hace años.

El objetivo de la evaluación de impacto ambiental es prevenir situaciones de deterioro, estableciendo las medidas más adecuadas para llevar a niveles aceptables los impactos derivados de acciones humanas y proteger la calidad del ambiente. Una definición más general la relaciona con un proceso de advertencia temprana que permite aplicar anticipadamente las políticas ambientales. Se entiende acá como política a la definición de principios rectores y objetivos básicos que la sociedad se propone alcanzar en materia de protección ambiental, conciliándolos con los aspectos económicos, sociales y de desarrollo. La política establece las bases sobre las cuales se elaboran las leyes y reglamentos, los que, a su vez, generan un conjunto legitimado de cuerpos normativos que son instrumentos para alcanzar los objetivos descritos en ella [ESP02].

De los puntos anteriores se deduce que la EIA propugna por un enfoque a largo plazo y supone y garantiza una visión más completa e integrada del significado de las acciones humanas sobre el medio ambiente. También implica una mayor creatividad e ingenio y una fuerte responsabilidad social en el diseño y la ejecución de las acciones y proyectos. Dicha creatividad puede estar soportada en la aplicación de nuevas tecnológicas (electrónica, telecomunicaciones e informáticas) que den valor agregado a los procesos de planificación y de toma de decisiones.

2.2.1 Conceptos Generales de la Evaluación de Impacto Ambiental

A continuación se listan los principales conceptos a tener en cuenta en un proceso de Evaluación de Impacto Ambiental [ESP02]:

- La EIA es un análisis sistemático, reproducible e interdisciplinario de los impactos potenciales, tanto de una acción propuesta como de sus alternativas, en los atributos físicos, biológicos, culturales y socioeconómicos de un área geográfica en particular.
- La EIA tiene como propósito asegurarse de que los recursos ambientales de importancia se reconozcan al principio del proceso de decisión y se protejan a través de planeamientos y decisiones pertinentes.
- La EIA es un proceso de advertencia temprana y de análisis continuo que protege los recursos ambientales contra daños injustificados o no anticipados.
- El proceso de EIA convierte el lenguaje y la intención de las leyes ambientales en un grupo uniforme de requisitos técnicos y de procedimientos que permiten un análisis sistemático de las acciones humanas mucho antes de su ejecución.
- La implementación eficiente de un proceso de EIA requiere esfuerzos significativos, premeditados y cooperativos entre muchas entidades y actores.
- La EIA ayuda a la resolución de problemas en el proceso de toma de decisiones. No la reemplaza, sino que provee las bases para decisiones más informadas respecto a los impactos positivos y negativos sobre el medio ambiente.
- La efectividad del proceso de EIA depende del grado de prioridad nacional, regional o local que tenga la calidad ambiental.
- La adopción de un sistema de EIA puede conllevar a decisiones económicas difíciles y a compromisos políticos y sociales fuertes. El interés y consenso ciudadano, privado y público por la calidad ambiental suministra una base sólida para su desarrollo efectivo.

De acuerdo a los principales conceptos de EIA se generan los componentes que permiten llevar a cabo un adecuada EIA [ESP02]:

- Proceso de Evaluación de Impacto Ambiental (EIA), entendido como el conjunto de requisitos, pasos y etapas que deben cumplirse para que un análisis ambiental preventivo sea suficiente como tal según los estándares internacionales.
- Sistema de Evaluación de Impacto Ambiental (SEIA), entendido como la forma de organización y administración del proceso de EIA según la realidad y capacidad de quien lo aplique.
- Estudio de Impacto Ambiental, entendido como el o los documento(s) que sustenta(n) el análisis ambiental preventivo y que entrega(n) los elementos de juicio para tomar decisiones informadas en relación a las implicaciones ambientales de las actividades humanas.

En este contexto, el proceso de evaluación de impacto ambiental permite anticipar los futuros impactos negativos y positivos de acciones humanas, aumentando los beneficios y disminuyendo las alteraciones ambientales no deseadas. De esta manera el proceso de evaluación está de acuerdo con el concepto que busca asegurar que las variables ambientales de interés se reconozcan desde el inicio y se protejan a través de decisiones pertinentes. Así mismo se asegura coherencia entre las acciones previstas y las políticas y regulaciones ambientales que se hayan establecido con la finalidad de proteger el entorno.

El concepto que define la evaluación de impacto ambiental como un sistema de advertencia, afirma que ésta es un proceso informado y objetivo de decisiones

concatenadas y participativas, que ayudan a identificar las mejores opciones para llevar a cabo una acción sin daños ambientales inaceptables.

En este sentido, es importante destacar que un proceso de evaluación de impacto ambiental debe ser considerado como un instrumento que está al servicio de la toma de decisiones y que permite alcanzar, anticipadamente, un conocimiento amplio e integrado de los impactos o incidencias ambientales derivadas de acciones humanas.

Un proceso de evaluación de impacto ambiental no es en sí mismo un instrumento de decisión, sino que genera un conjunto ordenado, coherente y reproducible de antecedentes que permiten al promotor de un proyecto, a la autoridad competente y a la ciudadanía, en cada caso, tomar decisiones informadas y certeras. Todo ello se hace posible cuando se presenta un estudio de impacto ambiental y la autoridad respectiva lo somete a un proceso participativo de revisión para calificar la calidad del análisis [ESP02].

Es importante resaltar el carácter preventivo del proceso, ya que orienta la toma de decisiones en las etapas anteriores a la ejecución de la acción en cuestión. En este sentido el proceso de evaluación de impacto ambiental es siempre desarrollado en forma previa a la acción prevista. No tiene sentido llevar a cabo una evaluación para acciones ya implementadas o en etapas avanzadas de desarrollo, salvo que sean modificaciones de las existentes y se las considere como nuevas acciones. Como herramienta de predicción, la EIA adquiere valor sólo si puede influir en la planificación y desarrollo futuro de cada acción humana. Su aplicación debe iniciarse en la fase más temprana posible de la toma de decisión.

La evaluación de impacto ambiental permite comparar las situaciones ambientales existentes con aquellas que surgirían como resultado del desarrollo de una acción en particular. La comparación sirve para identificar tanto los impactos positivos y los beneficios ambientales que surgen de realizar el proyecto que se está evaluando, como aquellos de carácter negativo que deben manejarse para evitar la degradación del medio ambiente. Lo más significativo es que se incorporen las medidas que aseguren la protección del medio ambiente y que hagan viable la acción; si ello no es posible, la acción no debe ser ejecutada.

El análisis se expresa a través de un documento denominado estudio de impacto ambiental, donde se entregan los argumentos que permiten, por ejemplo, decidir la localización óptima de una acción cuando existen alternativas desde el punto de vista ambiental. En el caso que haya diversas opciones de ubicación, aporta los antecedentes para establecer cuál de ellas es la más adecuada para proteger el medio ambiente. Si se trata de una acción humana que no permite su reubicación, propone las formas de disminuir o evitar los impactos ambientales negativos [ESP02].

El estudio de impacto ambiental es el informe que documenta el proceso de evaluación de impacto ambiental y sus distintas etapas. Contiene el análisis, pronóstico y medidas que se toman para que una acción en particular sea compatible con la protección del medio ambiente.

2.3 Importancia de la Información para los Procesos de Gestión Ambiental

La descripción del ambiente afectado es determinante para conocer los impactos significativos que resultan del desarrollo de una acción en particular, al presentar en detalle las características y variables que mejor reflejan las alteraciones. La información que se utiliza debe: ser suficiente para calificar los cambios que ocurrirán como resultado de la acción humana; estar disponible en el nivel de detalle adecuado; y utilizar los métodos de análisis ajustados a cada realidad en particular. La información debe estar orientada a la obtención de antecedentes que permitan reconocer, calificar y vigilar el comportamiento de los impactos ambientales. Dado que la información es específica caso a caso, ella depende fuertemente del tipo de acción y de las características del ambiente involucrado [ESP02].

En éste contexto, los Sistemas de información Ambiental son elemento fundamental para cualquier Evaluación de Impacto Ambiental. Éstos están constituidos por un conjunto de antecedentes elaborados a partir de los datos proporcionados por diferentes actores, los que previamente han acordado procedimientos y métodos comunes para la generación, almacenaje, procesamiento y distribución de la información. Esto permite disponer de antecedentes integrados, que favorecen la toma de decisiones estratégicas y tácticas, científicamente centrados en cuanto a sus causas e impactos ambientales.

El sistema debe compartir objetivos comunes claramente identificados y contener secuencias establecidas de procedimientos y flujos de información. Generalmente, el propósito de todo sistema de información es facilitar la evaluación global de los principales problemas y asuntos ambientales, proveyendo los datos científicos necesarios para gestionar racionalmente los recursos naturales y la calidad del medio ambiente. La colecta de datos, en consecuencia, no es un fin en sí mismo sino un medio para observar y vigilar el medio ambiente y elaborar evaluaciones sobre su estado. El objetivo es proporcionar antecedentes que permitan su mejoramiento y adecuado manejo [ESP02].

La mayor dificultad surge al establecer qué constituye el dato y la información ambiental. Se puede definir, operacionalmente, que los datos de interés ambiental son aquellos que permiten conocer la estructura, estado y funcionamiento de los elementos naturales y artificiales que se articulan en el espacio y tiempo para conformar un sistema global a través de sus interacciones. La complejidad de conocer el estado y los procesos que están teniendo lugar obliga a seleccionar indicadores de estado, de cambio y de tendencias. Desde luego que una de las mayores dificultades se encuentra en la fijación de líneas de base; es decir, en la conceptualización para simular una condición que, en sí misma, puede manifestar en el futuro importantes niveles de deterioro o daño ambiental.

Entre los objetos y atributos de naturaleza física, biológica y sociocultural, cuyas interacciones contribuyen a caracterizar al medio ambiente, se pueden mencionar como ejemplos: el aire (tiempo meteorológico y clima, contaminación, etc), el agua (ciclo hidrológico, características físicas y biológicas, manejo y uso del agua, manejo de cuencas), los recursos naturales renovables y no renovables (los suelos, la flora y la fauna, los recursos minerales-energéticos y biológicos, sus niveles de contaminación y degradación), y los aspectos socioculturales (cantidad y características estructurales y dinámicas de las poblaciones humanas, actividades económicas y culturales, niveles de calidad de vida) [ESP02].

Entre los elementos que más comúnmente se analizan en los estudios de impacto ambiental, se encuentran:

- Agua
- Aire
- Suelos
- Flora y fauna
- Paisaje
- Aspectos sociales, culturales y económicos

El conocimiento previo que se tenga del área afectada influye mucho sobre el detalle de la descripción. Si el conocimiento es escaso, es recomendable realizar una descripción rápida, dando una idea de sus características gruesas para, posteriormente, determinar lo que requiere de mayor precisión. Si existe la información científica adecuada, ésta debe incluir sólo aquellos aspectos necesarios para entregar un conocimiento acabado de las posibles componentes ambientales que serán afectados por la acción en cuestión [ESP02].

En la mayoría de los casos, los datos que se acumulan se transforman en información ambiental en la medida que se analizan: a) las interacciones y su aporte integral a la condición global, b) las modificaciones que experimentan por acción humana, y c) los efectos que generan las perturbaciones causadas por la sociedad sobre las diversas manifestaciones de la vida.

En consecuencia, no sólo se trata de la caracterización de los componentes a través de sus atributos por separado, sino que de la proposición de modelos conceptuales que den cuenta de la interacción entre ellos. Una forma de analizar las interacciones que contribuyen a la visión holística y global del medio ambiente y de definir el tipo de información que se requiere, consiste en recurrir a los numerosos modelos conceptuales que existen a nivel de disciplinas científicas específicas.

De ésta manera, una de las condiciones importantes para el desarrollo de cualquier sistema de información es conocer la disponibilidad, calidad, origen y características de los datos y antecedentes relevantes para el cumplimiento de las funciones que le son asignadas. En síntesis se necesitan observaciones integradas de los recursos y de las interacciones entre los diferentes componentes naturales, artificiales, socioeconómicos y culturales de los ecosistemas seleccionados. Lo expuesto requiere la creación de un sistema integrado de información ambiental en el que se pueda disponer de la información de la línea de base generada en el proyecto de evaluación de impacto ambiental, estableciendo una relación entre la instrumentalización, el diseño de las bases de datos concebidas, el levantamiento de la información en campo, y la tipificación de indicadores ambientales [GIT02].

3. SISTEMAS DE INFORMACIÓN AMBIENTAL Y TECNOLOGÍAS QUE FACILITAN LA CONSTRUCCIÓN DE UN ENTORNO DE INTEGRACIÓN DE MÚLTIPLES FUENTES DE DATOS

En este capítulo se elabora un estado del arte de los Sistemas de Información Ambiental y las tecnologías que facilitan la construcción de un entorno de integración de múltiples fuentes de datos con sistemas de análisis espacial sobre Internet: Servicios Web, Esquema Global para Múltiples Bases de Datos, Lenguajes para Múltiples Bases de Datos, Bases de Datos Federadas y Sistemas Basados en Wrappers y Mediadores.

3.1 Sistemas de Información para la Gestión Ambiental

Un Sistema de Información Ambiental (SIA) es una herramienta computarizada para la homogeneización, la actualización, el almacenamiento, el análisis y la divulgación de toda la información pertinente para la gestión, el control y el monitoreo ambiental. Es el banco de datos más adecuado para apoyar la toma de decisiones en un desarrollo económico sostenible. La información está compuesta de datos con distribución espacial (datos geográficos), y de datos generales, sin ubicación o extensión geográfica. También hay datos numéricos (computarizados) y analógicos (documentos en "papel"). Además, hay datos descriptivos y cuantitativos. El SIA debe tener la capacidad de integrar toda esa variedad de datos más los recursos científicos y técnicos para proveer un análisis de calidad y facilitar la divulgación de los resultados [GIT02].

Teniendo en cuenta esta definición de Sistema de Información Ambiental, los principales componentes que lo soportan son:

- Sistema de Información Geográfica (SIG) para el análisis espacial.
- Programas de gestión de bancos de datos para el análisis.
- Equipamiento para la grabación y digitalización de datos analógicos.
- Equipamiento para la producción de reportes y mapas.
- Sistema de monitoreo de variables ambientales.
- Acceso a recursos profesionales para el análisis de temas ambientales.
- Personal para la operación y la gestión del sistema.
- Medios de comunicación para la divulgación y socialización de resultados.
- Lazos inter-institucionales en los ámbitos regional, nacional e internacional.
- Instrumentos para compartir e intercambiar datos e información.

A continuación se hará referencia a los componentes que más influencia tienen al momento de desarrollar un Sistema de Información Ambiental: los Sistemas de Información Geográfica, los Programas de Gestión de Bancos de Datos, y los Instrumentos para compartir e intercambiar datos e información.

3.1.1 Sistemas de Información Geográfica

Un Sistema de Información Geográfica (SIG) es un sistema de información con capacidades específicas para la captura, manipulación, recuperación, análisis, representación, etc., de información georreferenciada. Se dice que una información está referenciada geográficamente cuando de manera inherente considera la posición espacial ocupada por los objetos del mundo real que modela (e.g. ríos, montañas, cuencas) [MFW95].

Uno de los aspectos fundamentales de los SIG es la información que almacenan. Sus entidades contienen los datos asociados a los objetos del mundo real (e.g. nombre, temperatura), su localización geográfica (coordenadas), y además su geometría, pues entre las funcionalidades más importantes de los SIG se encuentra la representación gráfica de dichos objetos. Por consiguiente, las bases de datos de los SIG contienen datos gráficos y alfanuméricos; cada uno de ellos tiene características específicas y diferentes requisitos para su eficaz almacenamiento, proceso y representación; además deben estar integrados para efectos de formar una única fuente de información.

Otra característica específica de los SIG, que los diferencia de los demás sistemas de información, es su capacidad para el procesamiento de datos espaciales, que incluye operaciones métricas (e.g. obtener la distancia entre dos ciudades), geométricas (e.g. obtener la pendiente de un terreno) y topológicas (e.g. establecer si un lago está dentro de cierta cuenca).

En el siguiente capítulo de esta monografía se profundizará en el tema de los SIG y se expondrá claramente los tipos de datos manejados por estos sistemas.

3.1.2 Sistemas de Gestión de Bases de Datos

Las bases de datos que dan soporte al análisis de los procesos de gestión ambiental contienen una enorme diversidad de información. Las variables comúnmente utilizadas en estos procesos pueden agruparse en dos grandes conjuntos: las variables temáticas (geológicas, hidrológicas, atmosféricas, biológicas) y las variables auxiliares (topográficas, socioeconómicas y de teledetección). No todas las características de estas variables son simplemente datos puntuales; por ejemplo, los datos que componen los módulos de gestión tanto geológico e hídrico, como mapas y sondeos, son datos que necesitan investigaciones complejas y por lo general están constituidos por otra infinidad de datos. Esta diversidad de datos se puede modelar como capas de información, que una sobre otra generan un nuevo dato o información procesada.

En el siguiente capítulo de esta monografía se abordará con mayor detalle los aspectos más relevantes de los Sistemas de Gestión de Bases de Datos.

3.1.3 Intercambio de Datos e Información

El intercambio de datos e información se ha convertido en un nudo Gordiano para los Sistemas de Gestión Ambiental, ya que depende de convenios inter-institucionales, que muchas veces no se logran concretar. Un Sistema Integrado de Información Ambiental

como el propuesto en el presente proyecto, debe contar, en primer término, con la iniciativa de las principales instituciones promotoras de la gestión ambiental en la región. El paso siguiente es la definición de un modelo común de gestión que se atempere a la definición dada para los SIA, y de una arquitectura que facilite el acceso a las diferentes fuentes de información por parte de las instituciones involucradas en el proceso [GIT02].

En la actualidad existen iniciativas a nivel nacional, que propenden por la generación de un Sistema Integrado de Información Ambiental, que permita intercambiar información entre diferentes entidades. Estas iniciativas son:

- Sistema de Información Nacional Ambiental (SINA). Este sistema, según el artículo 4 de la ley 99 de 1993, es el conjunto de orientaciones, normas, actividades, recursos, programas e instituciones que permiten la puesta en marcha de los principios generales ambientales contenidos en ella.
- ICDE (Infraestructura Colombiana de Datos Espaciales). Es la suma de políticas, estándares, organizaciones y recursos tecnológicos que facilitan la producción, el acceso y el uso de la información geográfica de cubrimiento nacional, para apoyar el desarrollo económico y social del país.

Aunque estas propuestas son robustas y ambiciosas, no han creado las condiciones conceptuales y tecnológicas para hacer operativo un sistema de articulación regional, que permita la construcción de bases de datos compartidas sobre las regiones, sus recursos y potenciales amenazas. La construcción de un sistema de esta naturaleza no parece ser una prioridad en los planes de desarrollo de los organismos responsables de estas acciones, pues continúan actuando de forma reactiva, más por la presión social que por la existencia de una política coherente y sistémica frente al problema; su accionar está definido en función de procesos de mitigación y no de prevención, visión que genera grandes costos para el desarrollo social, económico y ambiental de Colombia.

En esta misma dirección, se vienen adelantando iniciativas a nivel internacional como el Sistema de Información de la Diversidad Biológica y Ambiental de la Amazonía Peruana [MMP01], y el Sistema de Información Ambiental del Ministerio de Medio Ambiente de la República de Nicaragua [MMN01]. Estos sistemas centran sus esfuerzos en establecer la organización de los datos ambientales con el objetivo de generar planes de ordenamiento de los recursos naturales, mas tampoco se plantean la automatización del intercambio de información entre las instituciones.

Desde el punto de vista tecnológico, los esfuerzos a nivel internacional confluyen alrededor de la creación de componentes que brinden facilidades para el intercambio de datos espaciales, como en el sistema desarrollado en la Escuela de Ingeniería en Sistemas Computacionales de la Universidad de las Américas en Puebla (México) [CEA00], que tiene los siguientes objetivos:

- Realizar las consultas espaciales remotas que se requieran por múltiples usuarios en forma sencilla y segura.
- Diseñar y desarrollar una aplicación basada en el web para servir datos geográficos a través de Internet.
- Desarrollar un exportador de datos del formato propietario de ArcView (archivo shape) al estándar OpenGIS, para su posterior publicación en el web y su almacenamiento en la base de datos.

Estos trabajos cuentan con un soporte fuerte en las tecnologías de componentes distribuidos pero dejan de lado la gestión de la información ambiental, centrándose directamente sobre la arquitectura a implementar.

Es de observar que los desarrollos internacionales tratados en los párrafos anteriores dejan además a un lado el monitoreo de variables ambientales en tiempo real, elemento fundamental para la gestión de ecosistemas estratégicos protegidos.

En el campo comercial, existen productos que permiten la distribución de información geográfica vía Internet y la integración en tiempo real de datos procedentes de cualquier parte del mundo; por ejemplo, ArcIMS y MapInfo-MapServer son soluciones que proporcionan una plataforma común para este intercambio. ArcIMS permite el acceso a recursos en el World Wide Web y ofrece las siguientes características [ESR00]:

- Sencilla instalación, implementación y administración mediante asistentes y plantillas.
- Capacidad para servir imágenes y vectores.
- Integra datos locales con datos en Internet.
- Administración y mantenimiento sencillos de los servicios publicados.
- Disponible para Windows NT y Unix.
- Arquitectura del servidor altamente escalable.
- Representación cartográfica de alta calidad.

Aunque ofrecen una solución integral, estos sistemas presentan el gran inconveniente de establecer formatos propietarios para el intercambio de información, imposibilitando el uso de componentes de otros Sistemas de Información Geográfica disponibles en el mercado, y limitándose a establecer una arquitectura sólida para los productos de su casa matriz. De manera adicional, los costos en el mercado de estos productos son bastante elevados, dificultando su adquisición, especialmente en proyectos de bajo presupuesto. Finalmente, es complicado construir componentes adicionales con nuevas funcionalidades para estas arquitecturas.

Basado en el escenario descrito en los apartados anteriores, en la siguiente sección se describen los conceptos que dan soporte a un entorno de interoperabilidad de múltiples fuentes de datos que hacen uso de sistemas de análisis espacial sobre Internet.

3.2 Interoperabilidad de Sistemas de Información

Esta sección tiene como objetivo principal presentar los problemas críticos que aparecen en la interoperación de sistemas de información, y estudiar distintas propuestas de solución a los mismos. En primera instancia se presentarán los conceptos fundamentales, antes de abordar las arquitecturas de interoperabilidad.

3.2.1 Conceptos Fundamentales

Al comenzar a trabajar en esta temática es importante que tengamos en claro las diferencias entre términos similares que a veces utilizamos indistintamente cuando hablamos de Interoperabilidad.

Tomemos primero el término **Interconexión** de sistemas; éste hace referencia a la conexión a nivel de hardware de distintos sistemas entre sí. El objetivo de una interconexión es intercambiar información, la cual es alcanzada con éxito, si al menos se dispone de un protocolo común de comunicación. Por su parte **Integración**, hace referencia no sólo a intercambiar información sino también a unificar la información que se encuentre replicada, resolviendo posibles conflictos de heterogeneidad entre las diferentes fuentes de origen de la información [MOT02].

Finalmente, el término **Interoperación** se refiere no sólo a integrar información sino también a integrar funcionalidades. Hablamos entonces en forma general de **sistemas interoperables** cuando nos referimos a sistemas que cooperan intercambiando datos y sincronizando las ejecuciones de sus aplicaciones [MOT02].

Según [SSU91] distintos sistemas de información son interoperables cuando poseen la habilidad de requerir y recibir servicios, y operar como un único sistema en la resolución de un problema común.

Las características principales de los sistemas interoperables son:

- Existe distribución en un ambiente de Cliente/Servidor
- Existe comunicación a pesar de la heterogeneidad
- Cada uno de los sistemas usa las facilidades del otro

De esta manera, para lograr interoperación de sistemas, es necesario tener a los sistemas de alguna forma interconectados e integrados, además, debe existir alguna relación semántica entre ellos, es decir que algunos de los datos que ellos manejan sean correspondientes entre sí a nivel semántico. La semántica de los datos es "el mapeo entre el objeto modelado, representado y/o almacenado en un sistema de información y el objeto del mundo real que él representa" [ERS99]. Según esta definición la semántica de los datos se interpreta como la conexión entre la base de datos y el mundo real fuera de la base de datos. Desde una perspectiva intuitiva, es la correspondencia que podemos establecer entre el objeto del Mundo Real y la representación que de él estamos usando en la base de datos. Hablar entonces de dos sistemas que interactúan en algún nivel semántico es hablar de dos sistemas donde parte de la realidad que representan es la misma. De lo contrario, los conflictos semánticos ocurren cuando los dos sistemas no tienen la misma interpretación de la información. El uso de ontologías para la clarificación del conocimiento implícito en cada sistema es una alternativa que puede llegar a solucionar los problemas semánticos de heterogeneidad entre los sistemas.

Otro de los términos a considerar al trabajar la temática de interoperabilidad es la **distribución**, la cual hace referencia a la distribución de datos. Los datos distribuidos son un conjunto de datos, lógicamente relacionados y residentes en varias computadoras conectadas por una red de comunicación. La red de comunicación permite tener acceso centralizado, en el sentido de acceso lógico único. El acceso a datos distribuidos es resuelto a través de técnicas como HTTP, CORBA y Servicios Web [W3C02],[BSD03]

(Éstas técnicas se detallarán más adelante). De ésta manera, se pueden desarrollar aplicaciones que acceden a datos ignorando la ubicación física de los componentes que los proveen.

El término Autonomía se refiere a la capacidad de ser independiente. Los sistemas de información pueden cooperar entre sí teniendo autonomía. Para hablar de autonomía vamos a distinguir entre distintos tipos [ERS99].

Autonomía de Diseño. Ésta autonomía hace referencia a la libertad en la elección del sistema de su propio: Modelo de Datos, Lenguaje de consulta, Implementación, Restricciones, Sistema de Gestión de datos a ser usado y Funcionalidades soportadas.

Autonomía de Comunicación. La autonomía de comunicación se refiere a la capacidad de los componentes de decidir cuándo y cómo responder a peticiones de otros componentes.

Autonomía de Ejecución. La autonomía de ejecución de cada componente se refiere a la libertad que tiene cada componente de ejecutar sus propias operaciones sin interferir en la ejecución de operaciones no locales.

Autonomía de Participación. La autonomía de participación se refiere a la capacidad de los componentes de decidir cuánto de sus funciones, operaciones y datos compartir. También incluye la posibilidad de decidir entrar o salir de la cooperación cuando lo deseen.

Para preservar la autonomía de diseño y la de participación el sistema global no debe imponer ninguna restricción sobre los sistemas de bases de datos locales ni tampoco requerir ninguna modificación sobre ellos.

Para preservar la autonomía de comunicación y la de ejecución el sistema global no debe ejercer ningún control sobre las ejecuciones de transacciones locales (a excepción del envío de subtransacciones) [MOT02].

La preservación de la autonomía de ejecución debe asegurar que las operaciones locales (comandos o transacciones enviadas directamente por un usuario local del sistema manejador de bases de datos) son ejecutadas sin interferencia de las operaciones externas (operaciones enviadas por otro componente o por el sistema global) mientras que es el propio participante quien decide el orden en el cual ejecuta las operaciones externas. Por lo tanto, un sistema externo no puede forzar una orden de ejecución de los comandos sobre un sistema de gestión de bases de datos con autonomía de ejecución. Tampoco necesita el sistema participante informar al sistema global el orden en el cual las operaciones del sistema global son por él ejecutadas. Operacionalmente, un sistema participante ejercita su autonomía de ejecución tratando las operaciones externas en la misma forma que operaciones locales. Este último aspecto de autonomía puede, sin embargo, ser solucionado haciendo que el componente informe al sistema global del orden de ejecución de la transacción (o grafo de espera de la transacción) para permitir una administración más simple y más eficiente de las transacciones globales [MOT02].

Según [BE+96] los distintos tipos de autonomía también se pueden agrupar de acuerdo a aspectos estáticos o dinámicos. Autonomía de diseño y de participación son aspectos estáticos de la autonomía local, estas características son fijadas antes (autonomía de

diseño) o en el momento (autonomía de participación) de la integración con otros sistemas. En contraste, la autonomía de comunicación y de ejecución son aspectos dinámicos de la autonomía local; estas características definen el derecho de los sistemas participantes de tomar decisiones durante el proceso de integración.

Además de los términos de Semántica, Distribución y Autonomía, el término Heterogeneidad cobra vital importancia. Su significado se extrae de la definición de Homogeneidad. Bajo éste contexto, un sistema de información es homogéneo si el software que crea y manipula los datos es el mismo, y si la representación de los datos sigue el mismo modelo, y es parte del mismo universo de discurso. Si un sistema no cumple alguno de estos requerimientos se habla de un sistema heterogéneo.

Los distintos casos de heterogeneidades se clasifican según distintos autores. Por ejemplo, [HML93] clasifican las heterogeneidades en las siguientes clases: heterogeneidades de lenguaje de metadatos, de la terminología utilizada, de especificación de metadata, de comparación de objetos, de formato de datos, y de sistemas de gestión de bases de datos. Algunos autores colocan dentro de las heterogeneidades semánticas las diferencias de nombres de atributos, de tipos de datos, de diferentes números de atributos y de diferentes nombres de entidades. En [KSH96] se describe una medida de distancia entre entidades de diferentes esquemas llamada de proximidad semántica, y se aplica esta medida al análisis de las relaciones entre heterogeneidades semánticas y estructurales. En esta sección presentamos una clasificación de heterogeneidades basada en la clasificación de [KSH96]. Agrupamos las heterogeneidades en cuatro clases: Heterogeneidad de sistemas, sintáctica, semántica y estructural. A continuación definimos cada una de estas clases de heterogeneidades:

Heterogeneidad de sistemas: Corresponde al uso de diferentes plataformas de hardware y software, el uso de diferentes sistemas operativos, así como también diferentes protocolos de comunicación.

Heterogeneidad sintáctica: Corresponde al uso de diferentes modelos de datos (relacionales, jerárquicos, orientados a objetos, etc.). También se habla de heterogeneidad sintáctica o descriptiva refiriéndose a los nombres de los elementos con los cuales se representa la realidad, por ejemplo la existencia de sinónimos u homónimos.

Heterogeneidad semántica: Corresponde a diferencias en el significado, interpretación, o el uso deseado para los mismos datos o para datos relacionados. Por ejemplo, ocurre una heterogeneidad semántica entre atributos de dos esquemas que representan el costo de una comida en un restaurante, cuando en un esquema el concepto "costo comida" incluye el costo de propina, impuestos y comida, mientras que en el otro esquema con el mismo concepto "costo comida" no se incluye ni la propina ni los impuestos.

Heterogeneidad estructural: Corresponde a diferentes representaciones para la misma realidad.

3.2.2 Arquitecturas de Interoperabilidad

Debido a que una de las características inevitables al querer realizar interoperabilidad es la presencia de DISTRIBUCION de datos, esta sección se iniciará revisando el concepto de sistema de gestión de bases de datos distribuidas.

Una base de datos distribuida consiste en un conjunto de datos lógicamente relacionados residentes en varias computadoras conectadas por una red de comunicación entre las que existe una aplicación global. Cada sitio de la red tiene procesos autónomos y pueden ejecutar aplicaciones locales. Cada sitio también participa en la ejecución de por lo menos una aplicación global, la que requiere acceder a los datos de varios sitios usando un subsistema de comunicación [CPE84].

Esta definición enfatiza los siguientes aspectos importantes en un Sistema de Bases de Datos Distribuidas:

- **Distribución:** el hecho de que los datos no están residentes en el mismo sitio.
- **Lógicamente relacionados:** el hecho de que los datos tienen ciertas propiedades que los vinculan, así podemos distinguir Bases de Datos Distribuidas de un conjunto de bases de datos locales o archivos los cuáles residen en diferentes sitios o en una red de computadoras.
- La existencia de un único sistema de gestión de bases de datos el cual se encuentra distribuido entre los diferentes sitios.

Ahora bien, tomando como punto de partida los conceptos de distribución y la relación lógica de datos, a continuación se presentarán las arquitecturas propuestas para lograr interoperabilidad entre sistemas de gestión de bases de datos.

3.2.2.1 Esquema Global para Múltiples Fuentes de Datos

Por esquema global se entiende la integración completa de varias bases de datos heterogéneas para proveer una vista única [SSP94]. Su principal ventaja radica en que los usuarios tienen una única vista consistente y acceso uniforme a los datos. Además, la distribución y la heterogeneidad de los sistemas participantes son transparentes a los usuarios. La siguiente figura ilustra el hecho de que en esta arquitectura las diversas bases de datos que participan son lógicamente vistas por los usuarios del sistema global como una única base de datos.

Sin embargo, la construcción de un esquema global posee serias desventajas. La resolución de las heterogeneidades es realizada en forma manual por un “super usuario” capaz de tener el conocimiento de cada una de las bases de datos participantes, lo cual hace al sistema muy propenso a errores y muy costoso en horas-hombre dedicadas. Debido además a la dependencia de la comprensión del usuario y de la semántica en el contexto de cada sistema participante, resulta muy difícil probar la correctitud del esquema global, o sea, probar si representa completamente a toda la información de todos los esquemas participantes.

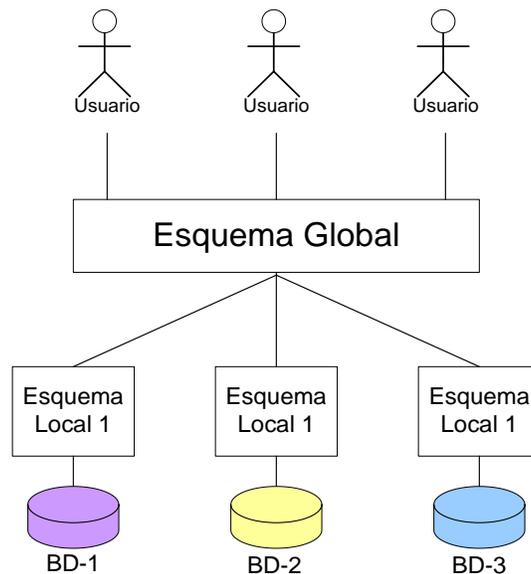


Figura 3.1 - Arquitectura del Esquema Global para Múltiples Fuentes de Datos

Otra desventaja, es que los sistemas participantes deben sacrificar su autonomía, especialmente la autonomía de participación, pues para lograr una visión lógica de una única base de datos el esquema global debe disponer de toda la información de los participantes a priori. Respecto a la evolución de los sistemas participantes, esta arquitectura no resulta adecuada ya que ante cambios en las bases de datos participantes todo el proceso de construcción del esquema global debe ser realizado nuevamente.

Una alternativa a la desventaja del costo de mantenimiento de estos sistemas ante la evolución y/o la escalabilidad de los mismos es la arquitectura de Lenguajes para Múltiples Bases de Datos [MOT02].

3.2.2.2 Lenguajes para Múltiples Bases de Datos

El objetivo de los lenguajes para múltiples bases de datos es proveer construcciones que permitan realizar consultas sobre más de una base de datos al mismo tiempo [REC+89] [LIT94].

Estos lenguajes tienen características que no poseen los lenguajes tradicionales de consulta, por ejemplo, un nombre global puede ser usado para identificar a un conjunto de bases de datos y las consultas pueden realizarse sobre datos que residan en cualquiera de las bases de datos participantes como lo ilustra la siguiente figura.

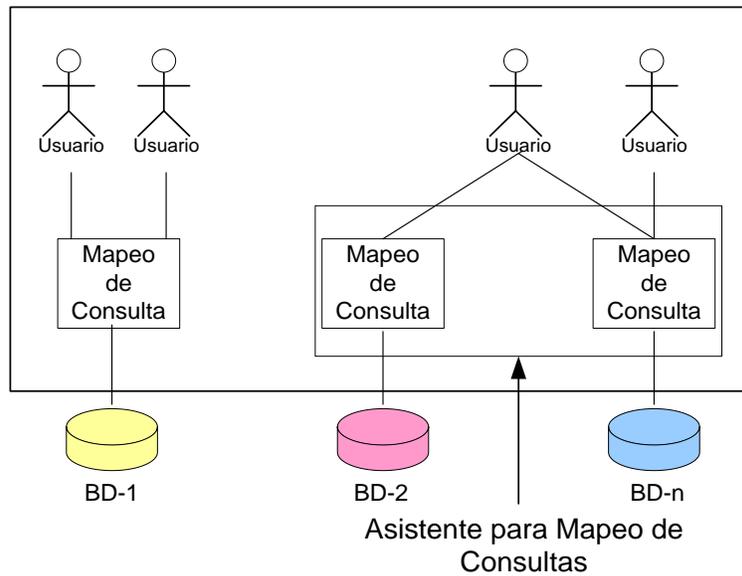


Figura 3.2 – Arquitectura para los Lenguajes Múltiples de Bases de Datos

Algunos ejemplos de propuestas de lenguajes para múltiples bases de datos son SchemaSQL [LSS96], MQL [WWG01] y nD-SQL [GLV98].

En [WWG01] se define el lenguaje declarativo Meta-Query Language como extensión de SQL, el cual posibilita reestructurar dinámicamente las consultas. Un elemento clave de la sintaxis de MQL es derivado de SchemaSQL, y es la posibilidad de incluir meta-variables en la cláusula FROM, para cada una de las cuales se recorre un rango de valores.

Debido a que en las arquitecturas de lenguajes para múltiples bases de datos es necesario conocer la estructura de los participantes para establecer la correspondencia de las consultas, el grado de transparencia en la distribución de los datos se ve muy reducido. Asimismo requieren un alto grado de responsabilidad del usuario para lograr sus objetivos con éxito.

3.2.2.3 Bases de Datos Federadas

Bajo la perspectiva de las bases de datos que componen un sistema, tenemos los sistemas Multi Bases de Datos (SMulBD), los cuales soportan sus operaciones en múltiples sistemas de bases de datos participantes (SBDP).

Desde el punto de vista de la heterogeneidad, un SMulBD es llamado homogéneo si todos los SGBD participantes son iguales, de lo contrario será llamado SMulBD heterogéneo. Igualmente, si consideramos la autonomía de cada SBDP, un SMulBD puede ser clasificado en dos tipos: Sistema de base de datos no-federado y Sistema de bases de datos federado.

Un sistema de bases de datos no federado es una integración de SGBDs participantes que no son autónomos. Esto significa que los SBDP al participar en una federación pierden su autonomía y cualquier operación debe hacerse sobre la base de datos global

[RMM99]. Un Sistema de este tipo no distingue entre usuarios locales y usuarios no locales.

Por el contrario, las bases de datos federadas son una colección de SBDP cooperativos, heterogéneos, que son autónomos y que permiten compartir todos o algunos de sus datos [SLA90].

La siguiente figura muestra un esquema de una federación donde se observa que los sistemas participantes pueden ser bases de datos centralizadas, distribuidas o a su vez otra federación [FRV95].

El recuadro pintado de amarillo de la siguiente figura señala la existencia de un esquema global, unificado, al cual acceden todos los usuarios de la federación. Este esquema global puede ser materializado (cuando se puebla con nuevas instancias extraídas de las bases de datos participantes) o puede ser virtual (cuando no se generan nuevas instancias sino que se mantienen correspondencias con las instancias de las bases de datos participantes).

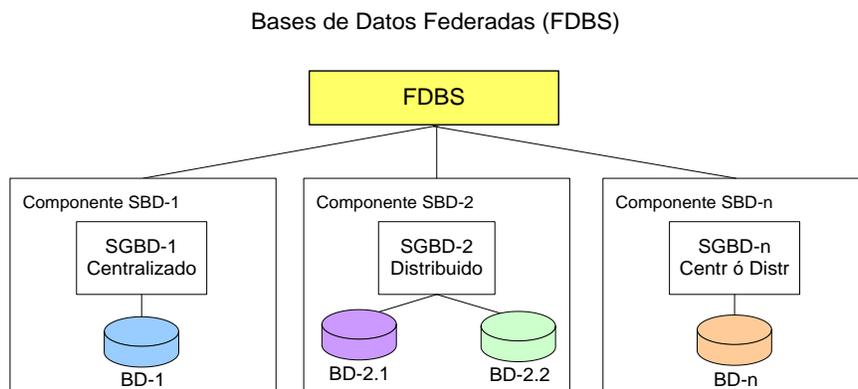


Figura 3.3 – Esquema de Federación

La participación de los sistemas fuentes en el esquema global puede ser en forma completa o parcial, indicando que el participante colabora en la federación con todo su esquema en el primer caso, o con cierto subesquema de su sistema en el último caso. Esto marca una fuerte diferencia con la arquitectura de esquema global donde siempre los participantes deben colaborar con la totalidad de su esquema. Además, si se realiza una comparación mas detallada de las bases de datos federadas respecto al esquema global, los sistemas de bases de datos participantes preservan su autonomía. De ésta manera, los sistemas participantes pueden continuar con sus operaciones locales mientras que al mismo tiempo participan de la federación, o sea participan en la ejecución de una operación global [MOT02].

Con base en la característica de la autonomía y de la forma en la cual se identifican los subesquemas equivalentes, o complementarios, entre los esquemas participantes se realiza la siguiente clasificación de sistemas federados:

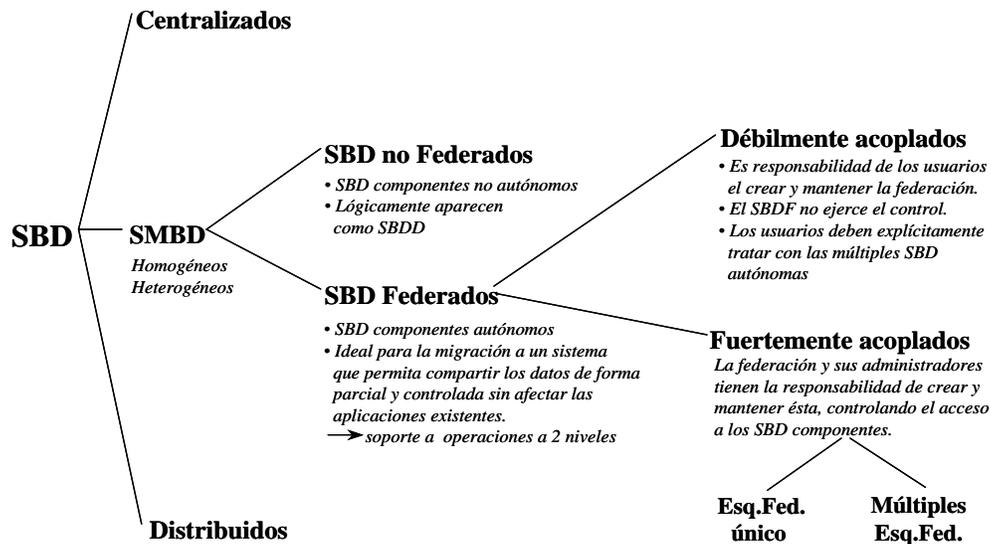


Figura 3.4 – Clasificación de Esquemas Federados

Los sistemas federados se clasifican según la forma de acoplamiento entre sus participantes en: fuertemente acoplados y débilmente acoplados. La forma de acoplamiento se refiere al procedimiento por el cual se identifican los subesquemas equivalentes, o complementarios, entre los esquemas participantes.

En los sistemas fuertemente acoplados existe un usuario experto de la federación, comúnmente el administrador de la base de datos federada, quien decide las correspondencias entre los sistemas participantes. Para que este usuario identifique estas correspondencias los sistemas participantes deben permitirle ver sus esquemas, acceder a sus descripciones, pero esto viola la autonomía de los participantes. Sin embargo la ventaja de este enfoque es que existe una visión centralizadora del problema que se encarga de asegurar la consistencia del sistema federado.

Por el contrario, en federaciones débilmente acopladas, cada uno de los sistemas participantes es el encargado de negociar con los demás participantes para identificar sus similitudes. Esta negociación par-a-par tiene la ventaja de no necesitar que el administrador de la federación vea sus esquemas completos, pero presenta la desventaja de la terrible duplicación de esfuerzos para negociar con cada uno de los participantes. Otro posible inconveniente es que requieren que sus participantes posean algún lenguaje común de acceso a datos para poder identificar las partes de los esquemas que tienen en común. Esta identificación se realiza generalmente a través de la definición de vistas, esto hace que para la actualización de datos posea los mismos problemas que en actualización de vistas [GMU95]. Como ventajas para los débilmente acoplados podemos citar el alto grado de autonomía y la mayor facilidad para manejar la escalabilidad del sistema federado.

3.2.2.4 Sistemas basados en Wrappers y Mediadores

Veamos ahora una variante de la arquitectura de sistemas federados, la arquitectura de Wrappers y Mediadores.

La palabra *wrapper* significa en inglés envoltorio sin embargo debido al extenso uso del término en inglés decidimos no traducirla. El término mediador fue introducido por Wiederhold en [WIE92]. Actualmente se le asocia con la “resolución de los problemas de heterogeneidad para lograr integración semántica de datos” [WIE94b].

Un sistema basado en *wrappers* y mediadores está constituido por diferentes sistemas participantes, donde puede existir sistemas no-bases de datos y donde un mediador puede ser a su vez un participante del sistema.

En ésta aproximación arquitectural, los datos permanecen en sus fuentes originales y un sistema intermedio, llamado mediador, se encarga de proporcionar a los usuarios la ilusión de que existe una única fuente en la que se encuentran todos los datos combinados y unificados de manera coherente de acuerdo a un único esquema global. Cuando el mediador recibe una consulta sobre el esquema global, ésta se reformula en diversas subconsultas que se realizan directamente sobre las fuentes originales. La interacción directa con las fuentes es delegada por el mediador a los llamados programas wrapper (o wrappers), cada uno de ellos especializado en el diálogo concreto con cada servidor. Estos se encargan de recibir las peticiones del mediador, traducirlas como subconsultas ajustadas al formato particular de cada fuente, ejecutarlas sobre la misma y obtener los resultados, retornándoselos al mediador. Es entonces cuando los resultados de las fuentes son reestructurados por el mediador para ajustarse al esquema global y ser devueltos entonces al usuario. De esta manera, éste obtiene la impresión de estar consultando un Sistema único. Este esquema se encuentra representado en la Figura 3.5.

A continuación se describen detalladamente los distintos módulos generales (Figura 3.6) que componen la arquitectura de un sistema mediador [RAP02].

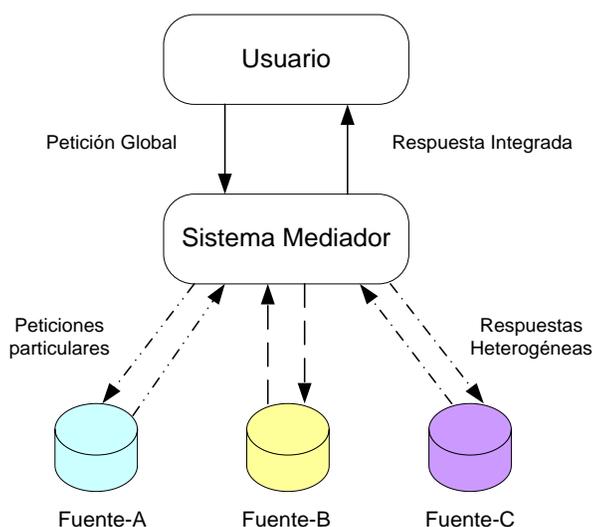


Figura 3.5 – Arquitectura del Sistema Mediador

Meta Datos: Almacena la definición de todas las estructuras de metadatos: esquemas globales, tipos, operadores, descriptores de las características de las fuentes (wrappers). Este módulo es importante para la Base de Datos Virtual, ya que los datos están almacenados en lugares remotos, y los únicos datos de la Base de Datos Virtual están almacenados en los Meta Datos.

Módulo de Administración: Facilita la manipulación de los metadatos. Las tareas fundamentales de este módulo son gestionar los procesos en las fuentes de datos y configurar los *wrappers*.

Módulo Intérprete de Consultas: Traduce la consulta de nivel de usuario a la representación interna de la Base de Datos Virtual.

Módulo de Cómputo de Capacidades del Esquema Mediador: Encargado de la generación de nuevas vistas (vistas globales de la Base de Datos Virtual) en función de la combinación de las capacidades de las fuentes. Las capacidades de las vistas fuente se generan a partir de la especificación de las restricciones de búsqueda que las fuentes encapsulan; en otras palabras, las capacidades de las vistas fuente son generadas por los *wrappers*. Estas capacidades son metadatos indispensables para el módulo de Ejecución de Consultas de la Base de Datos Global. En éste módulo de cómputo, cabe resaltar los procesos llevados a cabo por el módulo minimizador de métodos de búsqueda de las vistas, ya que existen un sinnúmero de métodos implementados por cada fuente.

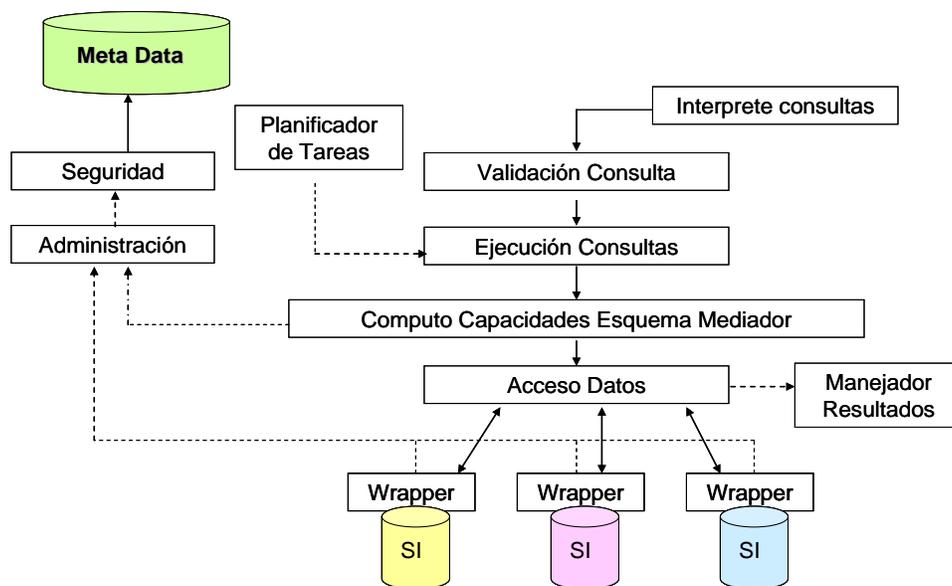


Figura 3.6 – Componentes del Sistema Mediador

Módulo de Validación de Consulta: Indica si una consulta puede o no ser resuelta por la Base de Datos Virtual.

Módulo de Ejecución de Consulta: Encargado de generar los planes de consulta para las consultas admitidas por el módulo anterior. En éste módulo se resaltan los siguientes sub-módulos:

- Planes de ejecución: Los cuales se forman mediante la unión de los distintos métodos de búsqueda que se activan para la consulta en cada una de las fuentes de datos.
- Optimizador de planes: su función es tener en cuenta la implementación de cada uno de los métodos de búsqueda activados, tal como su complejidad y capacidades de la fuente.

Módulo de Acceso a Datos: Encargado de la ejecución de los *Wrappers* de cada una de las fuentes de datos.

Módulo de Seguridad: Encargado de permitir o denegar el acceso a usuarios a determinadas operaciones sobre la Base de Datos Virtual.

Planificador de Tareas: Se puede considerar como un módulo independiente que opera “supervisando” a intervalos de tiempo la ejecución de las consultas sobre la Base de Datos Virtual.

Manejador de Resultados: Realiza acciones sobre los resultados obtenidos (enviar correo, etc.)

En conclusión, cada relación del sistema mediador se define mediante una consulta sobre las relaciones base, de forma similar a la definición de las vistas en una base de datos relacional. Las consultas se expresan en un lenguaje próximo a SQL en el que se escriben las consultas en fase de ejecución. Después de definir las relaciones globales el mediador debe ser capaz de propagar las consultas a través del árbol de vistas, obteniendo automáticamente las capacidades permitidas por las relaciones del esquema global en forma de una serie de métodos de búsqueda

3.3 Mecanismos de Acceso a Datos Distribuidos

Retomando lo expuesto en la sección 3.2.1 sobre conceptos fundamentales de Interoperabilidad de Sistemas de Información, en lo que hace referencia a las técnicas para el acceso a datos distribuidos, a continuación se expondrán los conceptos concernientes a Servicios Web.

Los Servicios Web son una tecnología emergente impulsada por el deseo de exponer de forma segura la lógica de negocios en Internet. A través de los Servicios Web las organizaciones pueden encapsular sus procesos de negocios existentes, publicarlos como servicios, suscribirse a otros servicios e intercambiar información entre organizaciones.

Un Servicio Web es visto como una noción abstracta que debe ser implementada por un agente concreto (Ver Figura 3.7). El agente es la entidad física (una pieza de software) que envía y recibe mensajes, mientras que el servicio es el conjunto abstracto de funcionalidades. Para ilustrar esta situación, se podría implementar un Servicio Web particular un día usando un agente (escrito en un lenguaje de programación x), y un agente diferente el día siguiente (escrito en un lenguaje de programación y). Al tiempo que el agente puede cambiar, el Servicio Web permanece igual [WSA03].

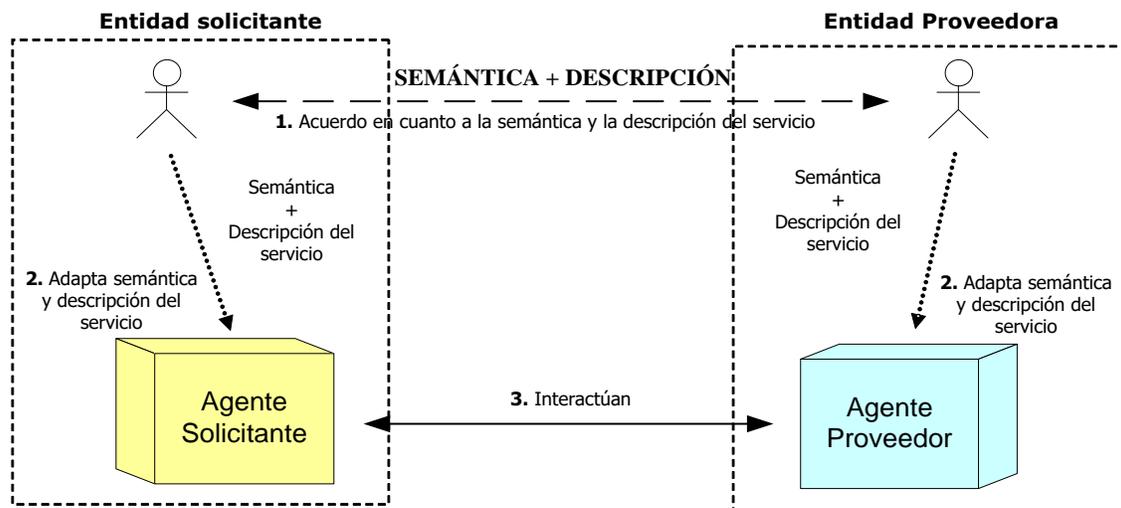


Figura 3.7 – Arquitectura de los Servicios Web

De la misma manera, el propósito de un Servicio Web es proveer alguna funcionalidad en representación de su dueño – una entidad legal, tal como un negocio o un individuo. La entidad proveedora es la entidad legal que provee un agente apropiado para implementar un servicio particular (ver Figura 3.7: Roles básicos de la arquitectura).

La entidad solicitante es una entidad legal que desea hacer uso del servicio Web de la entidad proveedora. Esta usará un agente solicitante para intercambiar mensajes con el agente proveedor de la entidad proveedora. Con el fin de que haya un intercambio de mensajes exitosos, la entidad solicitante y la entidad proveedora deben ponerse primero de acuerdo en la semántica y los mecanismos del intercambio de mensajes.

El mecanismo de intercambio de mensajes está (parcialmente) documentado en la descripción del servicio Web (Ver Figura 3.7). La descripción del servicio es una representación de los formatos de mensajes, tipos de datos y protocolos que deberían ser usados entre el agente solicitante y el agente proveedor. Ésta también especifica la ubicación de red del agente proveedor, y podría proveer alguna información sobre el patrón de intercambio de mensajes esperado.

La semántica del intercambio de mensajes representa el “contrato” entre la entidad solicitante y la entidad proveedora relativo al propósito y las consecuencias de la interacción. También incluye algunos detalles adicionales sobre los mecanismos de intercambio de mensajes que no son especificados por la descripción del servicio. Aunque este contrato representa el contrato completo entre la entidad solicitante y la entidad proveedora sobre cómo y por qué sus respectivos agentes interactuarán, esto no está necesariamente escrito o explícitamente negociado. Podría ser explícito o implícito, oral o negociado, procesable por la máquina u orientado a humanos.

Mientras que la descripción del servicio representa un contrato que gestiona los mecanismos de interacción con un servicio particular, la semántica representa un contrato que administra el significado y propósito de la interacción.

Aunque unos de los principales propósitos de los Servicios Web es automatizar procesos que serían de otra manera ejecutados manualmente, las personas todavía juegan dos roles importante en su arquitectura:

- Las personas necesitan ponerse de acuerdo en la semántica y la descripción del servicio. Ya que una persona (organización) es el dueño legal de cualquier Servicio Web, las personas deben implícita o explícitamente acordar la semántica y la descripción del servicio que gobernará la interacción.
- Las personas crean los agentes solicitante y proveedor (ya sea directa o indirectamente). Finalmente, las personas deben asegurarse de que esos agentes implementen los términos de la semántica y la descripción del servicio acordados.

La arquitectura de los servicios Web se basa en una serie de estándares que permiten llevar a cabo las operaciones mencionadas y facilitan la interacción entre los componentes de un Servicio Web y de los Servicios Web entre si [WSO01]. En la Figura 3.8 se ilustra la pila de protocolos de servicios Web.

La base de la pila de protocolos de los Servicios Web es el nivel de red. Los servicios Web deben ser accesibles a través de la red para poder ser invocados por un solicitante del servicio. HTTP (*HyperText Transfer Protocol*) es el protocolo de red estándar más utilizado para la implementación de Servicios Web. Aunque existen otros protocolos de Internet entre los que se encuentran SMTP (*Simple Mail Transfer Protocol*) y FTP (*File Transfer Protocol*) que pueden ser soportados por los Servicios Web, HTTP es el de mayor uso en la Internet.

El siguiente nivel, que corresponde a la mensajería basada en XML (*Extensible Markup Language*), representa el uso de XML como base del protocolo de intercambio de mensajes. De ésta manera los Servicios Web exponen funcionalidad útil a los usuarios mediante un protocolo Web estándar. En la mayoría de casos, el protocolo utilizado es *Simple Object Access Protocol* (SOAP) que es un dialecto de XML, el cual permite a las aplicaciones invocar métodos de objetos remotos, así como recibir las respuestas de los mismos, a través de un mensaje XML transportado en HTTP. SOAP define un mecanismo estándar para incorporar extensiones a los mensajes XML usando las cabeceras SOAP y el estándar de codificación de operación o función, soportando de ésta manera las operaciones publicar, encontrar y enlazar; básicas en la arquitectura de los Servicios Web.

Desde otro punto de vista, SOAP es *firewall friendly*, ya que al viajar por HTTP no es demasiado problema para que los administradores de la red dejen pasar un paquete SOAP a través del *firewall*, reutilizando en la mayoría de las veces el puerto 80. Esto ha contribuido al gran auge del concepto de Servicios Web sobre otras tecnologías distribuidas como CORBA, EJB, COM, etc., los cuales no soportan de manera nativa el acceso por un único puerto.

Además, las extensiones de SOAP (la cual puede ser hardware especializado, intermediarios SOAP, o librerías de código invocadas por el nodo de destino del mensaje SOAP) permiten proveer servicios de autenticación, encriptación, control de acceso, procesamiento de transacciones, enrutamiento, confirmación de entrega, etc. Las estructuras SOAP *envelope* (y *attachment*) y el modelo de procesamiento de cabecera ha demostrado ser una herramienta robusta para realizar dichas tareas.

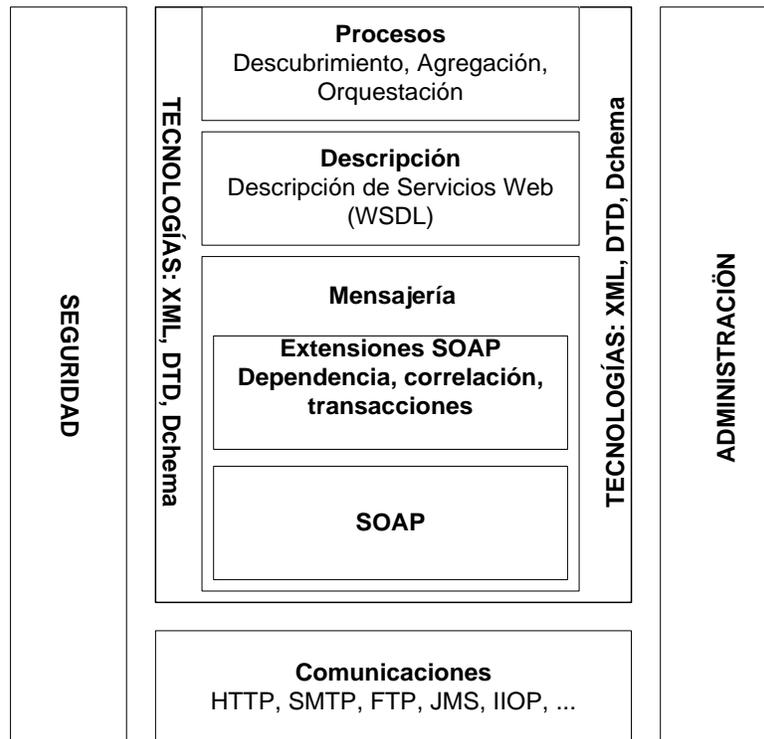


Figura 3.8 - Pila de Protocolos de los Servicios Web

Web Services Description Language (WSDL) es el estándar escogido para la descripción de servicios basado en XML. Este estándar no es el único, ya que el proceso de descripción también puede ser realizado por medio de descriptores de despliegue (*deployment descriptors*), que aunque no presentan tanto nivel de detalle como si lo ofrece WSDL, son una forma bastante sencilla de describir y publicar servicios. WSDL es la descripción del servicio necesaria para soportar la interoperabilidad de los servicios Web. WSDL define la interfaz y los mecanismos de interacción del servicio.

Como los servicios Web están definidos para ser accedidos por medio de Internet y representados por una descripción del servicio, los tres primeros niveles de la pila son requeridos para proveer cualquier servicio Web. La pila de protocolos más simple consistiría de HTTP para el nivel de red, el protocolo SOAP para el nivel de mensajería XML y WSDL para el nivel de descripción del servicio. A nivel de Intranet los servicios Web podrían ser soportados por otros protocolos de red y tecnologías de computación distribuida.

En el último nivel encontramos la publicación y descubrimiento de servicios Web. La publicación directa consiste en el envío por parte del proveedor del servicio, del documento WSDL al solicitante del servicio. En otra alternativa, el proveedor del servicio puede publicar el documento WSDL en un registro UDDI (*Universal Discovery Description and Integration*). Debido a que un servicio Web no puede ser descubierto a menos que haya sido publicado, el descubrimiento de servicios depende de la publicación de servicios. UDDI provee un mecanismo para el almacenamiento de descripciones de servicios Web. A la vez que UDDI es visto como un directorio, también define una

estructura de datos estándar para representar la información de la descripción de servicio en XML.

Desde ya se está trabajando en toda una variedad de descripción de procesos: el proceso de descubrir descripciones de servicios relacionadas por un criterio específico, el proceso de describir secuencias de mensajes, la agregación de procesos en procesos de más alto nivel, etc. En esta área encontraremos especificaciones como la *Web Service Choreography* (que todavía se encuentra en desarrollo), que busca especificar la interacción coordinada entre múltiples servicios Web en lo que se refiere al intercambio de mensajes, su composición y la secuencia en la cual estos son transmitidos y recibidos [WSC03].

En adición a las tecnologías de descripción y mensajería, la arquitectura define mecanismos para la gestión y la seguridad de los servicios Web, involucrados con los diferentes niveles de la pila de protocolos, por lo que se presenta transversal a todos ellos.

4. MODELO DE DATOS

En éste capítulo se hará la descripción de los modelos de datos elegidos (SIG-SGBD) para la implementación de la plataforma de integración.

4.1 *Sistemas de Información Geográfica*

Todos los sistemas de información geográfica (SIG) son representaciones digitales de las estructuras y procesos que observamos en el mundo real. Estas representaciones son creadas a través de un proceso conocido como construcción de modelos. Por lo tanto, su aspecto, estructura y comportamiento tienen que reflejar, como cualquier modelo, lo que nosotros conocemos u observamos. Así, el primer paso en cualquier proyecto de SIG es identificar los componentes del mundo real que pretendemos analizar, y posteriormente, representarlos en un modelo geográfico o espacial [GUI92].

A continuación, se examinará la construcción de modelos y los modelos espaciales. Luego, se presentarán las entidades espaciales y los modelos de datos, y más adelante se revisarán algunas de las estructuras de datos más utilizadas para representar entidades espaciales en el mundo de la informática. Por último, se tratarán los temas implicados en la selección del modelo espacial apropiado para una aplicación SIG.

4.1.1 Modelos Espaciales

El modelado espacial es un componente más del proceso de modelado. Los datos geográficos nos ayudan a entender y solucionar un problema concreto. Por ejemplo, tan simple como ir de 'A' a 'B' por la ruta más corta o tan complejo como un modelo de la atmósfera diseñado para simular el aumento global de la temperatura terrestre.

Antes de introducir la construcción de modelos espaciales con más detalle, se van a examinar tres modelos espaciales o geográficos que habitualmente se han trabajado [IGI91]:

- El mapa
- El modelo del arquitecto
- La fotografía aérea

Los mapas son documentos en los que se han representado mediante signos, símbolos gráficos y colores, toda una serie de datos que previamente se han recabado, analizado, depurado y sintetizado. De esta manera podemos afirmar que un mapa es una representación gráfica de la superficie de la tierra, o parte de la misma, dibujada a escala o plano [CCC99].

De otro lado, el modelo del arquitecto es uno de los modelos espaciales más fáciles de entender e interpretar, ya que utiliza la tercera dimensión (la altura) para representar características del mundo real. En éste tipo de modelos podemos ver los edificios, considerar sus localizaciones con respecto a otros y obtener una idea clara de la manera en que las construcciones interactúan con el paisaje. Aunque podemos tener más afinidad con este tipo de modelo que con un mapa, (porque se acerca más a la manera en que aparecen las cosas en el mundo real), sólo es un modelo más y representa, únicamente, una parte del mundo real.

El último modelo convencional es la fotografía aérea, ésta es una imagen del mundo real y, conceptualmente, puede ser pensada como modelo espacial del mismo modo que el mapa o el modelo del arquitecto. Mientras que para el mapa y el modelo del arquitecto, las líneas, símbolos y formas comunican una impresión del mundo real, mediante la fotografía aérea el mundo real se compone de diferentes sombras, tonos y colores. Una de las distinciones que podemos observar es que en la fotografía aérea la frontera entre características es borrosa y está representada por cambios sutiles de colores o tonos. En cambio, en el mapa y en el modelo del arquitecto las fronteras entre características están definidas de forma más precisa [CCC99].

Los tres modelos espaciales mencionados (la fotografía aérea, el modelo del arquitecto y el mapa) tratan en realidad dos dimensiones:

- La dimensión espacial; dónde se localiza un elemento.
- La dimensión temática; la naturaleza de la localización o del objeto que ocupa dicha localización.

Dichos modelos pueden ser utilizados para obtener una tercera dimensión:

- La dimensión temporal; la comparación de datos a lo largo del tiempo.

En ésta sección se presentará cómo los SIG modelan la dimensión espacial, y en el siguiente aparte (Modelo SGBD) examinaremos con detalle la dimensión temática y el modelado de la dimensión temporal.

A diferencia de otros tipos de información que habitualmente procesan los sistemas de información, los datos geográficos presentan cierto nivel de complejidad debido a que incluyen información sobre la posición, las posibles conexiones topológicas y los atributos de los objetos registrados [CRU93] Por lo tanto, podemos considerar los datos espaciales como valores, cadenas de caracteres o símbolos que proporcionan a su usuario información sobre la localización geográfica de entidades del mundo real.

Ahora bien, la mayoría de los SIG están organizados como una colección de temas o capas. Cada capa representa los valores de un atributo único del espacio geográfico. Donde cada capa puede fragmentarse por espacios, llamados objetos o entidades, los cuales son almacenados en una base de datos [JBS97]. Para cada capa geográfica se identifica la geometría y la topología. La geometría se refiere al conjunto de propiedades geométricas como son medidas, relación entre puntos, líneas, ángulos y superficies. La geometría puede ser vectorial o *raster*. La geometría vectorial esta basada en nodos, líneas y polígonos, los cuales representan el espacio ocupado por entidades del mundo real. La raster maneja un arreglo de puntos, donde cada punto representa el valor de un atributo por un área del mundo real [CML89].

Desde el punto de vista vectorial tendríamos que:

- Los Nodos, son objetos sin dimensiones que representan una unión topológica o un punto terminal y que especifican una localización geométrica. En cualquier caso, se trata de la entidad básica para representar entidades con posición pero sin dimensión (al menos a la escala escogida). En el formato vectorial se les denominan puntos.
- Las Líneas (o arcos), son objetos de una dimensión definidos por un nodo inicio y un nodo fin.
- Los Polígonos (o áreas), son objetos limitados y continuos de dos dimensiones.

Ahora bien, la topología se refiere a las propiedades de relación que existen entre entidades, como son la cercanía entre los objetos, la ubicación y la distancia.

De otro lado, retomando el término “modelo de datos”, muchos autores lo utilizan a menudo para describir los dos nuevos conceptos citados en los párrafos anteriores, vector y raster. Esto resulta algo confuso dado que el término modelado de datos es utilizado para describir también el proceso completo de representación de la realidad en un sistema informático. En éste caso, se utilizará el concepto de “modelo de datos espaciales” para referirse a los términos “raster” y “vectorial”, y los conceptos “modelo de datos” y “modelado de datos” para referirnos al proceso global de modelado.

A continuación se detallarán los dos modelos de datos espaciales, pero antes se presentarán dos analogías que permitirán comprender fácilmente estos dos conceptos. Estas analogías se construyen sobre dos juguetes comunes, el rompecabezas (acercamiento raster) y los dibujos Une-los-Puntos (acercamiento vector). Ahora imagínate que tienes que recrear un mapa utilizando ambos medios. En el rompecabezas lo harías uniendo las piezas de formas variadas sobre un tablero de base cuadrada. Con el juego Une-los-Puntos crearías tu mapa a partir de una serie de puntos con códigos numéricos que indican el orden para unir los elementos [COV98].

4.1.1.1 Estructura de Datos Vectorial

En la sección anterior se mostró que en el mundo vectorial los tres principales tipos de entidades geográficas (nodos - puntos, líneas, polígonos) y los tipos de entidades auxiliares como las redes y las superficies, pueden ser definidos mediante pares de coordenadas x-y. Los puntos son representados por dos pares de coordenadas, las líneas por arcos unidos a puntos, los polígonos por líneas cerrando áreas, las redes por líneas conectadas y las superficies por áreas que unen puntos y líneas [RSM87]. Sin embargo, es necesaria una información adicional muy importante. Consiste en la descripción de las relaciones geográficas entre puntos y líneas que son utilizadas para representar una entidad. Estas relaciones espaciales son llamadas “topología”.

De acuerdo con lo anterior, las estructuras de datos para representar una entidad en el mundo vectorial pueden ser divididas en dos grandes categorías [COV98]:

- Estructuras de datos con topología

- Estructuras de datos sin topología

Cuando se pasa de manejar figuras geométricas a manejar objetos geográficos en los que deben hacerse uniones, intersecciones y sustracciones de objetos con atributos, se recurre a la ayuda algebraica de la topología, la cual permite representar una configuración geométrica y sus relaciones.

Mediante la topología es posible mantener las relaciones geométricas existentes entre los objetos (representados por puntos, líneas y polígonos) de acuerdo con la estructura de los grafos. Por ejemplo, es posible mantener la relación espacial aunque se rote la figura.

La posibilidad de crear y almacenar relaciones topológicas brinda varias ventajas:

- La topología almacena datos más eficientemente.
- Agiliza procesos de análisis de datos.
- Brinda la posibilidad de combinar polígonos adyacentes para generar nuevas características y superponer figuras geográficas.

El modelo topológico posibilita que los objetos representados en el sistema conserven, no sólo su posición absoluta determinada por las coordenadas, sino también su ubicación relativa en relación con sus vecinos. Esto tiene gran importancia en muchas aplicaciones, por ejemplo, cuando se está trabajando con las vías de una ciudad: además de su posición absoluta, es también relevante la conectividad existente entre los diferentes elementos que forman la malla vial.

La topología reduce el riesgo de almacenar datos redundantes, principalmente cuando se trabaja con polígonos que comparten arcos (caso común), además esta estructura permite la identificación y corrección automática de errores cometidos durante la digitalización.

Las características principales de este modelo se pueden condensar de la siguiente manera:

- Enlaza segmentos de una red de límites.
- Verifica que los polígonos estén bien cerrados.
- Calcula automáticamente la superficie de los polígonos.
- Permite la asociación de atributos no espaciales a los polígonos.

Entre los principales modelos topológicos pueden mencionarse [COV98]:

- El modelo DIME (Dual Independent Map Encoding)
- La estructura Arc-Node o POLIVERT
- Estructura relacional
- El modelo DLG (Digital Line Graph Structure)

De lo anterior se puede concluir que un verdadero SIG debe utilizar una estructura de datos topológicos para representar las entidades espaciales, si pretende ir más allá de la simple visualización de las características de cada entidad, a tener capacidad de análisis espacial.

Si se tuviera una estructura de datos sin topología, la computadora sería capaz de reproducir la imagen, pero surgiría un problema tan pronto como se intentó utilizar esta información para realizar preguntas sobre una red, (las líneas simples no contienen, inherentemente, información espacial sobre su conectividad. Únicamente necesitan tener información topológica unida a ellas cuando forman parte de una red, un polígono o una superficie). Esto se debe a que la computadora no ha sido provista con ninguna información que le diga que el polígono, la línea o el punto poseen relaciones de vecindad, ubicación ó distancia. De ésta manera, estos vínculos espaciales únicamente son deducidos por la mente cuando las líneas aparecen en la pantalla del monitor y no están definidos explícitamente dentro de nuestro archivo de datos. Esta situación de apariencia visual de una estructura en pantalla pero sin presencia en el archivo de datos ha originado estructuras de datos vectoriales simples de este tipo, sin topología, llamadas "spaghetti" [GGO94] Los datos "spaghetti", en realidad, son simplemente una mezcla de entidades lineales. Hay dos problemas específicos de las estructuras de datos spaghetti que ilustran por qué la información topológica es importante. Primero, los datos spaghetti no contienen información de vecindad y, segundo, la estructura de datos es incapaz de enfrentarse con lo que se denomina agujeros o polígonos isla.

4.1.1.2 Estructura de Datos Raster

Los Sistemas de Información Raster basan su funcionalidad en una concepción implícita de las relaciones de vecindad entre los objetos geográficos. Su forma de proceder es dividir la zona de afección de la base de datos en una retícula o malla regular de pequeñas celdas (píxeles) y atribuir un valor numérico a cada celda como representación de su valor temático (Ver Figura 4.1). Dado que la malla es regular, el tamaño del píxel es constante y se conoce la posición en coordenadas del centro de una de las celdas; se puede decir que todos los píxeles están georreferenciados. De esta manera la geometría de un objeto sólo se puede describir por medio de las celdas vecinas que tengan el mismo atributo.

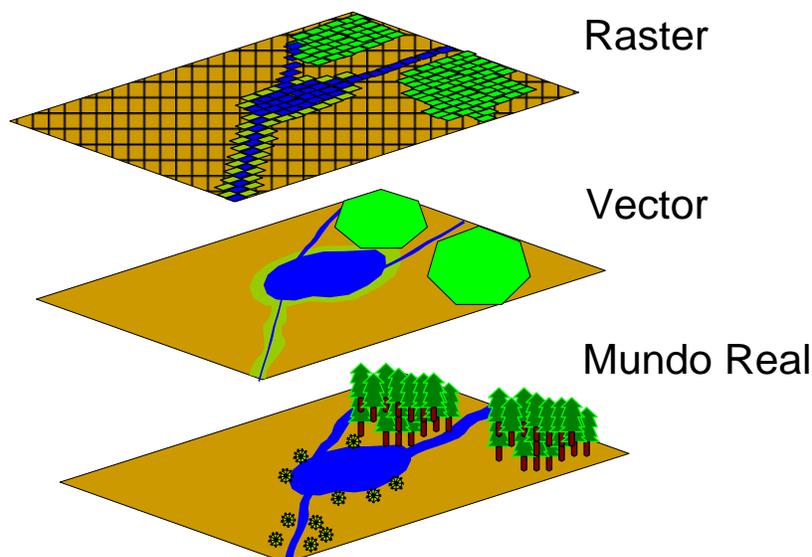


Figura 4.1 - Representación gráfica de los modelos Vector y Raster

A continuación se presentarán una serie de acercamientos del modelo raster, que incrementarán gradualmente su complejidad.

En un **raster simple** la información de la estructura de datos es almacenada para cada celda de la imagen. Esta simple información dice al computador si una característica está presente o no en una celda determinada. Un problema común que se da en todas las estructuras de datos raster es qué hacer con las entidades que ocupan parcialmente una celda de una malla. Dos de los métodos más comúnmente utilizados para solventar este problema son la regla del 50% y las normas de presencia/ausencia. La norma del 50 % establece que si una entidad geográfica ocupa más del 50% de un pixel dicha entidad será reconocida como presente en ese pixel. La norma de presencia/ausencia establece que si una entidad está presente en una celda de una malla, incluso aunque sea mínimamente, dicha entidad es considerada como presente en ese pixel.

Utilizando un acercamiento raster simple, para poder representar cada tipo de entidad en el computador, debemos tener una capa de información separada para cada clase de entidad.

Uno de los mayores problemas con los datos raster es su tamaño. Cada celda debe tener un valor, que será registrado y almacenado. Esto significa que el tamaño de almacenamiento para una imagen compleja formada por un mosaico de diferentes entidades (por ejemplo un mapa del suelo con 20 clases distintas) requiere la misma cantidad de espacio de almacenamiento que un mapa raster similar que muestre la localización de una única carretera en la misma área, a pesar del hecho que gran parte de esta última imagen raster contiene celdas que registran el valor 0 (no presentan una entidad). El almacenamiento de información innecesaria en el mundo raster es un tema que ha recibido una atención considerable. Se han desarrollado una serie de métodos de compactación de datos para poder reducir considerablemente el tamaño de un conjunto de datos raster, tales como:

- Compresión de valores consecutivos (Run length encoding)
- Codificación por bloques (Block coding)
- Codificación en cadena (chain coding)
- Árbol cuaternario (quadtree)

Comas y Ruiz (1993) explican que las aproximaciones raster y vectorial son dos métodos alternativos para el almacenamiento y representación de los fenómenos espaciales. Como modelos, tienen algunos puntos fuertes y débiles para describir las condiciones del mundo real. En la siguiente tabla se pueden observar las ventajas y desventajas comparativas entre los dos modelos.

Modelo Raster	Modelo Vectorial
Ventajas:	
Utiliza una estructura de datos muy simple.	Posee una estructura de datos muy compacta.
Las superposiciones de las diferentes coberturas se implementan de forma rápida y eficiente.	Codifica eficientemente la topología.
Permite una forma más eficiente de representación cuando la variación espacial es muy alta. El modelo raster es muy apropiado para el tratamiento de	La salida en papel presenta muy buenos productos

imágenes de satélite.	
Da la posibilidad de generar modelos de elevación del terreno.	
Desventajas:	
La estructura de datos es menos compacta	La estructura de datos es más compleja
Algunas relaciones topológicas son difíciles de representar	Las sobreposiciones son más complejas de realizar
La información original se generaliza una vez que se traspasa al sistema, tanto cuanto más grande sea la dimensión de las celdas.	Si la variación espacial es baja, resulta poco eficiente la aplicación.
La mayoría de estos SIGs se ven limitados por la cantidad de filas y columnas que pueden manejar, por tanto la resolución dependerá de éstas.	El procesamiento de imágenes digitales no puede ser realizado eficientemente en este tipo de formato.
La salida en duro no resulta de muy buena calidad.	

Tabla 4.1 – Comparación entre Modelos

4.2 Sistema de Gestión de Bases de Datos

En la actualidad, la mayoría de los SIG almacenan los datos alfanuméricos en un Sistema de Gestión de Bases de Datos Relacional (SGBD). Estos datos, hacen referencia a la dimensión temática de los Sistemas de Información Geográfica, la cual representa la naturaleza de los objetos que ocupan una posición en el espacio. De esta manera, el SGBD permite al SIG organizar y gestionar los datos almacenados en la base de datos; suministrando funciones típicas para relacionar lógicamente y físicamente dichos datos, recuperarlos y verificarlos, asegurar su integridad, actualizarlos y controlar su consistencia.

En esta sección se presentarán los modelos de datos utilizados para representar las relaciones que permiten estructurar una base de datos, de tal manera que el SGBD tenga una base conceptual y lógica que garantice su consistencia [BAT94].

4.2.1 Modelo Entidad – Relación

El modelo entidad-relación es el modelo más utilizado para el diseño conceptual de bases de datos. Fue introducido por Peter Chen en 1976. El Modelo Entidad-Relación está formado por un conjunto de conceptos que permiten describir la realidad mediante un conjunto de representaciones gráficas y lingüísticas [BAR94].

Este modelo considera como punto de partida la existencia de entidades, que representan objetos, personas, etc, sobre las que se quiere almacenar información relevante del mundo real. A las características necesarias para describir completamente cada tipo de entidad se les denominará atributo.

Las características fundamentales de este modelo son:

- Reflejan tan sólo la existencia de los datos sin expresar lo que se hace con ellos.
- Es independiente de las bases de datos y de los sistemas operativos
- Incluye todos los datos que se estudian sin tener en cuenta las aplicaciones que se van a tratar

Originalmente, el modelo entidad-relación sólo incluía los conceptos de entidad, relación y atributo. Más tarde, se añadieron otros conceptos, como los atributos compuestos y las jerarquías de generalización, en lo que se ha denominado modelo entidad-relación extendido.

En la Figura 4.2 se observa la representación gráfica de los componentes del modelo entidad relación. A continuación se explicaran cada uno de éstos términos.

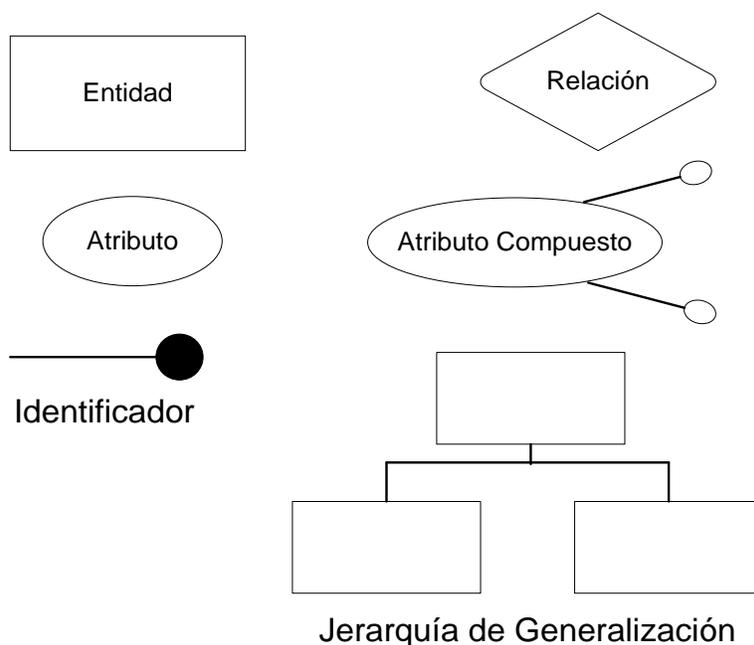


Figura 4.2 – Representación del Modelo Entidad-Relación

Entidad

Cualquier tipo de objeto o concepto sobre el que se recoge información: cosa, persona, concepto abstracto o suceso. Por ejemplo: coches, casas, empleados, clientes, empresas, oficios, diseños de productos, conciertos, excursiones, etc. Las entidades se representan gráficamente mediante rectángulos y su nombre aparece en el interior. Un nombre de entidad sólo puede aparecer una vez en el esquema conceptual.

Hay dos tipos de entidades: fuertes y débiles. Una entidad débil es una entidad cuya existencia depende de la existencia de otra entidad. Una entidad fuerte es una entidad que no es débil.

Relación (interrelación)

Es una correspondencia o asociación entre dos o más entidades. Cada relación tiene un nombre que describe su función. Las relaciones se representan gráficamente mediante rombos y su nombre aparece en el interior.

Las entidades que están involucradas en una determinada relación se denominan entidades participantes. El número de participantes en una relación es lo que se denomina grado de la relación. Por lo tanto, una relación en la que participan dos entidades es una relación binaria; si son tres las entidades participantes, la relación es ternaria; etc.

Una relación recursiva es una relación donde la misma entidad participa más de una vez en la relación con distintos papeles. El nombre de estos papeles es importante para determinar la función de cada participación.

La cardinalidad con la que una entidad participa en una relación especifica el número mínimo y el número máximo de correspondencias en las que puede tomar parte cada ocurrencia de dicha entidad. La participación de una entidad en una relación es obligatoria (total) si la existencia de cada una de sus ocurrencias requiere la existencia de, al menos, una ocurrencia de la otra entidad participante. Si no, la participación es opcional (parcial). Las reglas que definen la cardinalidad de las relaciones son las reglas de negocio.

A veces, surgen problemas cuando se está diseñado un esquema conceptual. Estos problemas, denominados trampas, suelen producirse a causa de una mala interpretación en el significado de alguna relación, por lo que es importante comprobar que el esquema conceptual carece de dichas trampas. En general, para encontrar las trampas, hay que asegurarse de que se entiende completamente el significado de cada relación. Si no se entienden las relaciones, se puede crear un esquema que no represente fielmente la realidad.

Una de las trampas que pueden encontrarse ocurre cuando el esquema representa una relación entre entidades, pero el camino entre algunas de sus ocurrencias es ambiguo. El modo de resolverla es reestructurando el esquema para representar la asociación entre las entidades correctamente.

Otra de las trampas sucede cuando un esquema sugiere la existencia de una relación entre entidades, pero el camino entre una y otra no existe para algunas de sus ocurrencias. En este caso, se produce una pérdida de información que se puede subsanar introduciendo la relación que sugería el esquema y que no estaba representada.

Atributo

Es una característica de interés o un hecho sobre una entidad o sobre una relación. Los atributos representan las propiedades básicas de las entidades y de las relaciones. Toda la información extensiva es portada por los atributos. Gráficamente, se representan mediante bolitas que cuelgan de las entidades o relaciones a las que pertenecen.

Cada atributo tiene un conjunto de valores asociados denominado dominio. El dominio define todos los valores posibles que puede tomar un atributo. Puede haber varios atributos definidos sobre un mismo dominio.

Los atributos pueden ser simples o compuestos. Un atributo simple es un atributo que tiene un solo componente, que no se puede dividir en partes más pequeñas que tengan un significado propio. Un atributo compuesto es un atributo con varios componentes, cada uno con un significado por sí mismo. Un grupo de atributos se representa mediante un atributo compuesto cuando tienen afinidad en cuanto a su significado, o en cuanto a su uso. Un atributo compuesto se representa gráficamente mediante un óvalo.

Los atributos también pueden clasificarse en monovalentes o polivalentes. Un atributo monovalente es aquel que tiene un solo valor para cada ocurrencia de la entidad o relación a la que pertenece. Un atributo polivalente es aquel que tiene varios valores para cada ocurrencia de la entidad o relación a la que pertenece. A estos atributos también se les denomina multivaluados, y pueden tener un número máximo y un número mínimo de valores. La cardinalidad de un atributo indica el número mínimo y el número máximo de valores que puede tomar para cada ocurrencia de la entidad o relación a la que pertenece. El valor por omisión es (1,1). Por último, los atributos pueden ser derivados. Un atributo derivado es aquel que representa un valor que se puede obtener a partir del valor de uno o varios atributos, que no necesariamente deben pertenecer a la misma entidad o relación.

Identificador

Un identificador de una entidad es un atributo o conjunto de atributos que determina de modo único cada ocurrencia de esa entidad. Un identificador de una entidad debe cumplir dos condiciones:

- No pueden existir dos ocurrencias de la entidad con el mismo valor del identificador.
- Si se omite cualquier atributo del identificador, la condición anterior deja de cumplirse.

Toda entidad tiene al menos un identificador y puede tener varios identificadores alternativos. Las relaciones no tienen identificadores.

Jerarquía de generalización

Una entidad E es una generalización de un grupo de entidades E₁, E₂,... E_n, si cada ocurrencia de cada una de esas entidades es también una ocurrencia de E. Todas las propiedades de la entidad genérica E son heredadas por las subentidades.

Cada jerarquía es total o parcial, y exclusiva o superpuesta. Una jerarquía es total si cada ocurrencia de la entidad genérica corresponde al menos con una ocurrencia de alguna subentidad. Es parcial si existe alguna ocurrencia de la entidad genérica que no corresponde con ninguna ocurrencia de ninguna subentidad. Una jerarquía es exclusiva si cada ocurrencia de la entidad genérica corresponde, como mucho, con una ocurrencia de

una sola de las subentidades. Es superpuesta si existe alguna ocurrencia de la entidad genérica que corresponde a ocurrencias de dos o más subentidades diferentes.

Un subconjunto es un caso particular de generalización con una sola entidad como subentidad. Un subconjunto siempre es una jerarquía parcial y exclusiva.

Una vez examinados los conceptos del modelo entidad-relación, a continuación se presenta un ejemplo gráfico que especifica la existencia de tres entidades: Profesor, Curso y Departamento. De acuerdo a la percepción de la realidad, un departamento tiene muchos profesores y un profesor puede dar muchos cursos, de tal manera que se puede definir una relación uno a muchos entre el profesor y los cursos, de la misma manera que entre el departamento y los profesores. Tal como se presenta en la Figura 4.3, para cada una de las entidades existen un conjunto de atributos que las describen, teniendo en cuenta que para cada una de las entidades existe un atributo que las identifica únicamente (palabra subrayada).

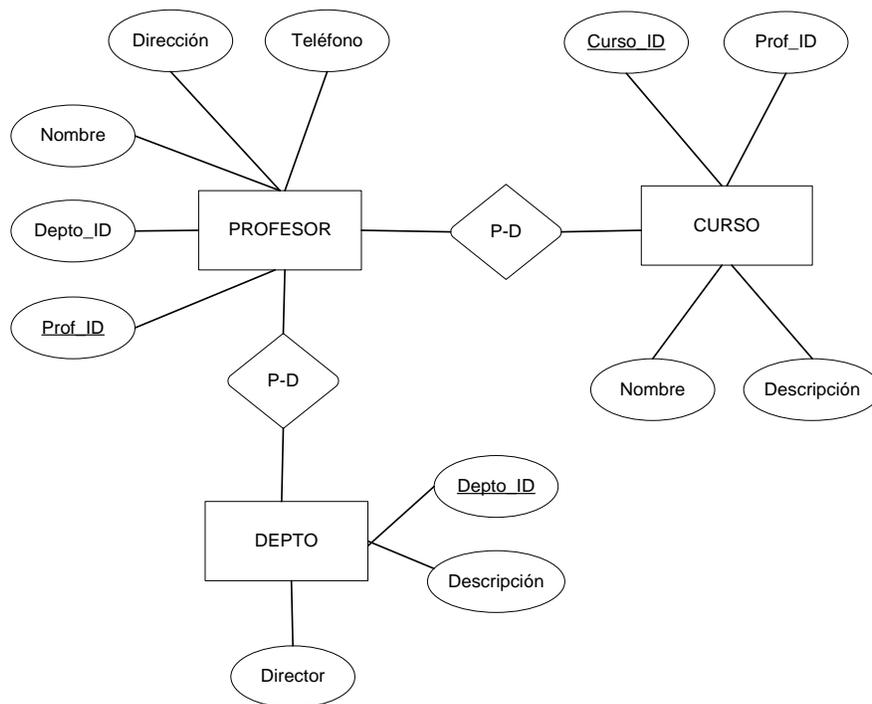


Figura 4.3 – Ejemplo del Modelo Entidad-Relación

Finalmente, las tres entidades contempladas son consideradas como independientes, aunque también se podría modelar la existencia de alguna de ellas como dependiente de otra; por ejemplo se podría establecer la restricción de no existencia de un profesor debido a que no esta adscrito a ningún departamento, o que un curso no puede existir sin un profesor que lo imparta.

4.2.2 Modelo Relacional

En 1970, el modo en que se veían las bases de datos cambió por completo cuando E. F. Codd introdujo el modelo relacional [COD90]. De esta manera, después de haber definido las entidades y sus atributos en el Modelo Entidad–Relación, éstos son representados lógicamente a través de tablas expresadas en el Modelo Relacional.

El modelo relacional representa los datos y las relaciones entre los datos mediante una colección de tablas, cada una de las cuales tiene un número de columnas con nombres únicos.

La estructura fundamental del modelo relacional es precisamente esa, "relación", es decir una tabla bidimensional constituida por filas (tuplas) y columnas (atributos). Cada instancia de la entidad encontrará sitio en una tupla de la relación, mientras que los atributos de la relación representarán las propiedades de la entidad. En realidad, siendo rigurosos, una relación es sólo la definición de la estructura de la tabla, es decir su nombre y la lista de los atributos que la componen. Cuando se puebla con las tuplas, se habla de "instancia de relación".

El Modelo Relacional esta compuesto por tres elementos [FHA89]:

- Un elemento estructural que describe la forma en que pueden guardarse los datos
- Un elemento de manipulación que describe un número de operadores que permiten a los usuarios procesar los datos guardados en formato relacional
- Un elemento de integridad que propone reglas para asegurarnos que los datos se gestionan con validez y consistencia

Considerando el elemento estructural, una base de datos es vista por los usuarios como una colección de tablas (y nada más que tablas) [FHA89]. Superficialmente, el elemento estructural del Modelo Relacional no podría ser más sencillo, ya que todos los datos parecen estar almacenados en tablas bidimensionales. Así, para los usuarios, el Modelo Relacional es un modelo conceptual muy familiar, permitiéndoles creer que los datos se guardan dentro del ordenador de una forma muy parecida a las tablas en papel con las que estamos tan familiarizados (ver Figura 4.4).

Las tablas contenidas en un SGBDR deben cumplir unos requisitos específicos para que el software funcione correctamente. Aunque el modelo conceptual de un usuario puede, simplemente, consistir en tablas de almacenamiento de datos, el diseñador de bases de datos debe ser consciente de las restricciones que puso Codd sobre el formato de las tablas relacionales y tiene que diseñar una estructura de datos que se adapte a estas restricciones. Una tabla relacional debe tener, especialmente, las siguientes características [REE99]:

- a. **Todas las columnas deben tener nombres distintos:** Los nombres de las columnas se utilizan como referencia para acceder a los valores que hay en ella (ver Figura 4.4).
- b. **Todas las entradas de una columna deben ser del mismo tipo:** El Modelo Relacional define que cada columna de una tabla sólo puede tener valores del mismo dominio. Un dominio es el conjunto de todos los valores

posibles que el elemento puede adoptar. Un ejemplo de éste concepto se presenta en la columna **num_ompl** de la Figura 4.4, la cual tiene como dominio los números enteros positivos. Si intentamos introducir caracteres o decimales en la dicha columna, se provocaría un error, ya que el tipo caracter no corresponde con el dominio definido. En el modelo de Codd sólo se pueden hacer comparaciones entre columnas que utilizan datos del mismo dominio. No tiene sentido en el mundo real preguntarse, por ejemplo, si el Día es igual a la Distancia, aunque los dos pueden ser representados como números enteros. El concepto de dominio es un intento de incorporar este tipo de conocimiento del mundo real en el Modelo Relacional. El concepto de dominio, por lo tanto, intenta evitar la entrada de datos ilegales y de prevenir comparaciones ilógicas.

- c. **Todas las entradas en una tabla relacional deben ser ‘atómicas’:** Como se muestra en la figura 4.4, en las tablas relacionales cada celda debe tener una única entrada. Llamamos ‘atómicas’ a las entradas porque para el propósito de la base de datos no tienen estructura interna, es decir, no pueden descomponerse.

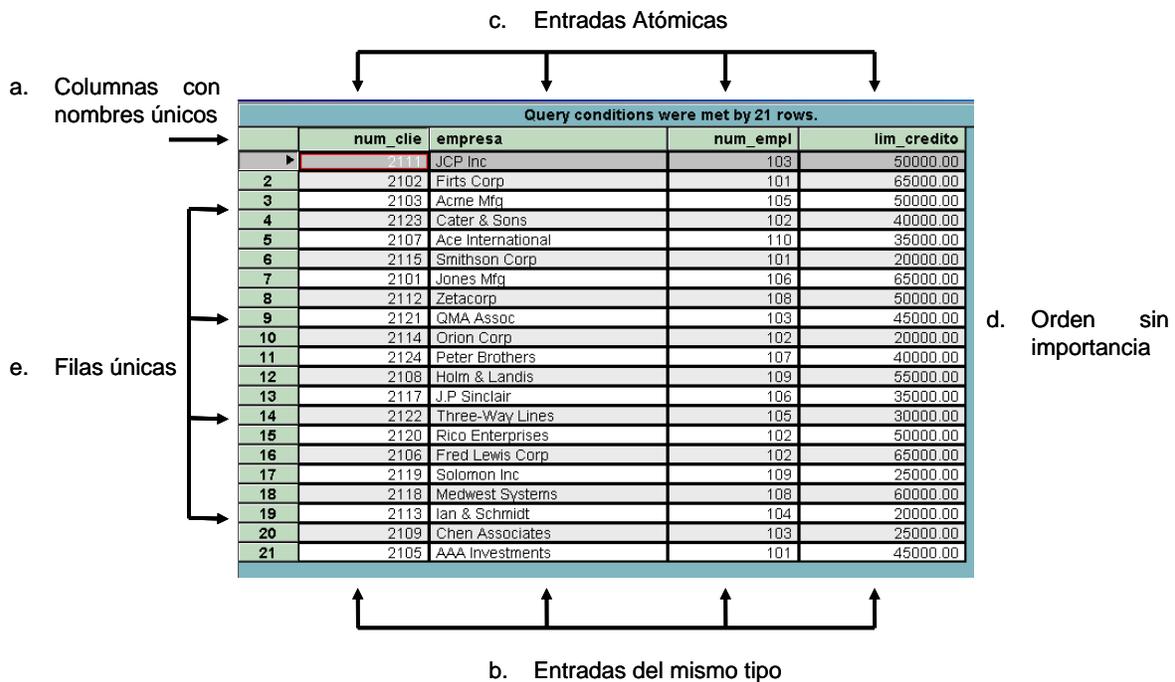


Figura 4.4 – Tabla Relacional

- d. **El orden de las columnas y filas no tiene importancia:** En las tablas en papel a menudo utilizamos el mismo orden en que entramos los datos para tratar la información. Podríamos, por ejemplo, entrar los datos ordenándolos por año o almacenar detalles personales ordenándolos por apellido. En una tabla relacional esto no está permitido. Se tiene que asumir que los datos entrados en las columnas y filas están ordenados al azar.

- e. **Cada fila de la tabla debe ser diferente de las demás:** En una tabla relacional no pueden existir filas duplicadas. En otras palabras, debe haber una columna (o combinación de columnas) que identifique de forma única cada fila de datos, y aquí es dónde se encuentra por primera vez el concepto de 'clave' (La columna **num_clie** de la tabla relacional de la Figura 4.4, presenta éste concepto). La clave es una columna (o combinación de columnas) con un valor único en cada fila. En ningún momento puede haber dos filas con el mismo valor. Por lo tanto, puede utilizarse para identificar de forma individual, a la información contenida en cualquier fila.
- f. **Valores Nulos:** Cuando en una tupla un atributo es desconocido, se dice que es nulo. Un nulo no representa el valor cero ni la cadena vacía, éstos son valores que tienen significado. El nulo implica ausencia de información, bien porque al insertar la tupla se desconocía el valor del atributo, o bien porque para dicha tupla el atributo no tiene sentido. Ya que los nulos no son valores, deben tratarse de modo diferente, lo que causa problemas de implementación.
- g. **Tablas Base y Tablas Vista:** En el Modelo Relacional las tablas pueden ser de dos tipos, tablas base y tablas vista. Una tabla base es 'real' en el sentido que dentro de la base de datos habrá una estructura física, normalmente un archivo, que contendrá los datos referidos en la tabla. Una tabla vista es una 'tabla virtual' formada por una combinación de filas y columnas seleccionadas a partir de una o más tablas base y/o a partir de vistas previamente definidas. Para el usuario una vista tiene el mismo aspecto que una tabla real y puede ser consultada utilizando los mismos comandos, pero las vistas no tienen la misma realidad física que las tablas base. El propósito de las vistas consiste en permitir que los datos sean presentados a los diferentes usuarios en el formato que interesa a cada uno de ellos.
- h. **Claves externas:** Hasta el momento se ha realizado una descripción de modelo relacional centrándose en las propiedades de una única tabla relacional, pero evidentemente en la mayoría de las bases de datos habrá más de una tabla y será necesario tener un método para enlazarlas. De nuevo, el modelo relacional dispone de un método sencillo e intuitivo para enlazar unas tablas con otras. En la figura 4.5, las tablas 1 y 2 están conectadas entre ellas mediante atributos compartidos. La clave primaria de la tabla 1 (columna **num_empl**) puede estar incluida como atributo en la tabla 2 y sirve de enlace entre las dos. De esta manera, la columna de una tabla que es la clave primaria en otra es llamada clave externa.

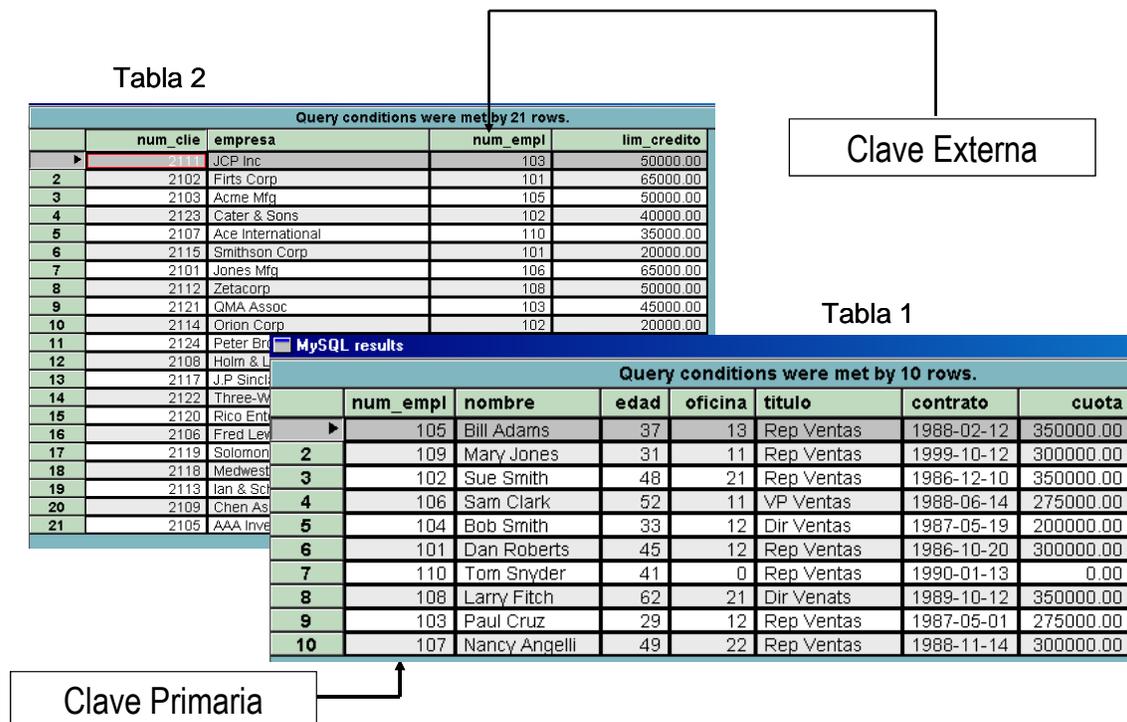


Figura 4.5 – Relaciones entre Tablas

En el Modelo Relacional de Codd el elemento de manipulación consiste en un conjunto de operadores conocido como álgebra relacional. Los usuarios de un software de BD relacional, sin embargo, no utilizan esta álgebra directamente. El álgebra relacional existe para declarar, matemáticamente, lo que es posible en el modelo relacional; en otras palabras, ofrece un criterio con el que evaluar los lenguajes SGBD implementados, como el SQL (Structured Query Language).

El álgebra relacional es sorprendentemente sencilla, pero su simplicidad puede tornarse compleja debido a los términos matemáticos que pueden utilizarse para describirlo. Un buen modelo mental que se puede utilizar, consiste en pretender que tenemos un número de tablas en papel convencional a partir de las cuales queremos crear nuevas tablas. Para hacer esto manualmente se necesitaría llevar a cabo un conjunto de operaciones de “cortar, pegar y copiar” hasta haber creado las nuevas tablas. El álgebra relacional puede ser vista como un servicio que ofrece el equivalente electrónico de unas tijeras, pegamento y fotocopidora para trabajar sobre las tablas electrónicas contenidas en el SGBD. Existen una gran cantidad de operadores relacionales, pero a continuación se presentarán los tres operadores relacionales fundamentales [GWE98]:

- **Select** (Selección): Este operador crea un subconjunto con todas las filas de una tabla que cumplen con un criterio determinado.
- **Project** (Proyección): Crea un subconjunto de columnas de una tabla.
- **Join**: Es el operador crucial que permite que los datos almacenados en tablas separadas se ‘peguen’ juntos temporalmente en una columna compartida.

El elemento de integridad de una base de datos se refiere a la exactitud y grado de corrección de los datos que están almacenados en ella. En el modelo relacional de Codd se proponen dos tipos de integridad:

- Integridad de entidades
- Integridad referencial

Integridad de Entidades: Esta hace referencia a la clave primaria de la tabla y simplemente consiste en que cada fila en la tabla relacional tendrá una entrada válida (o sea, no nulo) y única para los atributos de la clave primaria. Si no se impusiera esta regla podría suceder que las filas de una tabla existan pero no estén accesibles.

Integridad Referencial: Esta regla requiere que los valores de una clave externa existan como clave primaria en otra tabla. Si esta regla no existiera sería posible encontrar una fila con una clave externa que apunta a una fila de otra tabla que no existe.

5. ARQUITECTURA DE LA PLATAFORMA

Este capítulo contiene una descripción detallada de arquitectura a utilizar para la plataforma, la cual está basada en el estado del arte realizado en el capítulo 3. Se incluye también la descripción del proceso de construcción del prototipo final de la plataforma.

5.1 Introducción

Desde el punto de vista ambiental, la Plataforma permitirá la integración de tres componentes básicos que son:

- Módulo de Adquisición Remota de Información Ambiental (Clima).
- Módulo de Gestión de Información Hídrica
- Módulo de Gestión de Información Geológica

La interacción entre estos componentes se puede ver en el esquema de la Figura 5.1. Bajo esta perspectiva, la plataforma facilita la integración de la información y las aplicaciones de las entidades que tienen responsabilidades o interés en la gestión ambiental. En la parte superior se pueden observar los clientes que poseen la capacidad de realizar análisis espacial (Cliente1: INGEOMINAS, Cliente 2: UNIDAD ADMINISTRATIVA ESPECIAL DE PARQUES NACIONALES NATURALES (UAESPNN) y Cliente 3: UNIVERSIDAD DEL CAUCA), los cuales a través de la plataforma de integración soportada en Internet, pueden acceder a los datos ubicados en los servidores de Gestión de Información (e.g. geológica, hídrica, clima, faunística, de flora, etc.) [CRF03].

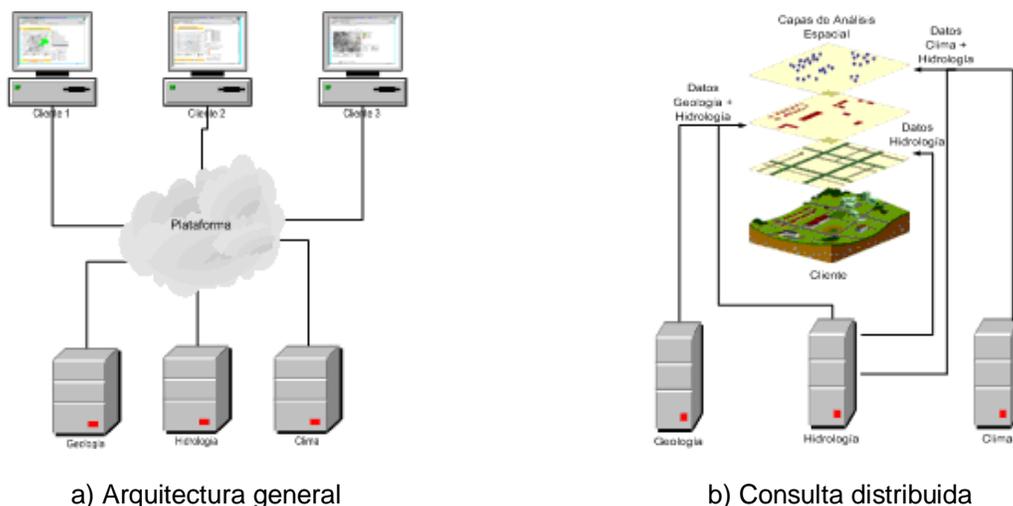


Figura 5.1 - Sistema Integrado de Información Ambiental

De esta manera un cliente puede generar una capa de análisis espacial con información proveniente de todas las fuentes de datos (Figura 5.1-b).

Si bien es cierto que la Plataforma propuesta, se enfoca en agilizar los procesos de intercambio de datos cartográficos y temáticos entre las diferentes entidades que generan información pertinente para realizar Gestión Ambiental de una zona en común, (INGEOMINAS – Módulo de Gestión de información Geológica, UAESPNN – Módulo de Gestión Información Hídrica, y Universidad del Cauca – Módulo de Gestión de Información de Clima), ésta debe permitir que los sistemas que la componen (Módulo de Gestión de Información Hídrica, Geológica y Clima) conserven la autonomía de cada institución a la cual pertenecen (INGEOMINAS, UAESPNN, Universidad del Cauca). De ésta manera, dicha plataforma se debe encargar de generar nuevas vistas de la Gestión Ambiental (Vistas Globales de la Base de Datos Virtual) en función de la combinación de las capacidades de cada sistema de gestión de información, dando solución a las heterogeneidades de los sistemas participantes y facilitando la evolución del sistema global enmarcado en el dominio de la Gestión Ambiental.

En la siguiente sección se presentará la arquitectura para la Plataforma de Integración de Múltiples Fuentes de Datos Descriptivos y Espaciales en un Sistema de Información Ambiental. Posteriormente se describirán las principales funciones que implementa la arquitectura propuesta, y finalmente se expondrán las tecnologías de desarrollo utilizadas.

5.2 Arquitectura Conceptual

Trasladando al campo técnico las necesidades expuestas por las instituciones participantes en el proyecto, se tiene que la arquitectura de la plataforma debe cumplir con las siguientes características:

Desde el punto de vista de la integración de datos:

- Las heterogeneidades deben ser resueltas automáticamente por la plataforma, de tal manera que se elimine la existencia de un “super usuario” con el conocimiento de cada una de las bases de datos participantes. Este super-usuario haría al sistema muy propenso a errores y muy costoso en horas-hombre dedicadas.
- Los sistemas participantes deben mantener su autonomía, especialmente la autonomía de participación que facilite alcanzar una visión lógica de una única base de datos.
- La plataforma debe garantizar la evolución de los sistemas participantes, de tal manera que ante cambios en las bases de datos componentes, no haya necesidad de construir nuevamente toda la plataforma.
- Debe existir un alto grado de transparencia en el mapeamiento de consultas entre la plataforma y los sistemas participantes. De ésta manera no hay necesidad de conocer la estructura de los sistemas componentes para establecer el mapeamiento de las consultas.

- Otra de las características de la plataforma se enmarca en la necesidad de lograr interoperabilidad con sistemas que no únicamente sean SGBD.

Desde el punto de vista de la conexión:

- Cada sistema debe ser expuesto como un servicio autónomo.
- El acceso a los sistemas componentes debe realizarse a través de interfaces estándar.
- Los usuarios deben poder hacer uso de los servicios de la plataforma a través de Internet (HTTP).
- La comunicación entre los sistemas componentes debe realizarse utilizando un lenguaje estándar y de fácil transporte a través de Internet (XML).
- La plataforma debe permitir que las invocaciones entre los diferentes sistemas se realicen por un solo puerto de conexión.

Basándose en las necesidades planteadas para la arquitectura de la plataforma, y considerando el contexto teórico descrito en el capítulo tres, se propone una arquitectura basada desde la perspectiva de integración en un Sistema Mediador y envoltorios (wrappers), y desde el punto de vista de conexión en las tecnologías de Servicios Web.

En la Figura 5.2 se muestra el esquema conceptual de la plataforma, donde se pueden apreciar tres módulos bien definidos: Mediador, Wrapper y Visualización. Los dos primeros constituyen el eje central del sistema, y el último permite visualizar los resultados de las consultas. Por un lado, el módulo Mediador, se encarga de la gestión y control central de todo el proceso de intercambio de información, permitiendo realizar gestión de fuentes de información, usuarios, permisos y demás. De la misma manera, el módulo Wrapper se encarga del encapsulamiento de cada sistema participante, logrando el mapeamiento de los datos estructurados de acuerdo a la perspectiva de un Sistema de información Geográfica (Sistema Componente), a un formato de datos comprensible para el mediador.

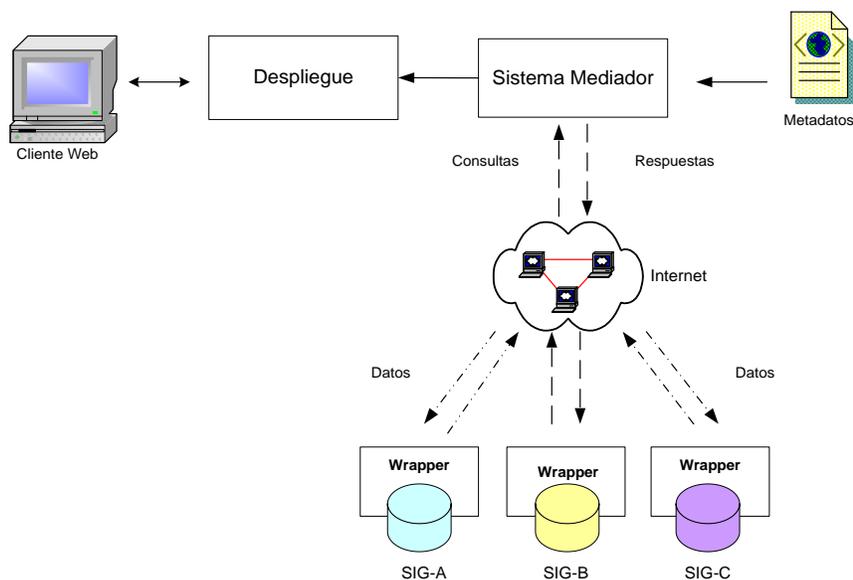


Figura 5.2 – Esquema Conceptual de la Arquitectura

5.3 Funcionalidades

Cada uno de los módulos principales (Mediador y Wrapper) trabaja como ente independiente. De esta manera, el módulo Mediador hace uso los servicios del módulo Wrapper y viceversa. Antes de abordar la descripción de cada uno de estos módulos, se presentará un glosario de términos que facilita el entendimiento de sus funcionalidades.

Glosario de términos

- **Wrapper:** (Envoltorio) Módulo encargado de realizar una interfaz entre el mediador y la base de datos de cada SIG local.
- **Mediador:** Módulo encargado de la gestión central del proceso de intercambio de información.
- **Capacidades de búsqueda:** Subconjunto de información de la fuente de datos que puede ser consultada a través del Wrapper.
- **Metadatos:** Definen un lenguaje común dentro del contexto ambiental para toda la plataforma. Se basa en el catálogo de objetos del ICDE (Infraestructura Colombiana de Datos Espaciales). En definitiva es un documento XML creado en el Mediador que describe todos los temas, objetos y atributos del catalogo de objetos.
- **Tema de los Metadatos:** Clasificación de los Metadatos en las que se agrupan los objetos.
- **Objeto de los Metadatos:** Cada uno de los elementos mas pequeños de los que se compone el catálogo de objetos, cuenta con un nombre y unos atributos, y es el elemento que se consulta a través de las sentencias SELECT que se emiten desde el mediador.
- **Atributo de los Metadatos:** Cada uno de los atributos que caracterizan un objeto de los Metadatos, cuenta con un nombre y un tipo, estos elementos se asocian con los campos de las vistas y con los parámetros de salida de los procedimientos almacenados de cada fuente de datos.
- **Elemento de la Fuente de Datos:** Objetos dentro de la base de datos que pueden ser asociados con los objetos de los Metadatos. Dichos elementos son las vistas y los procedimientos almacenados.
- **Descriptor Local:** Documento que describe los objetos que pueden ser consultados en cada fuente de datos local.
- **Metadatos Locales:** Copia local de los Metadatos en cada una de las fuentes de datos local.
- **Archivo de Configuración:** Para el correcto funcionamiento de la plataforma en cualquier sistema operativo, se requiere la configuración de ciertos parámetros en un archivo que contenga las rutas y variables de entorno.

5.3.1 Módulo Wrapper

5.3.1.1 Árbol de Funciones

A continuación se presentan las funciones que el módulo wrapper debe realizar:

- a. Configurar parámetros de conexión a la Fuente de Datos.
- b. Publicación de la Información local.
- c. Creación del Documento Descriptor.
- d. Gestionar usuarios del Sistema.
- e. Actualizar Metadatos.
- f. Validar Usuario.
- g. Resolver Consultas.
- h. Actualizar Descriptor Local.
- i. Visualizar Metadatos

Los usuarios que interactúan con el módulo wrapper son:

Administrador F. D: (Administrador de la fuente de datos) encargado de la configuración y mantenimiento de los sistemas de información integrados al sistema (Fuentes de datos). Especialmente se encarga de las labores de publicación y asignación de permisos a la información de su sistema.

Mediador: Sistema que realiza las peticiones de consultas al Wrapper, y de actualización del descriptor.

5.3.1.2 Descripción de Casos de Uso

A continuación se presenta una breve descripción de los casos de uso de acuerdo a las funciones identificadas (Ver Figura 5.3). Si desea ampliar ésta descripción, por favor remítase a [GIT04a].

Validar Administrador F.D

Actor: Administrador F. D.

Tipo: Primario

Descripción: Cuando el Usuario ingresa por primera ocasión al módulo wrapper, éste, ofrece una interfaz de ingreso de nombre de usuario y contraseña. Después de realizar el ingreso de los datos, éstos son validados en el sistema. Posteriormente se emite un mensaje de error en caso de existir problemas el proceso de validación, de lo contrario se conduce al usuario al menú principal de administración del módulo wrapper. Referencia cruzada: **d**.

Gestionar Usuarios

Actor: Administrador F. D.

Tipo: Primario

Descripción: Cuando el administrador selecciona la opción **Usuarios** desde el menú principal, se presenta un listado con los usuarios registrados en el módulo. Desde la interfaz de gestión de usuarios se puede agregar, eliminar o editar las características de los usuarios, asignado también, permisos sobre los datos que se consultan desde el módulo mediador. Los usuarios registrados NO cuentan con privilegios de Administrador

F. D., y sólo están autorizados para realizar consultas a través de la plataforma.
Referencia cruzada: **d**.

Publicar Información

Actor: Administrador Fuente de datos

Tipo: Primario

Descripción: Cuando el usuario selecciona la opción **Publicación** desde el menú principal, aparece una interfaz desde la que se puede observar los metadatos de la plataforma, en una interfaz explorable, en la que aparecen agrupados por niveles los temas que existen en el módulo wrapper y los objetos registrados en la plataforma. El usuario puede navegar en la interfaz hasta seleccionar el objeto que desee publicar a través de los metadatos de la plataforma y asociarlo con los elementos de su fuente de datos. De igual manera, una vez seleccionado un elemento de la fuente de datos a asociar con un objeto de los metadatos, se pueden seleccionar los Campos (si se asoció con una vista) o los Parámetros de Salida (Si se asoció con un procedimiento Almacenado). Una representación gráfica de estas funcionalidades se puede observar en las Figuras de la sección 6.2. Referencia cruzada: **b**.

Crear Descriptor Local

Actor: Administrador F. D.

Tipo: Primario

Descripción:

Al activar la opción de crear descriptor local, el administrador F.D puede sobrescribir el descriptor local existente o puede crear uno nuevo con las características actualizadas de la fuente de datos que esté analizando. Dicho documento será utilizado por el Mediador para conocer los elementos publicados en las Fuentes de Datos participantes. En la sección 6.3 se puede observar las representaciones gráfica que ilustran este proceso.
Referencia cruzada: **c**.

Actualizar Metadatos Locales

Actor: Administrador F. D.

Tipo: Primario

Descripción: Cuando el Administrador F. D. selecciona la opción **Actualizar Metadatos Locales** desde el menú Descriptor Local, aparece una ventana de confirmación para preguntar si se desea reemplazar los Metadatos Locales actuales por unos más actualizados. Este proceso es necesario cuando en la plataforma se hayan realizado modificaciones a los metadatos. Luego de realizar el proceso se mostrará un mensaje indicando el resultado de la operación o el indicador de error en caso de surgir alguno.
Referencia cruzada: **e**.

Configurar Conexión SIG

Actor: Administrador F. D.

Tipo: Primario

Descripción: Cuando el Administrador F. D. selecciona la opción **Parámetros** desde el Menú Principal, aparece una interfaz donde el Administrador F. D. puede ingresar los datos de conexión a la fuente de datos, URL, nombre de usuario y contraseña. Referencia cruzada: **a**.

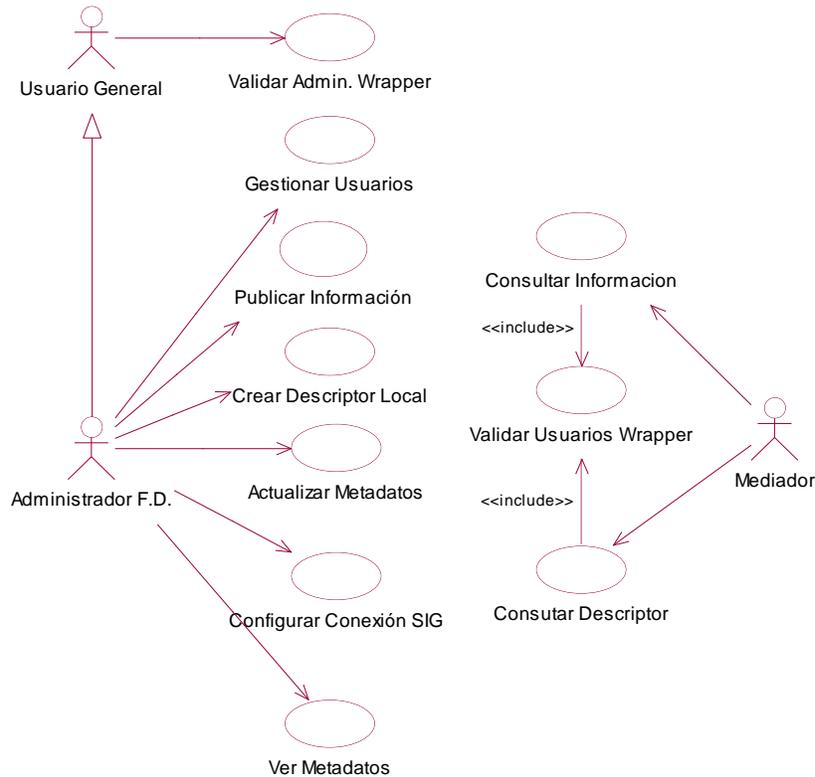


Figura 5.3 – Diagrama de Casos d+e Uso Módulo Wrapper

Ver Metadatos

Actor: Administrador F. D.

Tipo: Secundario

Descripción: Cuando el usuario selecciona la opción **Ver Catálogo** desde el Menú Principal, se muestra la ruta en la que se encuentra localizada una versión con formato HTML de los metadatos de la plataforma, basados en el catálogo de objetos del ICDE. De ésta manera el usuario puede visualizar el documento en un formato amigable. Referencia cruzada: i.

Validar Permisos Wrapper

Actor: Mediador

Tipo: Primario

Descripción: Permitir la validación del Mediador en el Sistema Wrapper. Referencia cruzada: f

Consultar Descriptor

Actor: Mediador

Tipo: Primario

Descripción: Una vez validado el Mediador, puede solicitar una copia del descriptor local que se creó en disco por parte del Administrador F. D., con el objetivo de actualizar la información existente en dicha fuente de datos en los registros del mediador. Referencia cruzada: h.

Consultar Información

Actor: Mediador

Tipo: Primario

Descripción: Una vez validado el Mediador, éste puede enviar peticiones de información en forma de consultas SQL (SELECT) al módulo wrapper, en donde los objetos que se consultan son los descritos en los metadatos. El wrapper realiza la traducción de la consulta del mediador a una consulta local y devuelve los resultados en un formato compatible con los datos del mediador. Dicho formato incluye los datos de respuesta, los Metadatos del resultado y un listado de errores presentados en la transacción. Referencia cruzada: g.

Si analizamos la descripción general de los casos de uso del módulo wrapper, podemos identificar que el caso de uso de mayor relevancia es consultar Información. Por tal motivo, a continuación se detallará su descripción. Si desea ampliar la representación de los casos de uso restantes, remitirse a [GIT04b].

Descripción Detallada Caso de Uso Consultar Información

Actores	Mediador
Propósito	Permitir al Mediador consultar la información de la Fuente de Datos local.
Resumen	Una vez validado en el sistema el Mediador puede enviar peticiones de información en forma de consultas SQL (SELECT) al sistema, en donde los objetos que se consultan son los descritos en el Esquema Global. El sistema realiza la traducción de la consulta a una consulta local y devuelve los resultados en un formato compatible con los datos del mediador. Dicho formato incluye los datos de respuesta, los Metadatos del resultado y un listado de errores presentados en la transacción
Tipo	Primario
Referencias Cruzadas	Funciones: g

Tabla 5.1 – Descripción caso de uso consultar información

Precondiciones

- El Mediador debe haber sido validado en el sistema

Flujo Principal

- El Mediador envía una sentencia SQL al Sistema Wrapper.
- El Sistema intenta conectarse a la B.D.(E1).
- El Sistema revisa la validez de la consulta SQL.(E2).
- El Sistema revisa la existencia de los objetos requeridos en la fuente de datos.
- El Sistema convierte la sentencia a una que pueda resolver la fuente de datos.
- El Sistema intenta conectarse a la Fuente de Datos.(E3)
- El Sistema consulta la información (E4).
- El Sistema da formato a la respuesta y la retorna al Mediador.

Flujos de Excepción

- E1: El Sistema informa al Mediador que no puede resolver su consulta, debido a problemas de conexión con la B.D.
- E2: El Sistema informa al Mediador que la consulta SQL no es válida
- E3: El Sistema informa al Mediador que no puede resolver su consulta, debido a problemas de conexión con la Fuente de Datos.
- E4: El Sistema informa al Mediador que no puede resolver su consulta, debido a problemas de ejecución de la consulta en la Fuente de Datos.

Diagrama de Clases

A continuación se presenta el diagrama de clases del Caso de Uso y una descripción general de cada una de las clases (Figura 5.4).

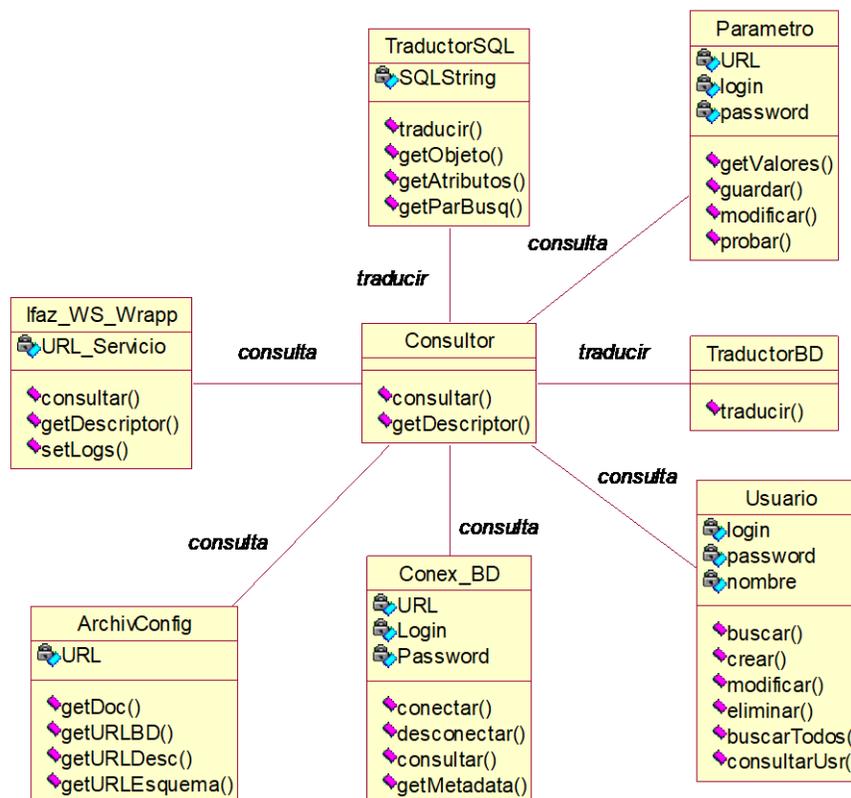


Figura 5.4 – Diagrama de clase C.U. Consultar Información

La Clase **IfazWSWrapp** provee una interfaz para que el mediador se conecte a cada uno de los wrappers, El mediador solo necesita saber la ubicación URL del servicio y el nombre de los métodos.

La Clase **Consultor** es la clase que gestiona todo el proceso de resolución de la consulta, para tal fin utiliza en primer lugar la clase **TraductorSQL**, la cual se encarga de chequear la validez de la sentencia SQL y de extraer sus parámetros. La clase **ArchivConfig**

permite conocer la ubicación de la base de datos del Wrapper, en esta base de datos el consultor busca los datos del usuario (nombre de usuario y contraseña) por medio del Bean **Usuario** (que encapsula todos los datos y la funcionalidad de la entidad Usuario en el sistema).

Por otro lado para la resolución de la consulta, es necesario revisar que los parámetros consultados se encuentren publicados en la Fuente de Datos. Este proceso lo realiza la clase **TraductorBD**, que crea una sentencia aceptable por la Fuente de Datos local.

Por último, la clase **Consultor** obtiene los parámetros para conectarse a la fuente de datos desde el bean **Parámetro**. Posteriormente la clase **consultor** realiza la consulta y compone la respuesta para devolverla al Mediador.

La clase **consultor** también rastrea todos los errores (campos no encontrados o campos sin información) presentados durante el proceso y los envía de vuelta al mediador, junto con los datos consultados y los metadatos de la consulta (nombre de los campos obtenidos y No de registros).

5.3.2 Módulo Mediador

5.3.2.1 Árbol de funciones

A continuación se presentan las funciones que el módulo mediador debe realizar:

- a. Actualizar parámetros de acceso a las Fuentes de Datos
- b. Actualizar el descriptor de cada fuente de datos.
- c. Gestionar Usuarios.
- d. Gestionar Metadatos
- e. Crear Documento de los Metadatos
- f. Ver objetos publicados
- g. Gestionar Tipos de datos
- h. Gestionar Cartografía
- i. Consultar Información

Los usuarios que interactúan con el módulo mediador son:

Administrador Mediador: (Administrador del Mediador) Usuario con privilegios para realizar la configuración de la Plataforma de Integración de Múltiples Fuentes de Datos Descriptivos y Espaciales en un Sistema de Información Ambiental.

Usuario Mediador: Entidad o Sistema que realiza consultas de información a la plataforma, a través de la interfaz destinada para la prestación de dicho servicio.

Wrapper: Sistema encargado del encapsulamiento de cada sistema participante

5.3.2.2 Descripción de Casos de Uso

A continuación se presenta una breve descripción de los casos de uso de acuerdo a las funciones identificadas (Ver Figura 5.5). Si desea ampliar ésta descripción, por favor remítase a [GIT04a].

Validar Administrador Med

Actor: Administrador Mediador

Tipo: Primario

Descripción: Cuando el Usuario ingresa por primera ocasión al módulo Mediador, éste, ofrece una interfaz de ingreso de nombre de usuario y contraseña. Después de realizar el ingreso de los datos, éstos son validados en el sistema. Posteriormente se emite un mensaje de error en caso de existir problemas el proceso de validación, de lo contrario se conduce al usuario al menú principal de administración del módulo mediador. Referencia cruzada: **c**.

Gestionar Tipo de Datos

Actor: Administrador Mediador

Tipo: Primario

Descripción: El Administrador del sistema tiene la opción de crear, editar y eliminar los tipos de datos que describen los Atributos de los Objetos de los metadatos. Referencia cruzada: **g**.

Gestionar Usuarios Mediador

Actor: Administrador Mediador

Tipo: Primario

Descripción: Una vez validado dentro en el módulo mediador, el Administrador cuenta con una interfaz para agregar, eliminar o modificar información de los usuarios que pueden consultar información a través de la plataforma. (Los usuarios del sistema Mediador que se gestionan aquí NO tienen privilegios de Administrador, solo de consulta). Referencia cruzada: **c**.

Gestionar Metadatos

Actor: Administrador Mediador

Tipo: Primario

Descripción: El Administrador cuenta con una interfaz desde la cual puede gestionar los elementos de los metadatos: Los Temas y sus Jerarquías, Los Objetos dentro de los Temas, Los atributos que integran los objetos y sus tipos de datos. Todo esto a través de una interfaz navegable y fácil de utilizar. Referencia cruzada: **d**.

Ver Objetos Publicados

Actor: Administrador Mediador

Tipo: Primario

Descripción: El Administrador tiene la posibilidad de observar todos los objetos que han sido publicados en cada una de las fuentes de datos, para conocer de manera general la información con que cuenta la plataforma. Referencia cruzada: **b, f**.

Gestionar F. D.

Actor: Administrador Mediador

Tipo: Primario

Descripción: Una vez validado dentro del sistema, el Administrador cuenta con una interfaz donde puede agregar, eliminar o modificar información de las fuentes de datos componentes, lo que le permite entre otras opciones, configurar los parámetros de acceso a la Fuente de datos (URL, Login, Password), y registrar la información que se encuentra publicada en cada fuente de datos a través de la importación del documento descriptor. Referencia cruzada: **a, b, i**.

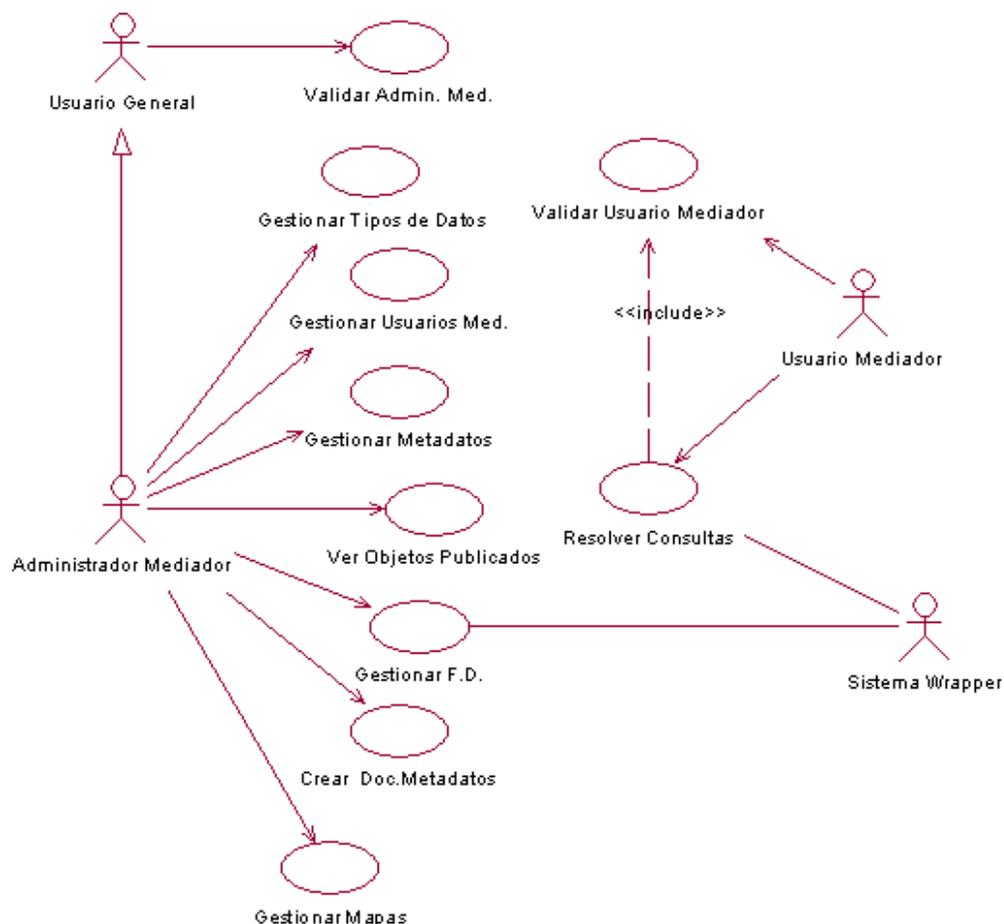


Figura 5.5 – Diagrama de Casos de Uso Módulo Mediator

Crear Doc. Metadatos

Actor: Administrador Mediator

Tipo: Primario

Descripción: Una vez que se han creado los elementos de los metadatos a través de la interfaz, es necesario consolidarlo para su publicación y distribución en un documento, que puede ser creado en cualquier momento por parte del administrador. Referencia cruzada: **d, e**.

Gestionar Mapas

Actor: Administrador Mediator

Tipo: Primario

Descripción: El Usuario del Mediator puede requerir que se consolide su consulta con información cartográfica, el Administrador puede agregar, eliminar o modificar los datos

de los archivos de mapas para que puedan ser utilizados en las consultas. Referencia cruzada: **h**.

Validar Usuario Mediador

Actor: Usuario Mediador

Tipo: Primario

Descripción: El usuario realiza el proceso de Autenticación (login y password) en el modulo mediador con el fin de ingresar a la sección que realiza la petición de los datos. Referencia cruzada: **c**.

Resolver Consulta

Actor: Usuario Mediador

Tipo: Primario

Descripción: Después de validarse, el Usuario del Mediador puede enviar a través de la plataforma las consultas de datos. El módulo mediador se encarga de validar los permisos del Usuario del Mediador en cada fuente, y posteriormente revisa en que posibles fuentes componente se encuentran los datos que se están consultando. A continuación, se envían las peticiones a cada fuente de datos. Finalmente, se revisan los datos obtenidos en las fuentes participantes y se compone un mensaje de respuesta para el Cliente Mediador. Referencia cruzada: **i, h, c, a, b, d**.

Si analizamos la descripción general de los casos de uso del módulo mediador, podemos identificar que el caso de uso de mayor relevancia es resolver consulta. Por tal motivo, a continuación se detallará su descripción. Si desea ampliar la representación de los casos de uso restantes, remitirse a [GIT04b].

Descripción Detallada Caso de Uso Resolver Consulta

Actores	Usuario Mediador
Propósito	Resolver las consultas de datos hechas al mediador con información contenida en las diversas Fuentes de Datos Adscritas al Sistema
Resumen	<ul style="list-style-type: none"> - El Sistema recibe una consulta (Compuesta de una sentencia Select, un login y un password). - El Sistema revisa en el Esquema Global que los campos solicitados correspondan a atributos de los Objetos del Catálogo. - El Sistema revisa que los Objetos sobre los que se requiere información se encuentren publicados. - El Sistema crea una sentencia SQL para cada una de las fuentes de Datos que alojan los datos que se requieren. - El Sistema consulta los datos de cada Fuente de Datos. - El Sistema compone una respuesta con los resultados obtenidos y agrega el Metadata del resultado (Nombre de los campos obtenidos) además de un listado de errores presentados. - El Sistema retorna dicha respuesta al Sistema que ejecutó la consulta.
Tipo	Primario
Referencias Cruzadas	Funciones: i, h, c, a, b, d.

Tabla 5.2 – Descripción caso de uso resolver consulta

Precondiciones

- El Servicio Web del Sistema Wrapper debe estar disponible.
- El Administrador Mediador debe haber registrado al menos una Fuente de Datos.
- El Administrador Mediador debe haber actualizado el Descriptor de al menos una de las Fuentes Registradas.

Flujo Principal

- El Sistema recibe la petición de una consulta compuesta por una sentencia SELECT SQL, acompañada de un login y un password.
- El Sistema verifica que el usuario sea un usuario válido (E1).
- El Sistema verifica que la sentencia sea válida en cuanto a su formato (E2).
- El Sistema verifica que el objeto y los atributos consultados se encuentren publicados (E3).
- El Sistema crea una consulta para cada una de las fuentes.
- El Sistema realiza la consulta a cada una de las fuentes (E4).
- El Sistema compone la respuesta en una estructura con los resultados obtenidos agregando el nombre de los campos, los valores de los campos obtenidos y un listado de errores.
- El sistema retorna la estructura al sistema que ejecutó la consulta.

Flujos de Excepción

- E1: El Sistema agrega el error a la respuesta : Usuario no válido
- E2: El Sistema agrega el error a la respuesta: Sentencia No válida
- E3: El Sistema agrega el error a la respuesta: Ninguno de los Atributos se encuentran publicados
- E4: El Sistema agrega el error a la respuesta: Ninguna de las fuentes se encuentran disponible por el momento

Diagrama de Clases

La Figura 5.6 muestra las clases más representativas del caso de Uso. La clase **lfaz_WS_Mediador** representa la interfaz del servicio de consulta del Mediador; está compuesta por el método **consultar()** cuyo objetivo principal es invocar la Clase **Consultor**, que es la encargada de todo el proceso de resolver la consulta hecha a través del mediador. La clase **Consultor** en primer lugar revisa la validez de la sentencia SQL a través de la clase **TraductorSQL**, la cual obtiene los elementos de la consulta. Estos elementos son revisados por la clase **Consultor**, la cual examina por medio de los beans **Objeto** y **Atributo** si estos elementos están publicados en el mediador, verificando su ubicación en cada fuente de datos.

Para cada fuente de datos, la clase **Consultor** crea un objeto **Consulta**, el cual tiene la ubicación del servicio proporcionado por el módulo Wrapper y los elementos a consultar. El método consultar de la clase **Consulta** obtiene los datos de cada fuente de datos, y los entrega a la clase **Consultor**, ésta última, revisa los datos, los errores y los metadatos, para posteriormente componer la respuesta.

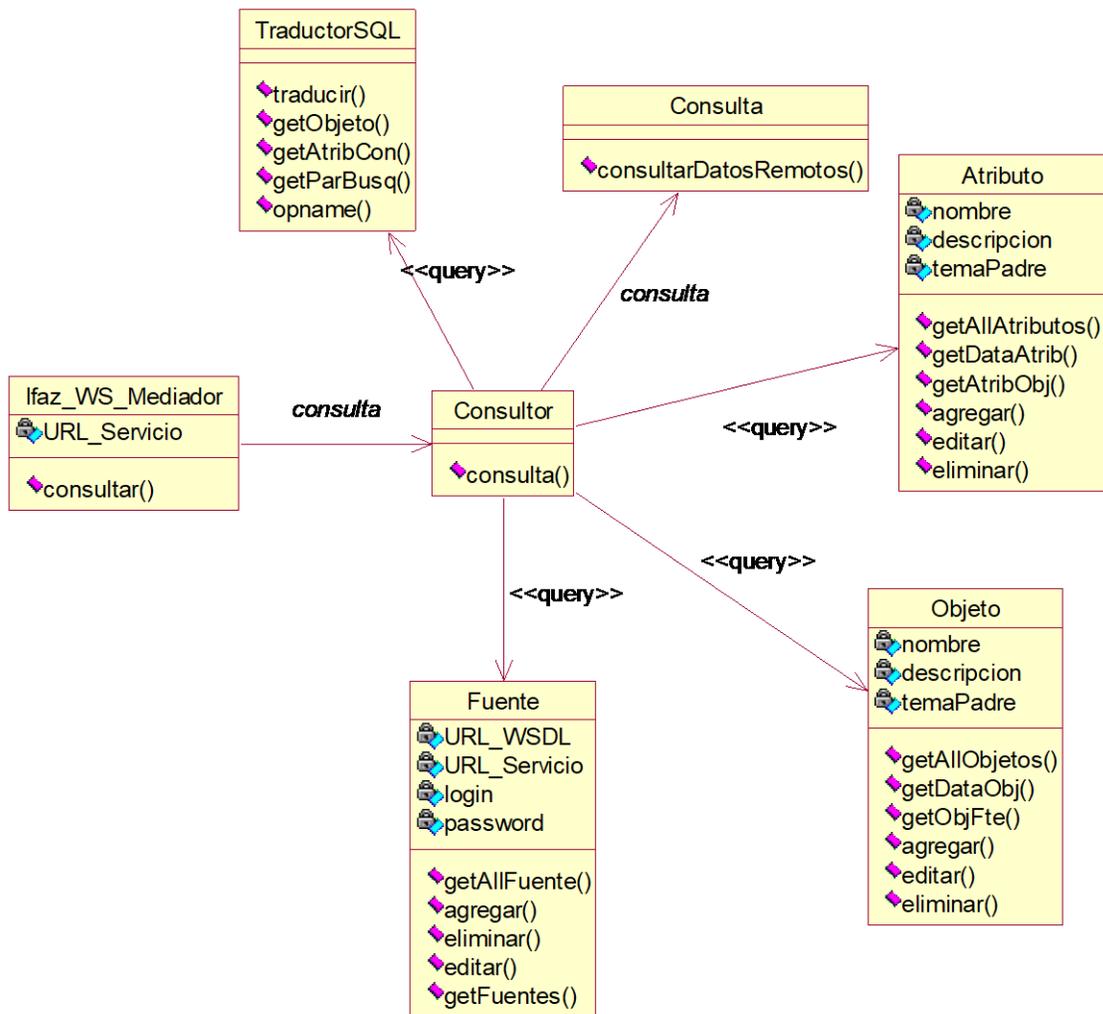


Figura 5.6 – Diagrama de clase C.U. Resolver Consulta

5.4 Arquitectura física

El lenguaje seleccionado para la implementación de la Plataforma de Integración de Múltiples Fuentes de Datos Descriptivos y Espaciales en un Sistema de Información Ambiental fue Java, dada sus potencialidades desde el punto de vista funcional para la construcción, mantenimiento y adaptación de aplicaciones complejas, que impliquen interacción con modelos de datos, transformaciones, publicación de servicios y distribución de funcionalidades. Además, está diseñado para operar en ambientes de red e Internet e interactuar con bases de datos, y en especial, para el desarrollo de aplicaciones en el campo de los servicios web. El JDK (Java Development Kit) utilizado es J2SDK1.4.1_02, el cual está compuesto por: el API de Java, el JRE (JVM), compilador de Java y otras funcionalidades definidas por SUN, que son utilizadas para generar programas básicos escritos en este lenguaje.

Bajo este contexto, las estructuras software implementadas fueron [GIT04c]:

- Las interfaces de usuario fueron implementadas utilizando JSP (Java Server Pages).
- La lógica del sistema fue desarrollada utilizando servlets, beans y clases Java.
- Todo el manejo de Mapas se realizó utilizando el formato ShapeFile de ESRI [ESR98].
- El manejo de los archivos ShapeFile a través de Internet: creación, manipulación y visualización se realizó utilizando el API Java Geotools [GTO03], la cual es un conjunto de herramientas de java de distribución libre, que sirve para desarrollar mapas geográficos interactivos.
- El manejo de Servicios Web se realizó utilizando el API Axis 1.1 [AXI03]. Apache Axis es una implementación de la especificación SOAP. Es usada para la construcción y el despliegue de diversos sistemas incluyendo Servicios Web.
- El motor de Bases de Datos seleccionado fue Firebird 1.5 [FIR03], el cual además de ser gratuito, cuenta con capacidades de rendimiento y manejo de Procedimientos Almacenados, Triggers y Vistas.
- El servidor de aplicaciones seleccionado fue Apache Tomcat 5.0.27.
- El Servidor Web seleccionado fue Apache Web Server 2.0.49.

A continuación, se presenta las capas de aplicación, lógica de mediación y software, para los módulos Wrapper y Mediador. Posteriormente, se tratará la distribución de componentes por cada nodo que conforma la arquitectura física de la plataforma.

5.4.1 Descripción de las Capas de Aplicación, Lógica de Mediación y Software

5.4.1.1 Módulo Wrapper

En la Figura 5.7 se muestran la interacción de los subsistemas representados en los paquetes generales de la aplicación y los productos comerciales seleccionados para la implementación del módulo Wrapper.

La capa de aplicación se compone de un subsistema de Administración (Paquete Administración Wrapper) y un subsistema que representa el servicio de Consulta (Paquete Servicio Consulta Wrapper). El primero permite tanto la configuración de los parámetros de funcionamiento del Wrapper, como el proceso de Publicación de Información de cada fuente de datos. Por otro lado, El paquete de "Servicio Consulta Wrapper" se encarga de la resolución de consultas provenientes del Mediador.

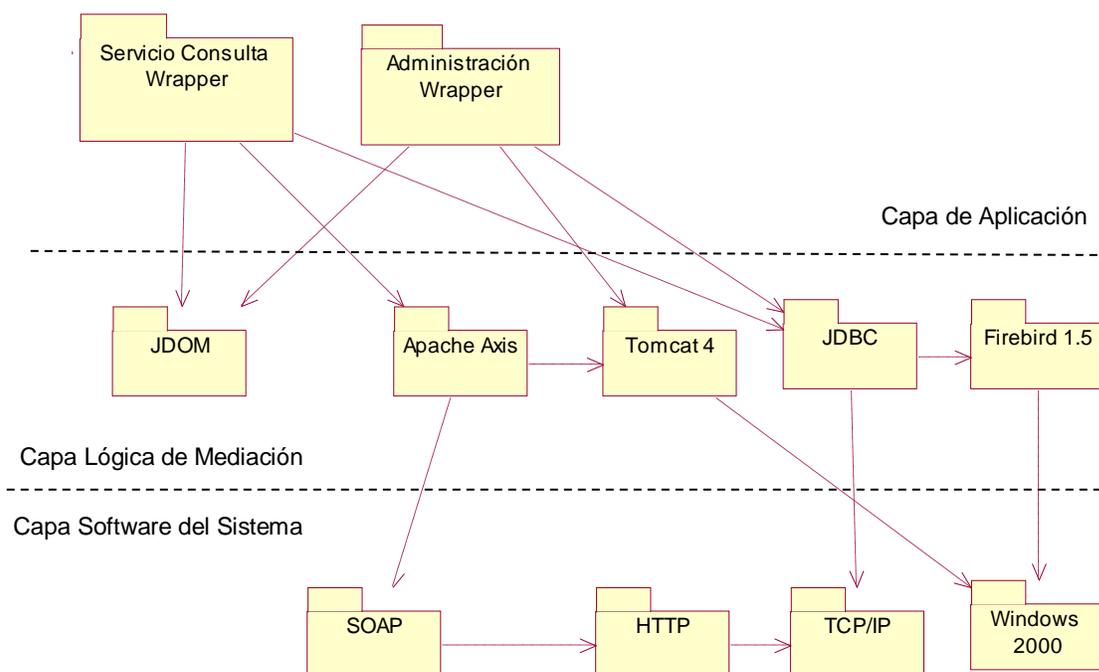


Figura 5.7 – Diagrama capas de aplicación, lógica de mediación y software módulo wrapper

La capa lógica de Mediación esta compuesta de los siguientes paquetes:

- El paquete JDOM, que permite la creación, lectura y modificación de documentos XML generados a partir de las fuentes de datos locales (e.g: descriptores).
- El paquete Apache Axis, que facilita la creación y publicación del módulo wrapper como Servicios Web.
- El paquete Tomcat 4 representa al servidor web y de aplicaciones, necesario para que el módulo wrapper este desplegado constantemente en la web y proporcione facilidades de invocación a la información residente en las fuentes de datos locales a través de él.
- El paquete JDBC (Java Data Base Connectivity) proporciona una interfaz estándar entre los subsistemas de capas superiores con el servidor de base de datos.
- Finalmente el paquete que representa el servidor de bases de datos es Firebird 1.5.

En la capa de Software del Sistema, se encuentran los paquetes TCP/IP, HTTP, SOAP y Windows 2000.

- HTTP (Hypertext Transfer Protocol) como protocolo cliente-servidor que articula los intercambios de información entre los clientes Web y los servidores HTTP.

- TCP/IP (Transfer Communication Protocol / Internet Protocol), el cual permite conectar a diversas aplicaciones a través de Internet.
- SOAP (Simple Object Access Protocol) es el protocolo base de los Servicios Web que permite al mediador invocar métodos de objetos residentes en el wrapper, así como enviar las respuestas, a través de un mensaje XML transportado en HTTP.
- Web Services Description Language (WSDL) es el estándar escogido para la descripción de servicios Web (wrapper) basado en XML.
- En la implementación actual, el Sistema Wrapper se encuentra sobre la Plataforma Windows 2000. Sin embargo, ésta no constituye una restricción del sistema, pues dada la implementación en Java, puede montarse en cualquier otro Sistema Operativo.

5.4.1.2 Módulo Mediador

En siguiente figura se muestran la interacción de los subsistemas representados en los paquetes generales de la aplicación y los productos comerciales seleccionados para la implementación del Sistema Mediador.

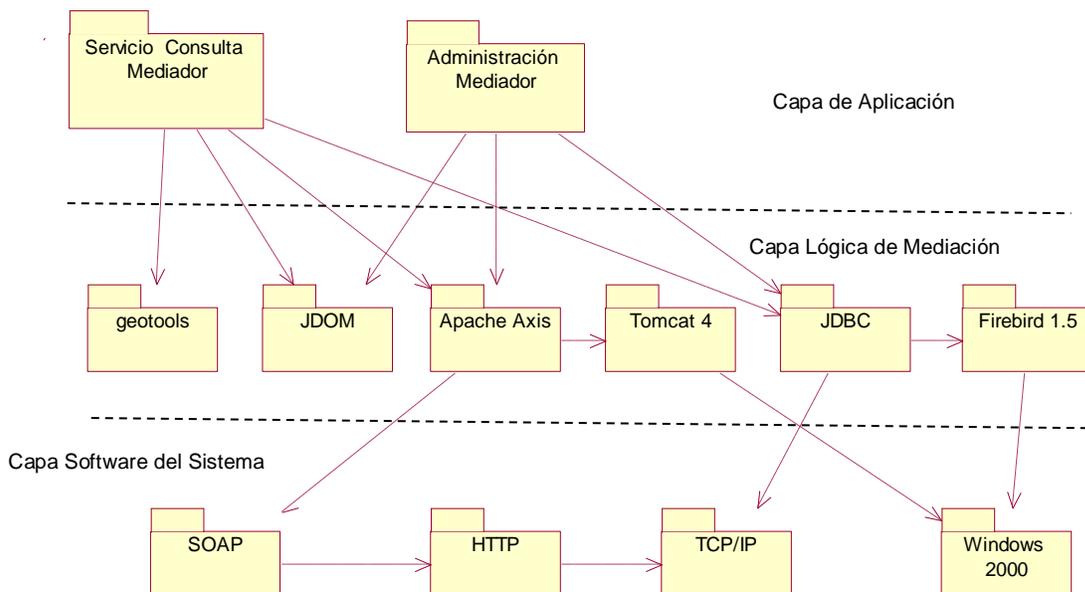


Figura 5.8 – Diagrama capas de aplicación, lógica de mediación y software módulo mediador

La capa de aplicación se compone de una parte de Administración (Paquete Administración Mediador) y otra de Servicio de Consulta (Paquete Servicio Consulta Mediador). El Paquete de Administración Mediador permite la configuración de todos los parámetros para el correcto funcionamiento de la plataforma (Permisos, Fuentes de Datos, Mapas).

El paquete de Servicio de Consulta, permite la resolución de consultas emitidas por parte de los usuarios del Sistema Mediador

La capa lógica de Mediación integra los siguientes paquetes:

- El paquete Geotools, que provee funcionalidades para la creación y actualización de los mapas.
- El paquete JDOM, que permite la creación, lectura y modificación de documentos XML generados a partir de los metadatos.
- El paquete Apache Axis, que facilita la creación y publicación del módulo mediador como Servicios Web.
- El paquete Tomcat 4 representa al servidor web y de aplicaciones, necesario para que el módulo mediador este desplegado constantemente en la web y proporcione facilidades al usuario mediador de realizar invocaciones a la información residente en las fuentes de datos locales.
- El paquete JDBC (Java Data Base Connectivity) proporciona una interfaz estándar entre los subsistemas de capas superiores con el servidor de base de datos.
- Finalmente el paquete que representa el servidor de bases de datos es Firebird 1.5.

En la capa de Software del Sistema, se encuentran los paquetes TCP/IP, HTTP, SOAP y Windows 2000.

- HTTP (Hypertext Transfer Protocol) como protocolo cliente-servidor que articula los intercambios de información entre los clientes Web y los servidores HTTP.
- TCP/IP (Transfer Communication Protocol / Internet Protocol), el cual permite conectar a diversas aplicaciones a través de Internet.
- SOAP (Simple Object Access Protocol) es el protocolo base de los Servicios Web que permite al mediador invocar métodos de objetos residentes en el wrapper, así como enviar las respuestas, a través de un mensaje XML transportado en HTTP.
- Web Services Description Language (WSDL) es el estándar escogido para la descripción de servicios Web (mediador) basado en XML.
- En la implementación actual, el Sistema Wrapper se encuentra sobre la Plataforma Windows 2000. Sin embargo, ésta no constituye una restricción del sistema, pues dada la implementación en Java, puede montarse en cualquier otro Sistema Operativo.

5.4.2 Diagrama de Implantación

En la figura 5.9, se muestra la distribución de componentes por cada nodo que conforman la arquitectura física de la plataforma. A continuación se hará la descripción de la interacción entre los módulos de Visualización, Mediador y Wrapper. Es necesario aclarar, que únicamente se muestra la interacción del Mediador con un solo Módulo Wrapper, mientras en la implementación se cuenta con varios de estos módulos para encapsular la información de las diversas fuentes de datos integradas a través de la plataforma.

La figura 5.9, presenta los dos tipos de clientes que pueden obtener información haciendo uso de la plataforma: el primer cliente lo hace a través del componente **Browser** y el segundo por medio de una aplicación consumidora de servicios Web (**Cliente de Servicios Web**). De ésta manera, un usuario que accede a la plataforma haciendo uso de un browser requiere de un módulo intermedio de Visualización, que realiza las consultas al Mediador utilizando SOAP/HTTP, y posteriormente el componente **Visualizador Consultas** estructura los resultados en páginas HTML para su presentación final.

Servidor de Aplicaciones de Visualización

Éste módulo es el encargado de la visualización de la información consultada al Mediador a través de un browser. En este nodo se encuentra el componente genérico **Visualizador Consultas** que se encarga de realizar las peticiones al Mediador (mas específicamente al componente **Consultor Mediador**) y generar una interfaz que publica los datos recuperados, asociándolos a las capas cartográficas que procesa el sistema. Las solicitudes al Mediador son realizadas a través de Servicios Web utilizando el protocolo SOAP / HTTP.

Servidor de Aplicaciones Mediador

En el Servidor de Aplicaciones Mediador, se ejecuta tanto la lógica de Administración como de resolución de consultas del módulo Mediador.

En este nodo, los componentes más importantes son:

- **Administración Mediador:** provee una interfaz Web para que el Administrador por medio de un browser, pueda configurar el funcionamiento de la plataforma.
- **Consultor Mediador:** Este componente recibe las peticiones de los clientes haciendo uso de una interfaz de Servicios Web. Las peticiones realizadas pueden originarse, tanto en el componente **visualizador de consultas** como en el **Consumidor Servicios Web**. El Consultor Mediador revisa en los registros, la validez de la consulta, la existencia de los datos consultados y la ubicación de los mismos en el **Servidor de BD Mediador**. Posteriormente, realiza las consultas a cada una de las fuentes de datos a través del **Servidor de Aplicaciones Wrapper** utilizando SOAP/HTTP, tal como se muestra en la figura 5.9. Finalmente compone la respuesta para retornarla al Cliente.

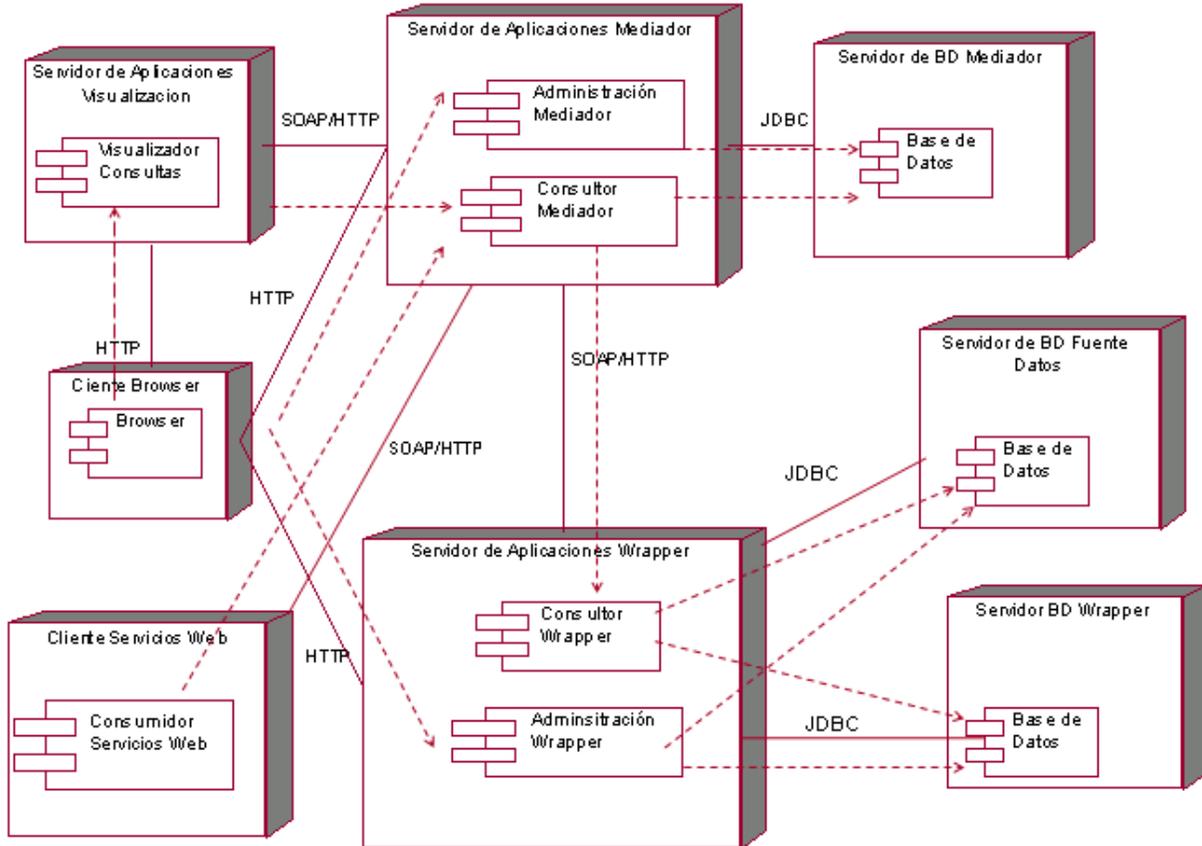


Figura 5.9 – Diagrama de Implantación

Servidor de Aplicaciones Wrapper

En el Servidor de Aplicaciones Wrapper, se ejecuta tanto la lógica de Administración como de resolución de consultas del módulo Wrapper.

En este nodo, los componentes más importantes se describen a continuación:

- **Administración Wrapper:** provee una interfaz Web para que el Administrador de la Fuente de Datos por medio de un browser pueda configurar los permisos y la información que se puede acceder desde el módulo mediador (Publicación).
- **Consultor Wrapper:** Este componente recibe las peticiones del Mediator y revisa en el **Servidor BD Wrapper** la validez de la consulta y la existencia de los datos consultados. Posteriormente transforma la consulta en una sentencia válida para el motor de BD local. Finalmente, emite una nueva consulta al motor de BD (utilizando JDBC como se muestra en la figura 5.1) y compone la respuesta para retornarla al Mediator.

Servidor de Base de Datos Mediador

Contiene la información persistente del Sistema Mediador: Usuarios Válidos, Administradores, Metadatos, Objetos Publicados y Fuentes de Datos.

Servidor de Base de Datos Wrapper

Contiene la información persistente del Sistema Wrapper: Usuarios Válidos, Administradores, Metadatos, Parámetros de Conexión a la Fuente de Datos y la información de publicación.

Servidor de Base de Datos Fuente de Datos

Contiene la información de la institución adscrita a la plataforma que es accedida por medio de la arquitectura propuesta. Los detalles de implementación dependen de cada entidad. El acceso se realiza utilizando el driver JDBC más apropiado para la tecnología empleada en la BD.

Cliente Browser

Éste nodo hace referencia al PC que permite el acceso a cualquiera de los clientes que hacen uso de un navegador: Wrapper (Administración Wrapper), Mediador (Administración Mediador) o de Visualización.

6. PROCESO DE INTEGRACIÓN DE LOS SISTEMAS DE GESTIÓN DE INFORMACIÓN

En éste capítulo estará consignado el proceso de integración de los sistemas de gestión de información (Hídrica y Geológica) a través de la Plataforma para la Integración de Múltiples Fuentes de Datos Descriptivos y Espaciales en un Sistema de Información Ambiental. Ésta descripción se abordará siguiendo el proceso de Publicación de Información, Creación de Descriptor Local y Configuración de Conexión. Además, se indicará los pasos a seguir para que un usuario Web pueda realizar una consulta haciendo uso de la plataforma.

6.1 Configurar Conexión SIG

Esta opción del **Módulo Wrapper**, permite realizar la configuración de los parámetros de acceso a la Fuente de Datos. Al hacer clic en la opción Parámetros del Menú Principal, se muestra la página de gestión de los Parámetros de Conexión. (Figura 6.1).

The screenshot shows a web browser window titled 'Sistema de Publicación de Información @ Ariadna - Microsoft Internet Explorer'. The address bar shows 'http://localhost/Wrapper/ParametroEditar.jsp'. The page header features the 'Plataforma de Integración' logo and 'ARIADNA - Sistema Integrado de Información Ambiental' and 'Universidad del Cauca'. The main content area is titled 'Parámetros Conexión Fuente' and contains a description: 'A continuación se muestran los parámetros que utilizará la plataforma para el acceso a la Fuente de datos Local.' Below this, there are three input fields: 'URL' with the value 'jdbc:interbase://172.16.130.130/C:/UB1C_BD.gdb', 'Login' with the value 'root', and 'Password' with masked characters. There are three buttons: 'Probar', 'Guardar', and 'Valores iniciales'. A 'Volver' link is at the bottom right. The footer contains 'Derechos Reservados © ARIADNA Sistema Integrado de Información Ambiental' and 'Intranet local'.

Figura 6.1 - Interfaz de Parámetros de Conexión

Desde aquí el usuario puede modificar los parámetros de conexión a las fuentes de datos, tales como:

URL: (Uniform Resource Locator) Ubicación de la fuente de datos, incluyendo el protocolo de conexión y su ubicación. Ej: **jdbc:interbase://172.16.130.130/C:/UBIC_BD.gdb**

Login: Identificador del Usuario de la base de datos

Password: Contraseña de acceso a la Base de datos

La interfaz de la figura 6.1, muestra 3 botones: Probar, Guardar y Valores Iniciales.

Al presionar Probar se prueba la conexión a la fuente de datos utilizando los parámetros introducidos en el momento. Posteriormente se muestra un mensaje con los resultados de la conexión (Figura 6.2).

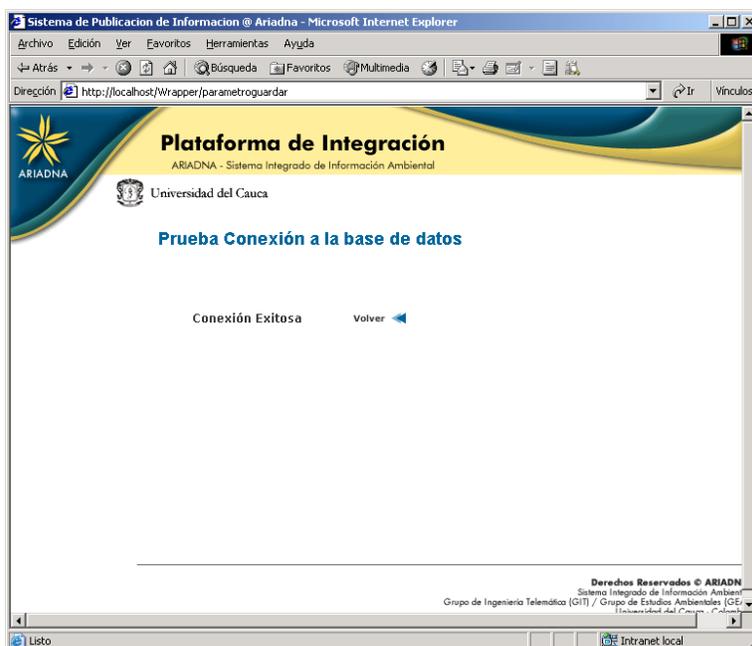


Figura 6.2 - Mensaje de resultado de la conexión a la Fuente de Datos

Al presionar **Guardar** en la Interfaz de Parámetros de Conexión (Figura 6.1), se guardan los valores que se están desplegando actualmente.

Al presionar **Valores iniciales**, se muestra nuevamente la Interfaz con los valores guardados en la base de datos actualmente.

6.2 Publicar Información

Esta opción del **Módulo Wrapper**, permite realizar la publicación de la información de la fuente de datos local en la Plataforma. Para tal efecto el usuario del Wrapper cuenta con una interfaz que le permite navegar por los Temas del Catálogo de Objetos (basado en el ICDE, Infraestructura Colombiana de Datos Espaciales), seleccionar el objeto a publicar y asociarlo con elementos de la fuente de datos, como vistas o procedimientos almacenados. El sistema permite además, asociar los atributos de un objeto con los

campos que representan los elementos de la fuente de datos asociada. La interfaz principal de Publicación se muestra al hacer clic en la opción Publicación del menú principal (Figura 6.3).

Para que la publicación se realice satisfactoriamente, se necesita realizar la Creación del Descriptor Local (Sección 6.3), inmediatamente después de la asociación descrita en esta sección. Posteriormente, el Administrador de la plataforma debe actualizar dicho Descriptor desde el Mediador.

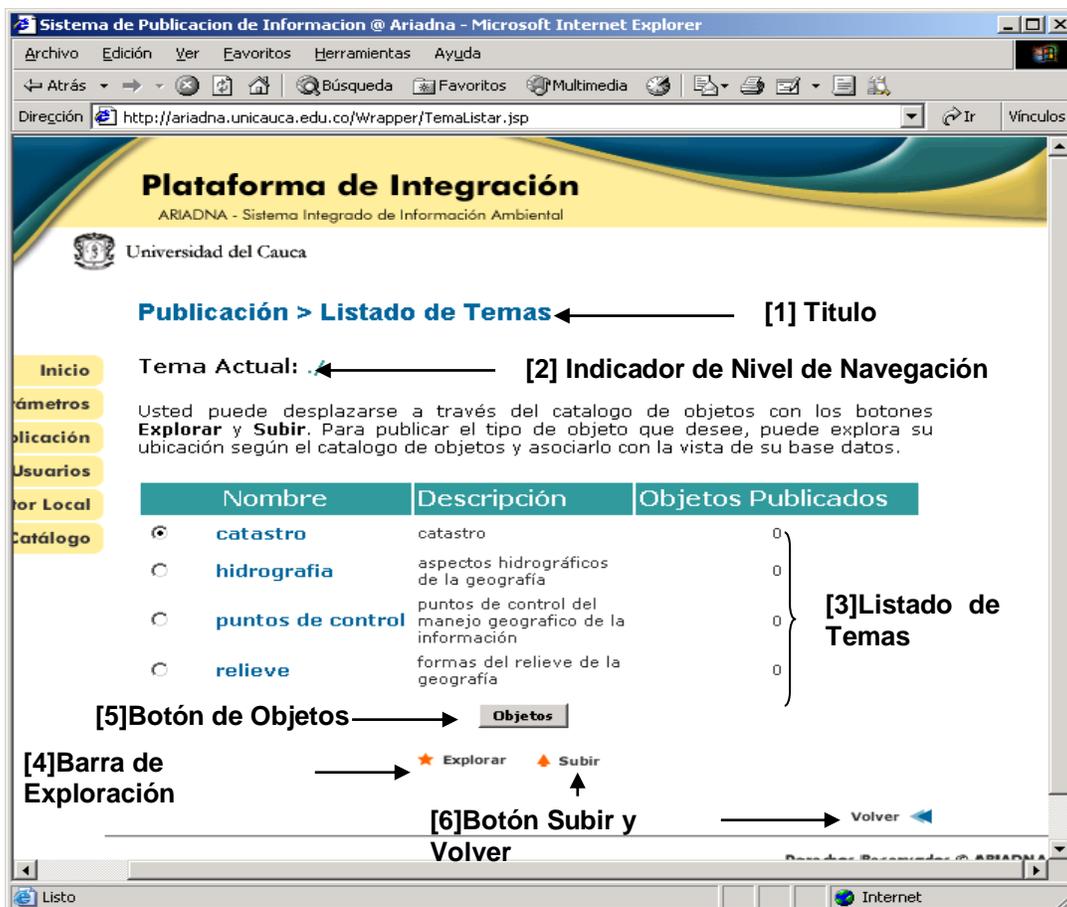


Figura 6.3 - Interfaz del Menú de Publicación

6.2.1 Navegación en los Metadatos

En la parte superior de la Figura 6.3, ítem [2], el sistema muestra el nivel de navegación, facilitando de esta manera la localización del administrador en los temas que contiene el catálogo de objetos. Inicialmente se despliega el carácter “/”, pero a medida que se va explorando dicho catálogo a través de los botones de navegación, se va modificando el nivel: “/Tema_1/Tema_2”.

En la parte inferior, se muestra el listado de temas del catálogo, ítem [3]. Dicho listado presenta la descripción y el número de objetos publicados por cada tema. Los temas son las clasificaciones generales del Catálogo de Objetos, mientras que los Objetos

constituyen los elementos básicos. Estos últimos son los ítems que pueden publicarse en la Plataforma.

EL ítem [4] muestra la barra de Exploración, con la cual el administrador puede moverse dentro de la jerarquía de temas del catálogo de objetos. El administrador tiene las opciones de **Subir** y **Volver** (ítem 6), si selecciona Subir se mostrarán los temas de nivel Superior, mientras que si se selecciona Volver se volverá a la página de inicio.

En todas las interfaces donde se encuentren Temas listados aparecerá el botón **Objetos**, el cual mostrará los objetos existentes dentro de cada mapa. (Figura 6.4)

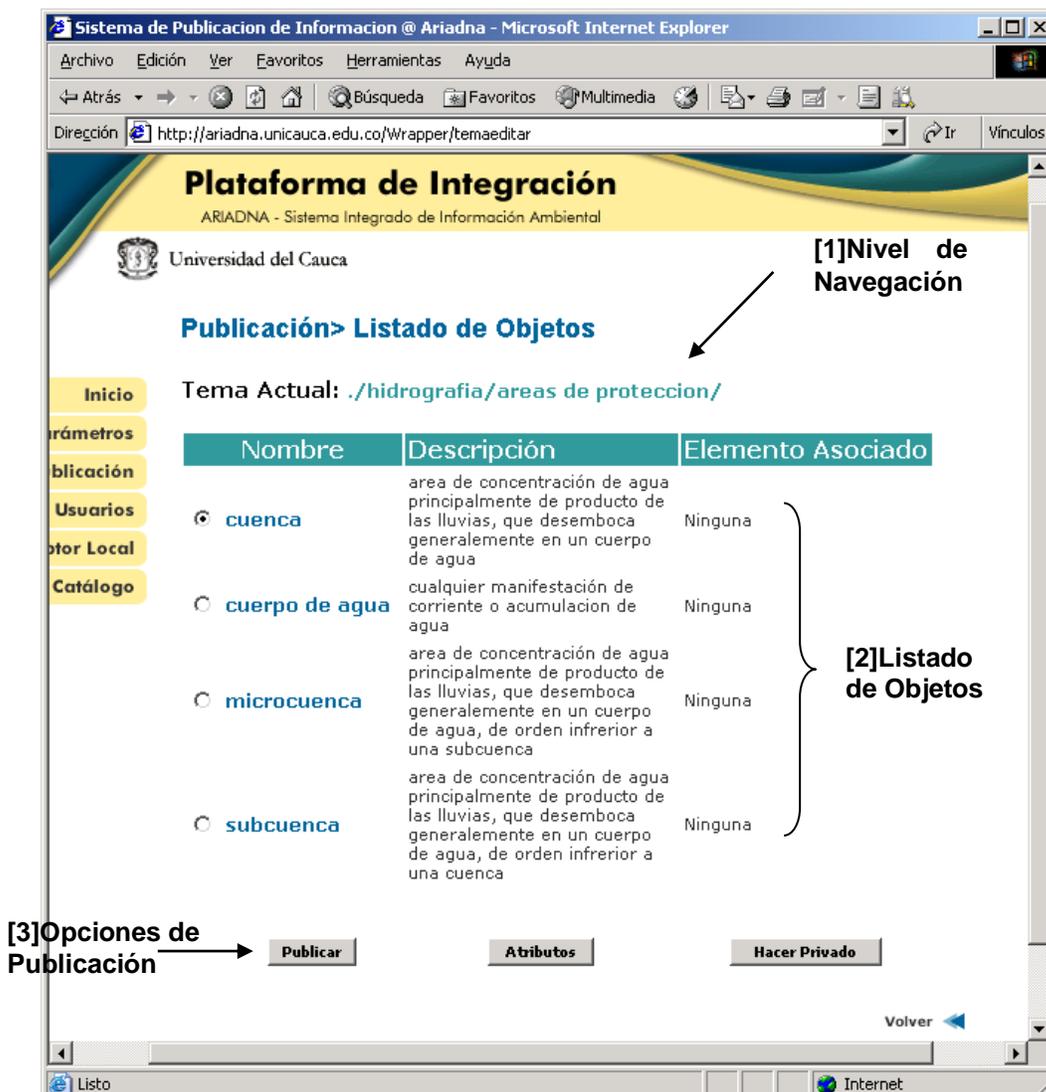


Figura 6.4 - Interfaz de Listado de Objetos

En la figura 6.4 ítem [2], se muestra el listado de los objetos existentes dentro del tema “Áreas de Protección” con su descripción y el elemento de la base de datos asociado. En esta interfaz se muestra también el Nivel de Navegación (ítem [1]) y tres opciones de

publicación que pueden ser seleccionadas: **Publicar, Atributos y Hacer Privado** (ítem [3]).

6.2.2 Publicación de Objetos

Al hacer clic en la opción **Hacer Privado** de la Figura 6.4, el objeto seleccionado pierde la asociación con el elemento vinculado. De esta manera el objeto no aparecerá publicado en el momento de generar el Descriptor.

Cuando el administrador selecciona la opción **Publicar** se muestra una interfaz que permite asociar los objetos del catálogo con los elementos de la base de datos (Vistas o Procedimientos).

Si en la base de datos existen vistas o procedimientos almacenados, éstos se listan como en la Figura 6.5.

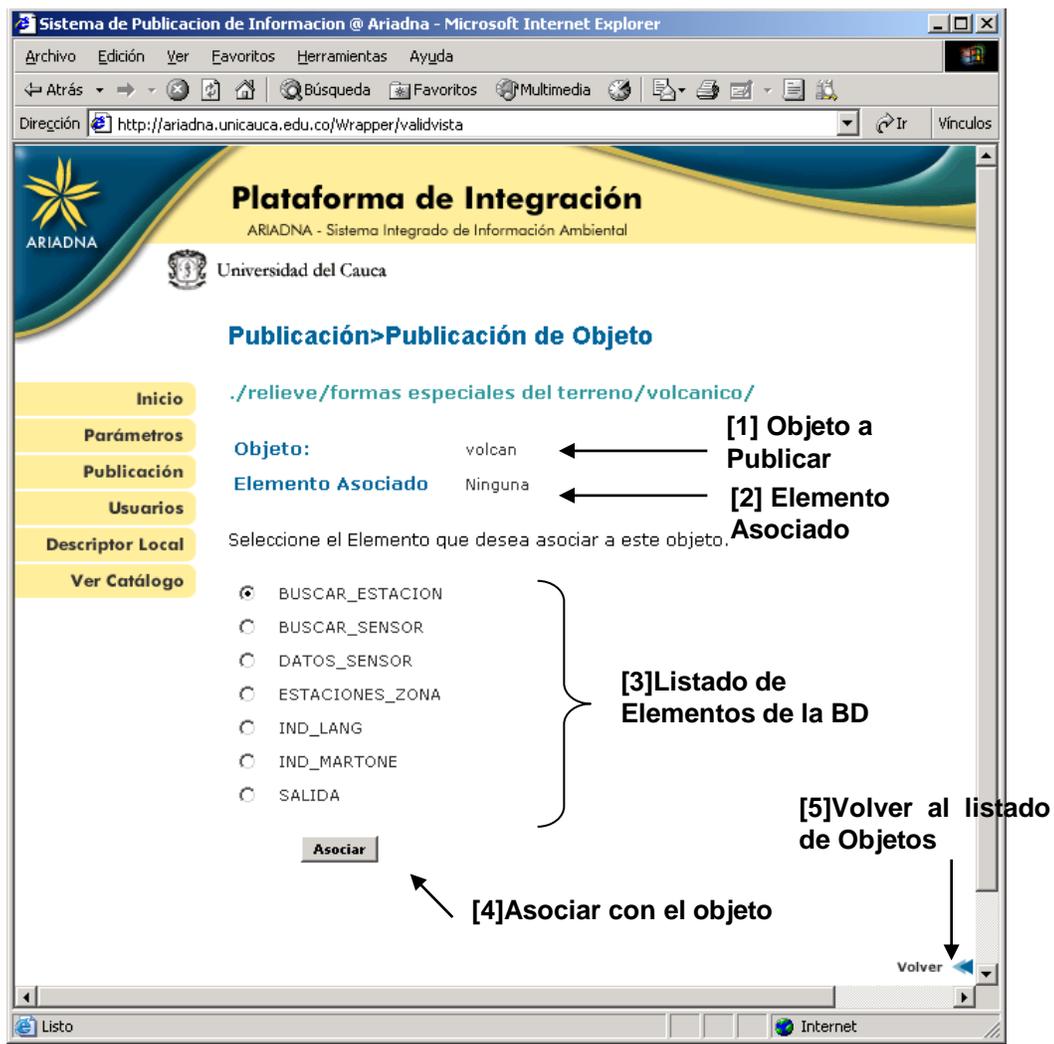


Figura 6.5 - Interfaz de Publicación de Objetos

En la Interfaz de Publicación de Objetos (figura 6.5), se muestran dos opciones: **Volver [5]** y **Asociar [4]**.

Si el administrador presiona la opción **Volver [5]**, el Sistema muestra el Listado de Objetos (Figura 6.4).

En el ítem **[2]** Elemento Asociado, aparece la palabra Ninguna, debido a que no se ha asociado ningún elemento. Por lo tanto, si se selecciona la opción **Asociar [4]**, el sistema asocia el objeto seleccionado en el ítem **[1]** de la Figura 6.5 (Volcán), con el elemento seleccionado en el ítem **[3]** (BUSCAR_ESTACIÓN), mostrando finalmente la interfaz de la Figura 6.6.

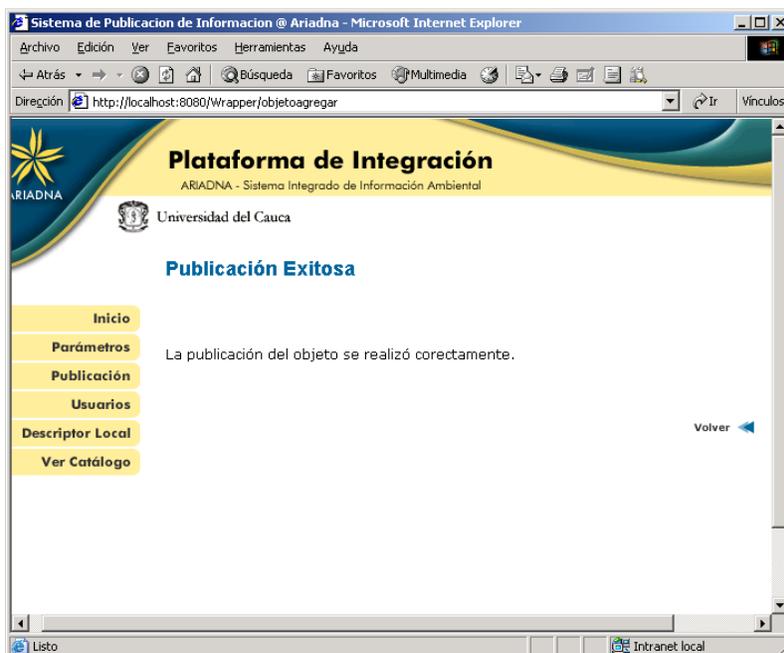


Figura 6.6 - Mensaje de Publicación Exitosa

6.2.3 Publicación de Atributos

Al hacer clic en la opción **Atributos** de la Figura 6.4, se mostrará un listado de todos los Atributos pertenecientes a un Objeto en particular.

La Figura 6.7 ítem **[3]**, muestra un listado de todos los atributos, la definición y el nombre del tipo de dato, así como el Campo al que se encuentra asociado. Dado que el objeto no ha sido previamente publicado aparece un mensaje informativo **[4]** indicando que sólo se puede publicar un Atributo perteneciente a un objeto publicado.

El enlace **Volver [5]** conduce al Listado de Objetos en el nivel en el que se encontraba.

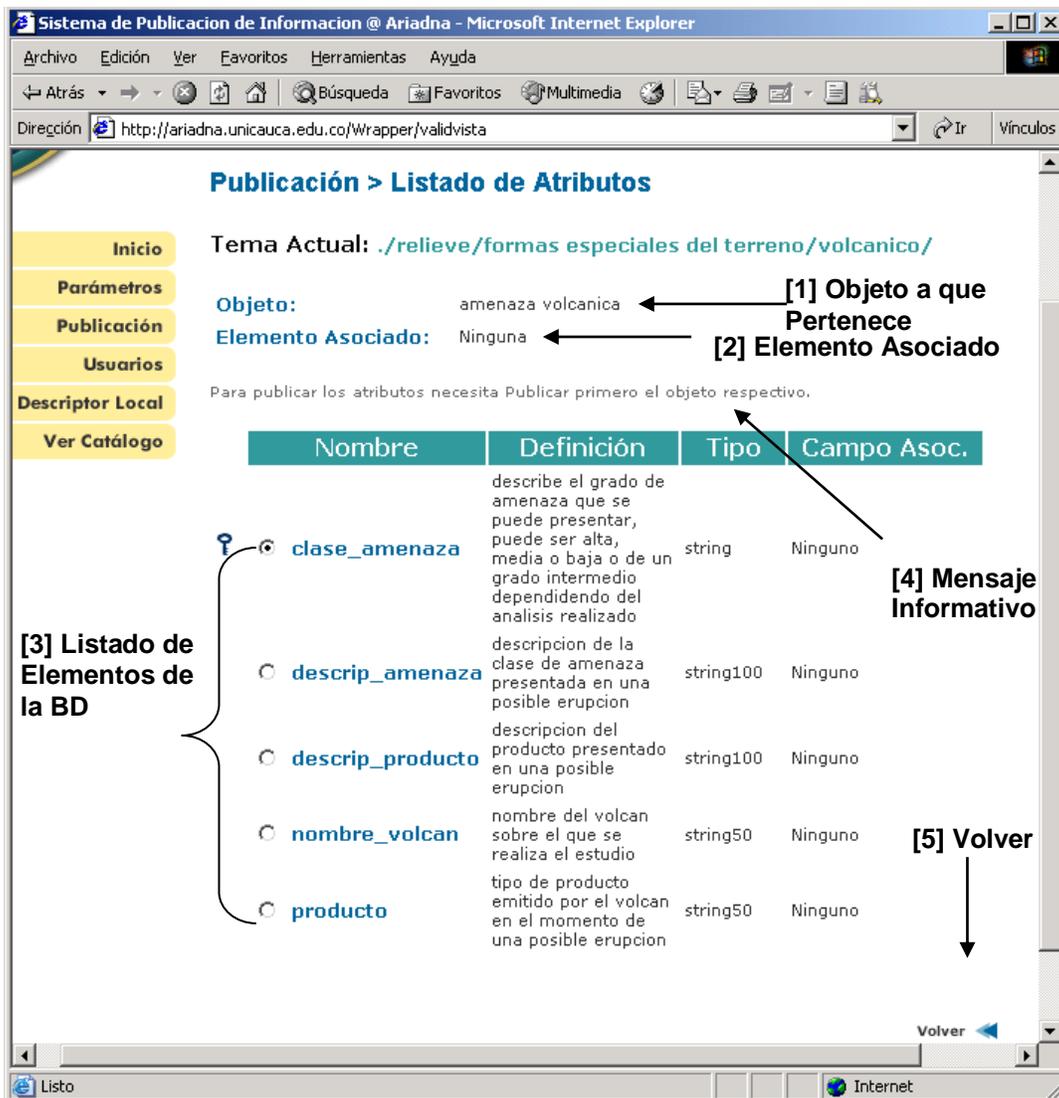


Figura 6.7 - Interfaz de listado de Atributos de un objeto no publicado

Si se consultan los atributos de un objeto previamente publicado, se muestra una interfaz similar a la de la Figura 6.8, la cual sólo difiere de la interfaz anterior (Figura 6.7) en el hecho de que aparece el nombre del elemento de la base de datos asociado (ítem [3] de la Figura 6.8) y las dos opciones: **Publicar** y **Hacer Privado**.

Además, en el ítem [1] de la Figura 6.8, se señala el atributo principal del objeto, para indicar que es el atributo utilizado en la plataforma como índice para asociar el objeto con la cartografía.

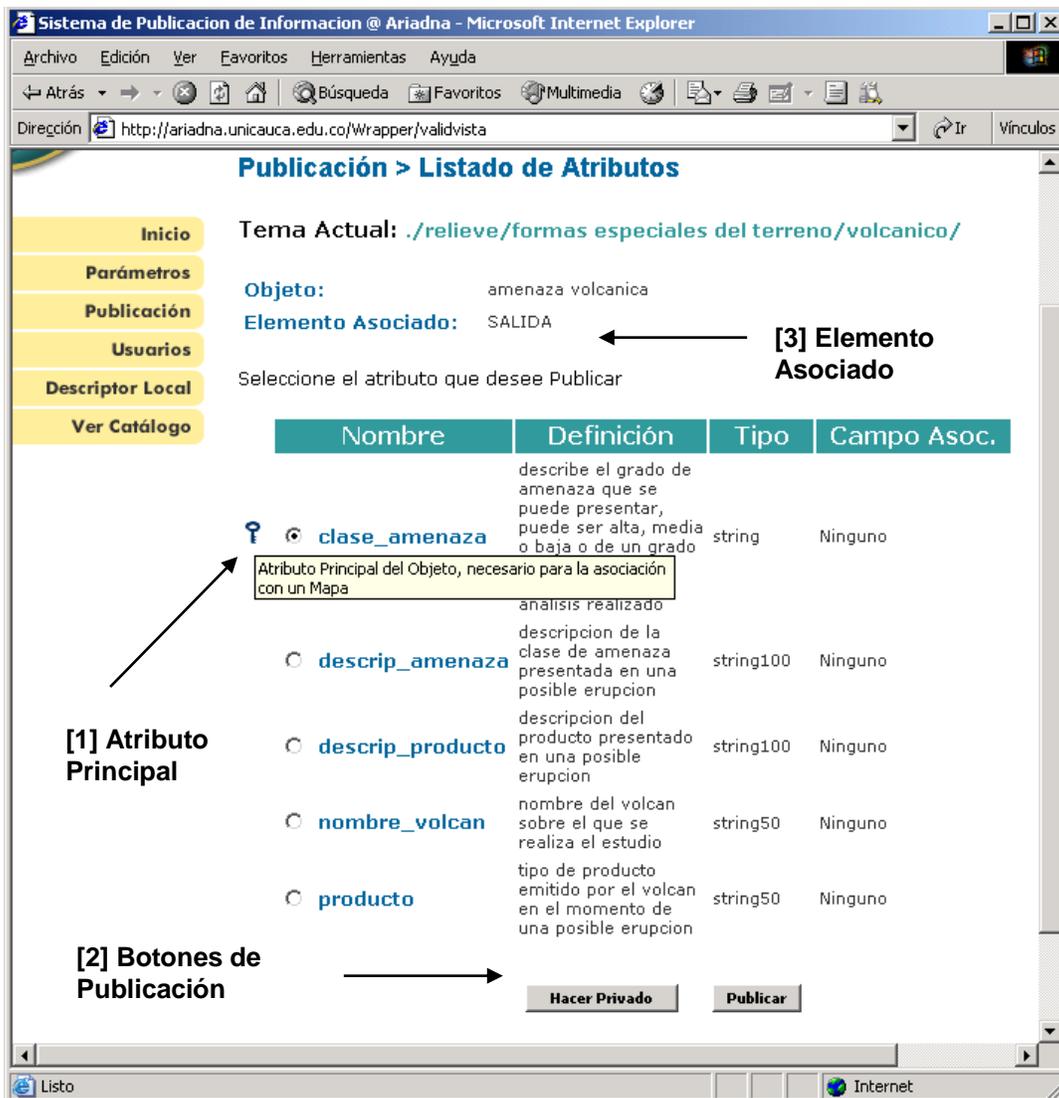


Figura 6.8 - Interfaz de Listado de Atributos de un Objeto publicado

Si se hace clic en **Hacer Privado** (Figura 6.8) se elimina la asociación entre el atributo seleccionado y el elemento de la base de datos con que se encuentre seleccionado.

Si hace clic en **Publicar**, se muestra un listado de los campos del elemento de la base de datos, con que se encuentre asociado el objeto padre. Como se muestra en la Figura 6.9.

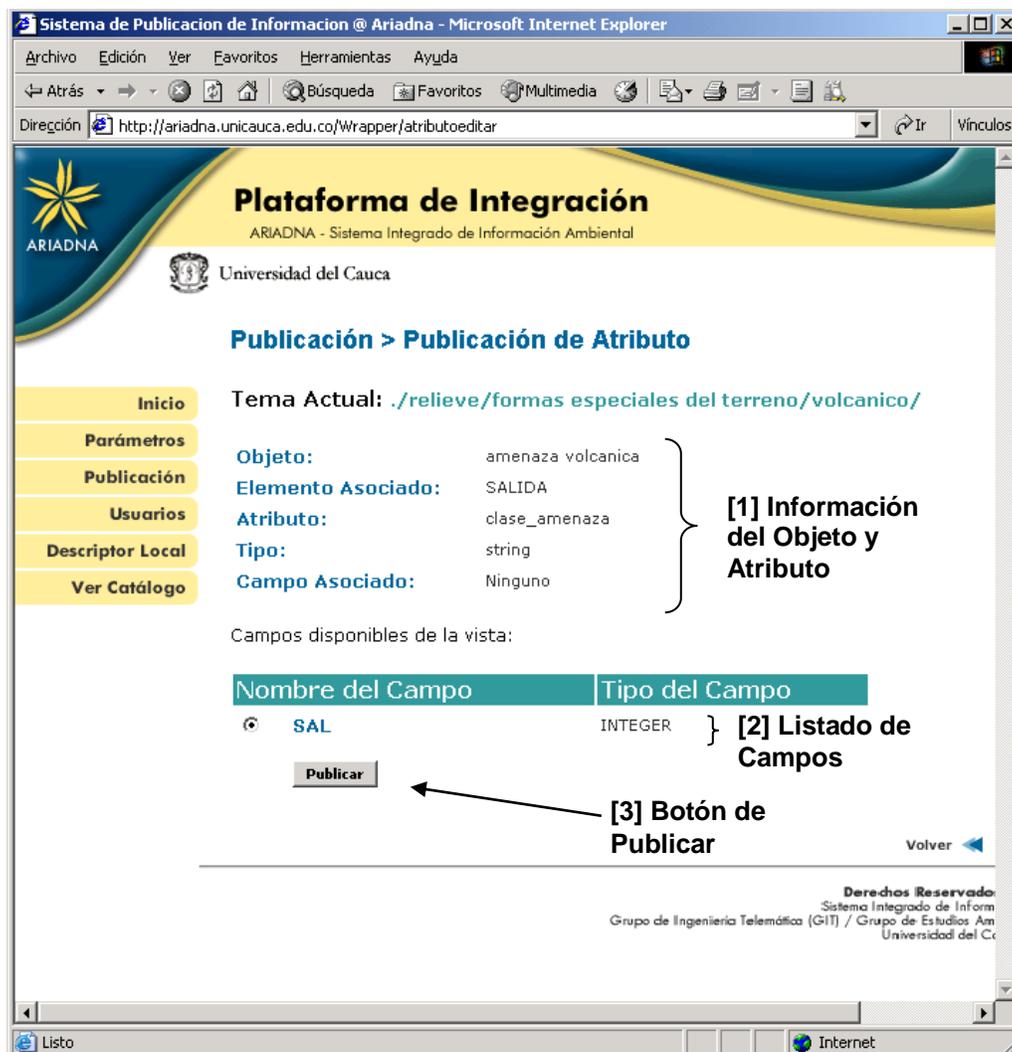


Figura 6.9 - Interfaz de Publicación de Atributos

En la Figura 6.9 se muestra la información para la publicación de un atributo. El ítem [1] muestra la información del objeto padre y de su atributo. De este último se muestra el tipo y el campo asociado en el momento.

También se muestra un listado de campos del elemento de la base de datos local (ítem [2]) con su tipo. Se resalta la importancia de los tipos puesto que el sistema antes de publicar revisa que los tipos de datos de los atributos del objeto padre y de los campos del elemento de la base de datos local, sean compatibles.

Para la correcta ejecución de las consultas es necesario que los tipos concuerden, dicho proceso se realiza en el momento de la instalación del sistema

En general los tipos numéricos de cualquier base de datos (MySQL, InterBase, SQLServer) son compatibles con el tipo NUMERIC de la plataforma. Mientras que los tipos VARCHAR, CHAR y BLOB de estos motores son compatibles con el tipo STRING de la plataforma, de la misma manera que los DATE.

Si se intenta asociar un Atributo con un campo que no sea compatible, como en el caso de la Figura 6.9, donde se intenta asignar el atributo clase_amenaza que es de tipo STRING, al campo SAL de tipo INTEGER, se generará un mensaje de Error indicando la situación. De lo contrario el administrador puede publicar los atributos eligiendo la opción **Publicar** (Figura 6.10).

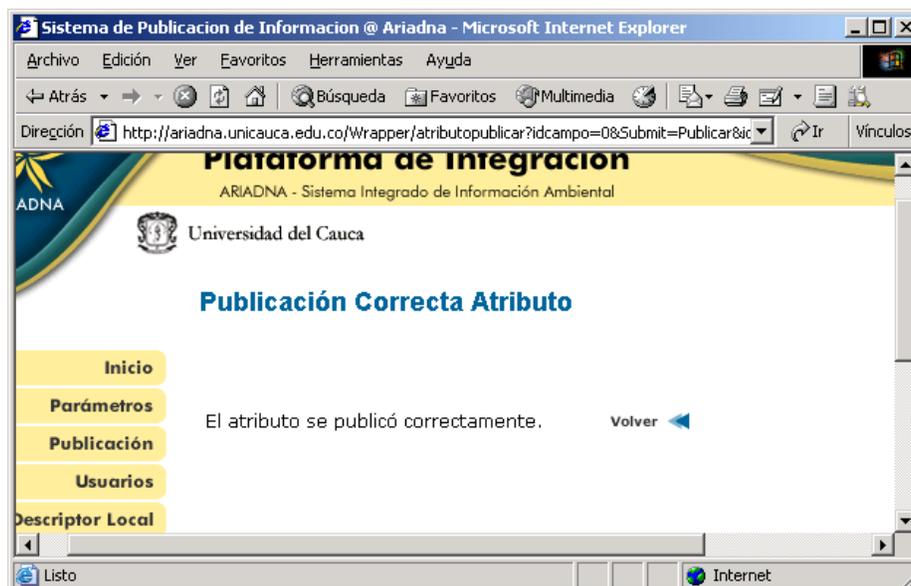


Figura 6.10 - Mensajes Obtenidos del proceso de Publicación

6.3 Descriptor Local

Esta opción del Módulo Wrapper, permite realizar las operaciones de actualizar metadatos y crear descriptor local.

La Figura 6.11 muestra la manera como se puede actualizar la base de datos del módulo wrapper de acuerdo a los metadatos definidos desde el módulo mediador. De la misma manera, se puede realizar la creación del descriptor local, el cual contiene la información que se ha publicado de la fuente de datos local (Sección 6.2 Publicar información).

6.3.1 Actualizar Metadatos

Al seleccionar la opción **Actualizar Metadatos** se revisa la ruta donde se encuentra la carpeta de configuración del sistema y se obtiene toda la información del documento **arquitectura.xml**, actualizando la base de datos del wrapper con los temas y tipos de objetos de los nuevos Metadatos.

Al realizar la actualización, todas las publicaciones previas dejan de ser válidas, por tal motivo el sistema muestra un mensaje de confirmación al Administrador del Sistema.

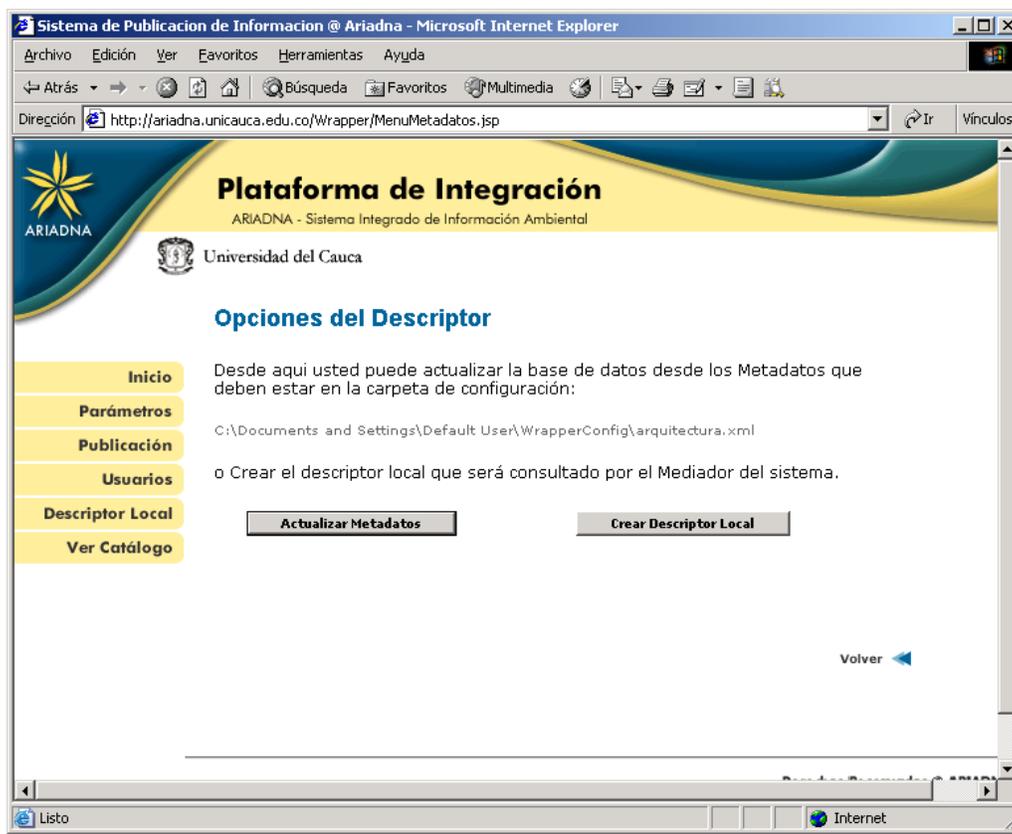


Figura 6.11 - Menú Descriptor Local

6.3.2 Creación del Descriptor Local

Si el Administrador del Sistema presiona **Crear Descriptor Local**, el Sistema crea un documento XML, llamado **metadata.xml** en la carpeta de configuración. Dicho documento contiene la información publicada por la fuente de datos local. **metadata.xml** es utilizado por el mediador para conocer las potencialidades que ofrece la fuente de datos.

Al finalizar la operación de creación del descriptor local, se muestra un mensaje con el resultado de la operación, y un enlace al documento generado. Este documento es buscado por el Mediador en la ubicación establecida en el documento de configuración del Módulo wrapper [GIT04d].

6.4 Acceso al Mediador

En esta sección se describe la forma como se pueden realizar consultas al Mediador, tanto para la visualización de Mapas como para el acceso a la Información [GIT04e].

El Mediador esta implementado como un servicio Web, por lo tanto para el acceso a sus funcionalidades se requiere conocer los parámetros de conexión al Servicio Web, que se encuentran especificados en el documento WSDL:

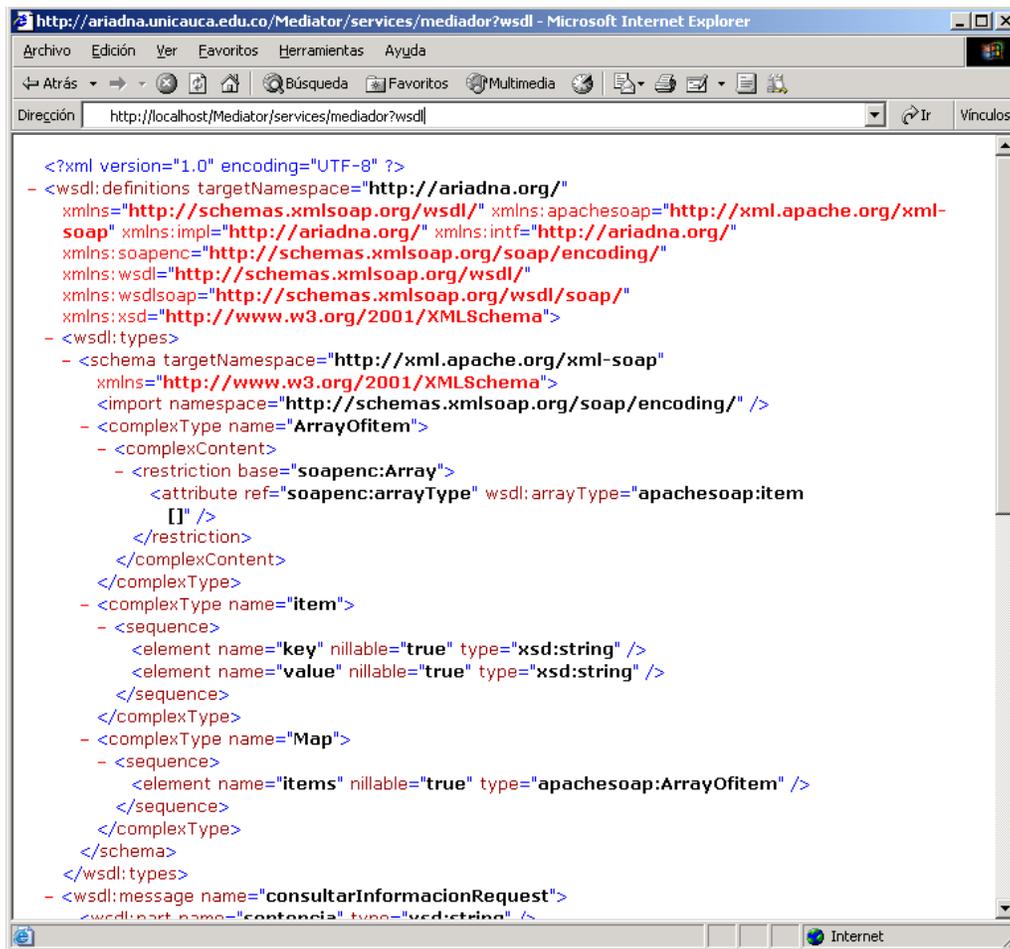


Figura 6.12 - Documento WSDL de Acceso al Módulo Mediator

El Mediator provee dos métodos, que pueden accederse vía SOAP a la dirección del Servicio Web (Nota: aunque como Métodos de un servicio Web, son independientes de la Implementación, a continuación se describen utilizando la Sintaxis de Java):

- public HashMap consultarInformacion(String sentencia, String login, String password) {}
- public byte[] consultarMapa(String sentencia, String login, String password) {}

El primer método devuelve una estructura de datos con la información resultante de las consultas, mientras que el segundo, retorna un arreglo de bytes de un archivo Shapefile construido a partir de la información consultada.

El login y el password corresponden a las credenciales de acceso entregadas al Cliente, mientras que la sentencia corresponde a una Sentencia SQL que se construye teniendo como base el Catalogo de Objetos del ICDE. Un usuario consumidor de servicios web puede revisar los Objetos sobre los cuales puede realizar consultas, a través de la página del proyecto:

- **Amenaza por flujos de lodo:** El mapa para esta clase de amenaza, visualiza información de las poblaciones y ríos que se verían afectadas en caso de ocurrencia del fenómeno de avalanchas ocasionadas por el flujo de lodo.
- **Mapa geológico:** En esta sección se presenta el mapa geológico, mostrando información de las unidades litológicas que lo componen.
- **Mapa de sismos:** En esta sección se presenta los sismos ocurridos en la zona en estudio durante un tiempo determinado. Este mapa asocia a la información de cada sismo, las poblaciones aledañas al punto de ocurrencia de dicho fenómeno, y la altura sobre el nivel del mar en el cual sucedió.
- **Mapa de cuencas:** El mapa de cuencas, muestra la información de las cuencas, subcuencas y microcuencas de la zona de estudio, además de la influencia de todas las posibles amenazas que puedan ocurrir.
- **Mapa de estaciones:** Este mapa presenta las estaciones climatológicas de la zona de estudio y los sensores asociados a cada una de ellas.

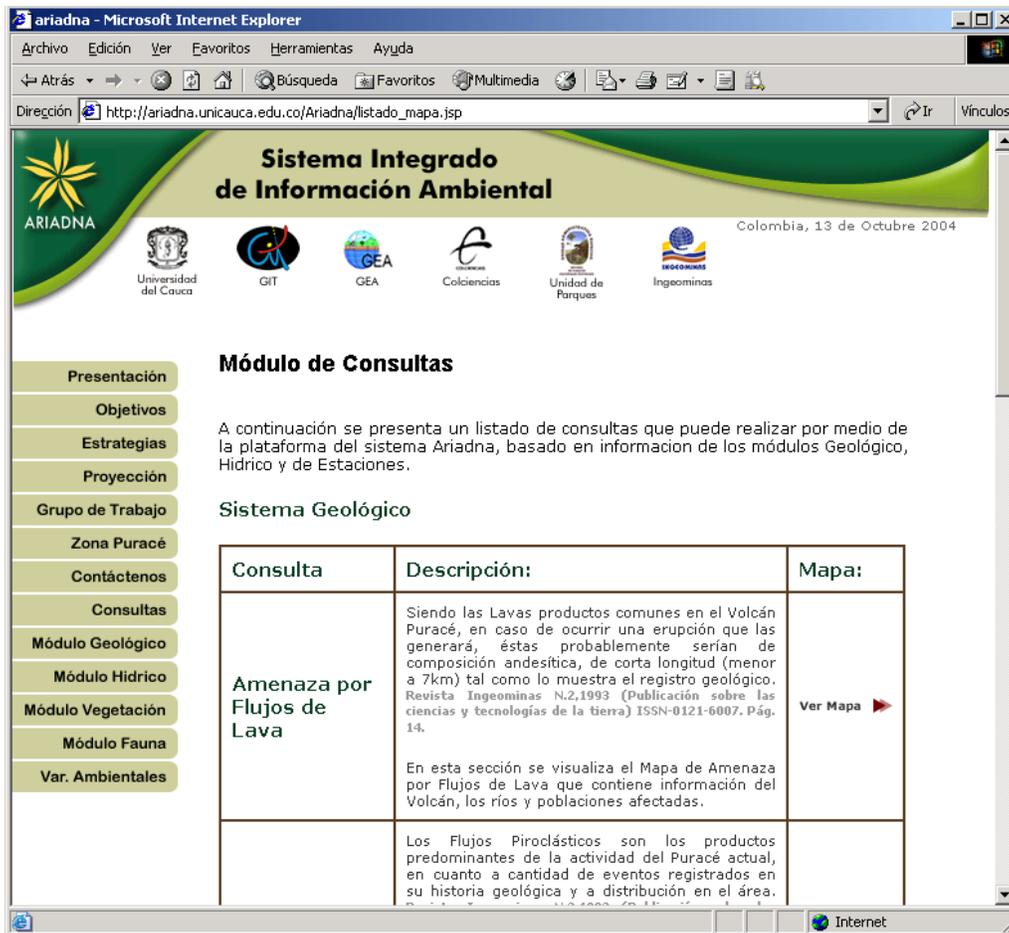


Figura 6.14 – Listado de mapas

Al hacer clic en cualquiera de los enlaces “Ver Mapa” se presenta el mapa consultado. A continuación se ilustra el proceso de consulta de información que acompaña la visualización de dos mapas disponibles a través de la plataforma, el “Mapa de sismos” y el “Mapa de cuencas”.

En la Figura 6.15 se presenta el mapa de sismos. Este mapa exhibe los sismos ocurridos en la zona en estudio durante un tiempo determinado, asociando a la información de cada sismo las poblaciones aledañas al punto de ocurrencia de dicho fenómeno, y la altura sobre el nivel del mar en el cual sucedió.

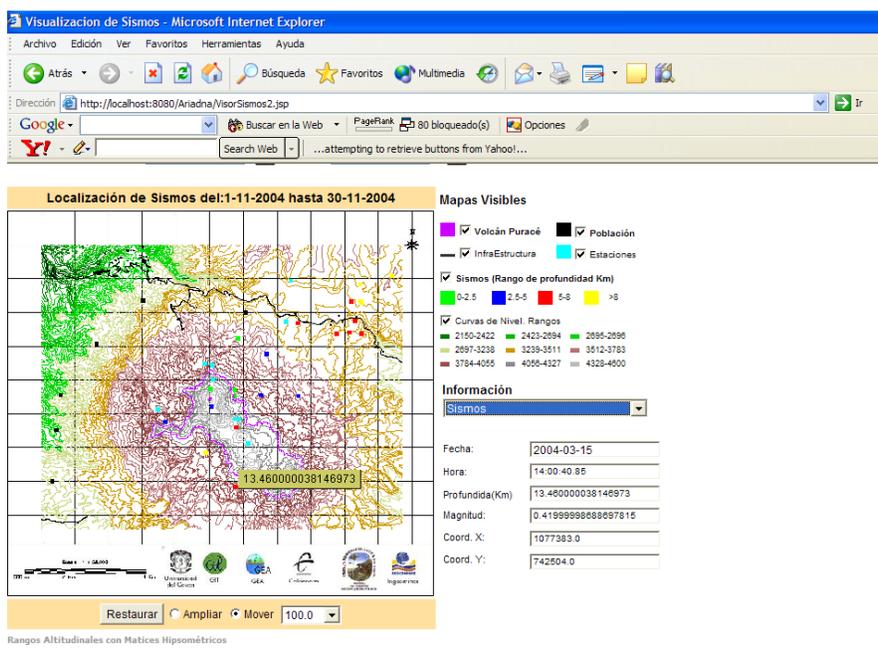


Figura 6.15 – Mapa de Sismos

El proceso de consulta de información que sigue la plataforma y que finalmente se ve reflejada en la visualización del mapa de sismos es el siguiente:

- El usuario selecciona la opción “**Ver Mapa**” de sismos.
- El sistema mediador verifica la sintaxis de la consulta solicitada a través de la opción “**Ver Mapa**” de sismos.
- El sistema Mediador verifica la localización de las fuentes de datos locales participantes en dicha consulta. Para este caso en particular la Fuente de datos local participante es la Fuente de datos geológicos [GIT04f].
- El sistema ejecuta la consulta sobre el módulo Wrapper que envuelve la fuente de datos geológica.
- Los módulos wrappers verifican el acceso a dicha fuente.
- El módulo wrapper ejecuta la consulta sobre la fuente de datos geológicos.
- El wrapper recupera la información de dicha fuente de datos local y la entrega al módulo Mediador. Los datos recuperados son: propiedades de los sismos ocurridos (fecha, hora, profundidad, magnitud y coordenadas de ocurrencia del sismo), atributos de las poblaciones de la zona en estudio (Nombre de la población, habitantes, altura, tipo de población y distancia en kilómetros de la población al volcán), y propiedades de las estaciones que monitorean el sismo (Nombre, tipo, estado, altura, y coordenadas de la estación) [GIT04f].
- El sistema Mediador estructura la respuesta a partir de la información recuperada de la fuente de datos consultada.

- El módulo de visualización presenta el mapa consultado (a los usuarios no consumidores de servicios web), el cual esta compuesto de las capas cartográficas de curvas de nivel (para representar la altura sobre el nivel del mar), la capa de infraestructura (vías de acceso y poblaciones de la zona en estudio), la capa de puntos que representa la ocurrencia de sismos seleccionadas, y todos los datos extraídos de la fuente de datos geológica.

La Figura 6.16 presenta el mapa de cuencas. A diferencia del mapa anterior, éste es estructurado a partir de la consulta de dos fuentes de datos: Fuente de datos geológica [GIT04f] y la Fuente de datos hídrica [GIT04g]. El objetivo de este mapa es presentar la información de las cuencas, subcuencas y microcuencas de la zona de estudio, además de la influencia de todas las posibles amenazas que puedan ocurrir.

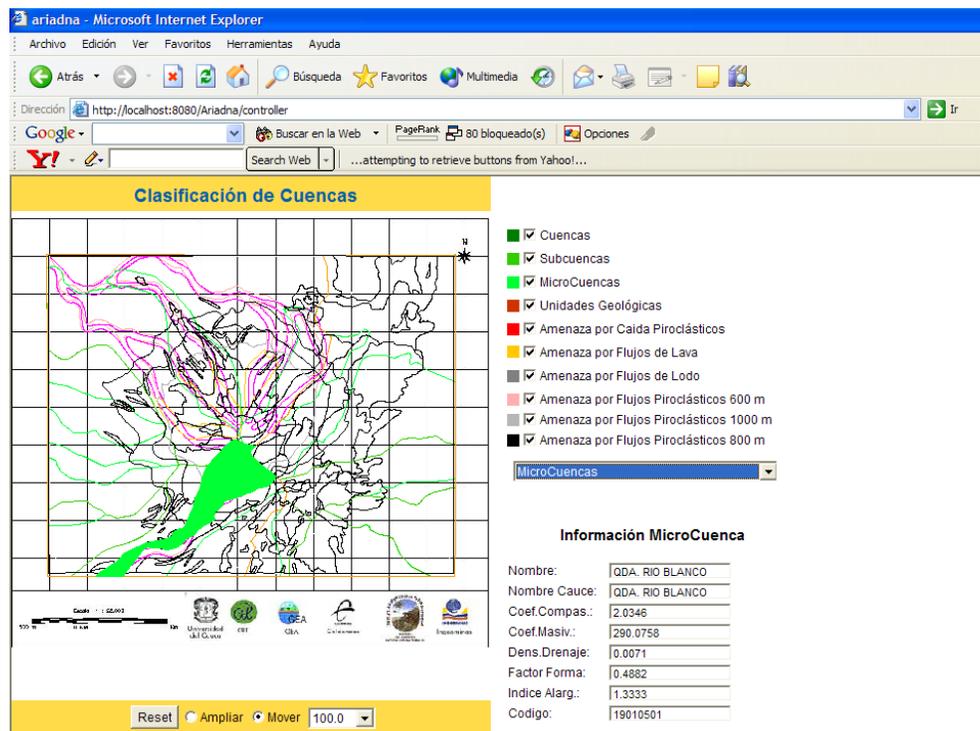


Figura 6.16 – Mapa de Cuencas

El proceso de consulta de información que sigue la plataforma y que finalmente se ve reflejada en la visualización del mapa de cuencas es el siguiente:

- El usuario selecciona la opción “**Ver Mapa**” de cuencas.
- El sistema mediador verifica la sintaxis de la consulta solicitada a través de la opción “**Ver Mapa**” de cuencas.
- El sistema Mediador verifica la localización de las fuentes de datos locales participantes en dicha consulta: Fuente de datos geológica [GIT04f] y Fuente de datos hídrica [GIT04g].
- El sistema ejecuta la consulta sobre los módulos Wrappers de cada fuente de datos local respectivamente.
- Los módulos wrappers verifican el acceso a cada fuente de datos que están representando.

- Cada módulo wrapper ejecuta la consulta sobre la fuente de datos local que envuelve.
- El wrapper recupera la información de cada fuente de datos local y la entrega al módulo Mediador. Los datos recuperados de la fuente de datos geológica son: Unidades geológicas (litologías y área), Amenaza por caída de piroplastos (600, 800 y 1000 metros), Amenaza por flujo de lodo y Amenaza por flujo de lava. Los datos recuperados de la fuente de datos hídrica están clasificados en tres temas: Cuencas, Subcuencas y Microcuencas. De cada una de estos temas se extraen las siguientes propiedades: Nombre (de la cuenca, subcuenca o microcuenca), nombre del cauce principal, coeficiente de Compasividad, Coeficiente de Masiv, Densidad de drenaje, Factor de forma, Índice de Alarg, código.
- El sistema Mediador estructura la respuesta a partir de la información recuperada de las múltiples fuentes de datos consultadas.
- El módulo de visualización presenta el mapa consultado (a los usuarios no consumidores de servicios web), el cual está compuesto de las capas cartográficas de cuencas, subcuencas, microcuencas, geológica, de amenazas (piroplastos, flujo de lava, de lodo) y los datos extraídos de cada fuente de datos local [GIT04f], [GIT04g].

7. CONCLUSIONES Y RECOMENDACIONES

Para dar mayor claridad a las conclusiones que se han originado en el presente trabajo, éstas se han clasificado en tres secciones. La primera define un conjunto de conclusiones desde la perspectiva de las instituciones que integran sus fuentes de datos a través de la plataforma, la segunda presenta las conclusiones técnicas del sistema, y la última hace referencia al punto de vista de los especialistas de medio ambiente.

Perspectivas de las Instituciones

Partiendo de la premisa que indica que la generación de información que alimentan los procesos de gestión ambiental requiere que las instituciones involucradas en dicho proceso dediquen gran cantidad de horas hombre y dinero para su generación; y contrastando esta premisa con la realidad nacional en donde los recursos para la conservación del medio ambiente asignados a las entidades estatales es limitado, es preponderante y urgente la definición de mecanismos estándares que permitan a las instituciones organizar, estructurar y compartir la información, con el fin minimizar la replicación de esfuerzos institucionales, y permitiendo el acceso a la misma de manera oportuna.

La plataforma para la integración de múltiples fuentes de datos descriptivos y espaciales en un sistema de información ambiental, da solución a los inconvenientes presentados, ya que:

- En primer lugar facilita el proceso de publicación de la información de cada entidad participante en las funciones de gestión ambiental, de manera fácil y segura.
- Por otro lado, permite un acceso oportuno a la información publicada por cada institución, a través de un navegador o en caso de requerir datos mas puntuales, por medio de un servicio web que puede ser accedido fácilmente haciendo uso del protocolo SOAP, superando los complicados procesos de solicitud de información a nivel interinstitucional, que además de los largos procedimientos administrativos involucran también los procesos de transformación de formatos o conversión de coordenadas.
- Como se ha mencionado en el párrafo anterior, la plataforma optimiza el acceso oportuno a la información de las diversas instituciones, permitiendo así, la reducción del tiempo en los procesos de toma de decisiones frente a la ocurrencia de emergencias o desastres.

Perspectiva Técnica

Una de las grandes potencialidades desde el punto de vista técnico de la Plataforma, radica en la solución dada a los conflictos semánticos entre la información cartográfica y temática suministradas por las diferentes fuentes de datos de cada institución. Para dar solución a dichos conflictos, la plataforma incorpora en su arquitectura un modelo de integración de datos basado en un sistema mediador y wrappers. Esta arquitectura gira en torno a los metadatos, que para esta solución particular toma como dominio de aplicación el medio ambiente, por tal motivo, estos metadatos están basados en el estándar de datos espaciales para Colombia dictados por el ICDE (Instituto Colombiano de Datos Espaciales).

Sin nos fijamos detenidamente en la arquitectura de la plataforma, es muy fácil apreciar que el componente que determina el dominio en el cual puede ser utilizada son los metadatos, de ahí que la plataforma esta concebida como una solución genérica adaptable a cualquier dominio de aplicación de acuerdo a los metadatos utilizados.

Desde el punto de vista de la conexión, al estar la plataforma basada en servicios web, permite que una entidad que pretenda integrarse a través de la plataforma no tenga necesidad de conocer nada acerca de su implementación, dado que es suficiente conocer su definición WSDL. De ésta manera, la plataforma proporciona un alto nivel de integración e interoperabilidad, dejando de lado la implementación de los sistemas de cada institución. Desde la perspectiva de un usuario que hace una consulta a través de la plataforma existe independencia del lenguaje de programación, ya que no es necesario que del lado del usuario y del lado de la plataforma estén escritos en el mismo lenguaje.

Finalmente, al estar la plataforma basada en XML, ésta es fácil de adaptar, extender y personalizar.

Si bien es cierto que la plataforma soluciona los inconvenientes de intercambio de información tanto a nivel institucional, de análisis de medio ambiente y técnico; todas las consultas realizadas a través de ésta se soportan en la posición geográfica, lo cual hace que un sistema que realice una consulta a través de la plataforma y que involucre múltiples fuentes de datos, este supeditado a conocer las coordenadas del área en estudio para poder localizar los datos temáticos provenientes de las múltiples fuentes. Lo ideal sería poder realizar consultas a la plataforma sin necesidad de incluir los parámetros espaciales, de tal manera que la plataforma tenga la capacidad de interpretar la consulta solicitada por el sistema sin tener en cuenta la posición geográfica, y asociarla después al componente cartográfico correspondiente. Éste será un tema a abordar para una futura versión de la aplicación.

Perspectivas de los especialistas en medio ambiente

Desde el punto de vista de los especialistas en medio ambiente la plataforma para la integración de múltiples fuentes de datos descriptivos y espaciales en un sistema de información ambiental, les permite alcanzar análisis que conducen a:

- Disminuir la presión humana sobre los ecosistemas, permitiendo identificar y evaluar indicadores que pongan en evidencia situaciones de riesgo sobre el recurso natural,

bien por disminución del mismo o por la baja rentabilidad en su utilización en procesos de producción.

- Disponer de información para la toma de decisiones, al igual que el procesamiento de datos en tiempo real que permitan acciones rápidas y efectivas frente a potenciales amenazas o peligros que pueden ser generados por fenómenos naturales
- Elaborar programas de ordenamiento del paisaje para la conservación biológica y de mantenimiento de servicios ecológicos (preservación del agua, suelo, fauna, vegetación).
- Elaborar planes de manejo y ordenamiento ambiental que favorezcan la valorización del paisaje a través del turismo.

BIBLIOGRAFÍA

- [ARR96] Arrow, K. y otros, 1996. Economic Growth, Carrying Capacity and the Environment. Washington, D.C.: The Ecological Society of America.
- [AXI03] Axis 1.1 API Documentation. 2003. <http://ws.apache.org/~dms/axisdiff/newdocs/>
- [BAR94] Barker, R. El Modelo Entidad-Relación (Case Method). Madrid: Addison-Wesley. 1994.
- [BAT94] Bartino, C. et. al. Diseño conceptual de bases de datos, un enfoque de entidades-relaciones. Argentina, Addison-Wesley. 1994.
- [BE+96] O. A. Bukhres, A. K. Elmagarmid, F. Gherfal, X. Liu, K. Barker and T.Schaller. The Integration of Database Systems pág. 37- 56 en Object-Oriented Multidatabase Systems: A Solution for Advanced Applications editado por Omran A. Bukhres y Ahmed K. Elmagarmid. Prentice-Hall, 1996.
- [BSD03] B. Benatallah, Q. Sheng, M. Dumas. The Self-Serv Environment for Web Services Composition. IEEE Internet Computing – January/February 2003.
- [CAM93] Campiglio, L. y otros. 1993. The Environment After Río. London: Graham & Trotman.
- [CCC99] S. Carbo., D. Comas., M. Compte., I. Compte., M, Yanes. Modelado de Datos Espaciales. Master Internacional en Sistemas de Información Geográfica. Universidad de Girona. España. 1999.
- [CDM90] CDMAALC (Comisión de Desarrollo Sostenible y Medio Ambiente de América Latina y el Cari-be), 1990. Nuestra Propia Agenda. Washington, D. C.: BID, UNDP.
- [CEA00] Consultas espaciales en una arquitectura de componentes GIS. Universidad de las Américas. Puebla, México. 2000. http://mailweb.udlap.mx/~tesis/loranca_m_o/
- [CPE84] Ceri S. and Pelagatti C. Distributed Databases, Principles and Systems McGraw-Hill, 1984.
- [CML89] J.A. Cebrián., D.M. Mark., J.P. Lauzon. Jerarquías de Datos para Sistemas de Información Geográfica de Grandes Dimensiones. En Estudios Geográficos, nº 196, pp 369-380.1989.
- [COD90] Codd, Edgar F. The Relational model for database management: version 2, Addison-Wesley, cop. 1990.
- [COV98] J.C. Corrales Muñoz, I.D. Delgado Vallejo. "Digitalización de Mapas de Curvas de Nivel y Estructuración de Bases de Datos en un Sistema de Información Geográfica". Trabajo de Grado. Universidad del Cauca. 1998.
- [CRF03] J.C Corrales M, A. Rendón G, A. Figueroa C. "Plataforma para Integración de Datos entre Sistemas de Información Geográfica". II seminario Internacional de Sistemas de Información Geográfica. Santa Fe de Bogotá. Septiembre de 2003.
- [CRU93] D. Comas., E. Ruíz. Fundamentos de los Sistemas de Información Geográfica. Editorial Ariel S. A. Barcelona. 1993.

- [DES98] Desay, U., ed. 1998. Ecological Policy and Politics in Developing Countries. State of New York University Press, New York.
- [EAS92] Escobar, A. y Álvarez, S., eds. 1992. The Making of Social Movements in Latin America. Identity, Strategy and Democracy. Colorado: Westview Press.
- [ERS99] A. Elmagarmid, M. Rusinkiewicz y A. Sheth. Management of Heterogeneous and Autonomous Database Systems. Morgan Kaufmann, 1999. (ISBN 1-55860-216-X)
- [ESC99] Escobar, A. 1999. El Final del Salvaje, Naturaleza, Cultura y Política en la Antropología Con-temporánea. Santafé de Bogotá: CEREC, Instituto Colombiano de Antropología.
- [ESP02] Espinoza .G. Gestión y Fundamentos de Evaluación de Impacto Ambiental. Cooperación Técnica N° ATN/JF-6618-RG "Programa de Apoyo para el Mejoramiento de la Gestión Ambiental en los Países de América Latina y el Caribe". Banco Interamericano de Desarrollo – BID. Santiago de Chile. 2002.
- [ESR98] Environmental Systems Research Institute. ESRI Shapefile Technical Description. ESRI White Paper. Printed in the United States of America. 1998. <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>
- [ESR00] ESRI-Venezuela. ArcIMS. 2000
http://www.esriven.com/paginas/productos/Prod_ArcIMS.htm
- [FHA89] C. C. Fleming, B. Halle. Handbook of Relational Database Design. Addison-Wesley Publishing Company. ISBN 0201114348. 1989
- [FIO95] Fiorino, D.J. 1995. Making Environmental Policy. Berkeley: University of California Press.
- [FIR03] Firebird Project. 2003. <http://firebird.sourceforge.net/>
- [FNA01] FNAC (Foro Nacional Ambiental, Colombia), 2001. Memorias del Foro Electrónico: Bosques y Conflicto. Santafé de Bogotá: FNAC.
- [FRV95] Floresu D., Rachid L., Valdúriez P. Using Heterogeneous Equivalences for Query Rewriting in Multibases Systems. Proceedings of the third international conference on cooperative information systems. 1995.
- [GGO94] Gutiérrez J., Gould M. SIG: Sistemas de información geográfica. Madrid. Síntesis, 1994.
- [GIT02] Grupo de Ingeniería Telemática. Grupo de Estudios Ambientales. ARIADNA - Sistema Integrado de Información Ambiental. Propuesta para la participación en la Convocatoria Convenio 071/082 – 2001, Presentada a: Conciencias Programa Nacional de Electrónica, Telecomunicaciones e Informática. Ministerio de Comunicaciones de Colombia. Popayán, julio de 2002. Sin Publicar.
- [GIT04a] Grupo de Ingeniería Telemática., Universidad del Cauca. "Análisis – Sistema Integrado de Información Ambiental". Proyecto ARIADNA – Sistema Integrado de Información Ambiental. (Contrato 026/02 suscrito entre el Centro de Investigación de las Telecomunicaciones - CINTEL y la Universidad del Cauca). Popayán. Octubre de 2004.
- [GIT04b] Grupo de Ingeniería Telemática., Universidad del Cauca. "Diseño – Sistema Integrado de Información Ambiental". Proyecto ARIADNA – Sistema Integrado de Información Ambiental. (Contrato 026/02 suscrito entre el Centro de Investigación de las Telecomunicaciones - CINTEL y la Universidad del Cauca). Popayán. Octubre de 2004.

- [GIT04c] Grupo de Ingeniería Telemática., Universidad del Cauca. "Implementación – Sistema Integrado de Información Ambiental". Proyecto ARIADNA – Sistema Integrado de Información Ambiental. (Contrato 026/02 suscrito entre el Centro de Investigación de las Telecomunicaciones - CINTEL y la Universidad del Cauca). Popayán. Octubre de 2004.
- [GIT04d] Grupo de Ingeniería Telemática., Universidad del Cauca. "Manual de Usuario – Módulo Wrapper". Proyecto ARIADNA – Sistema Integrado de Información Ambiental. (Contrato 026/02 suscrito entre el Centro de Investigación de las Telecomunicaciones - CINTEL y la Universidad del Cauca). Popayán. Octubre de 2004.
- [GIT04e] Grupo de Ingeniería Telemática., Universidad del Cauca. "Manual de Usuario – Módulo Mediador". Proyecto ARIADNA – Sistema Integrado de Información Ambiental. (Contrato 026/02 suscrito entre el Centro de Investigación de las Telecomunicaciones - CINTEL y la Universidad del Cauca). Popayán. Octubre de 2004.
- [GIT04f] Grupo de Ingeniería Telemática., Universidad del Cauca. "Módulo de Gestión de Información Geológica". Proyecto ARIADNA – Sistema Integrado de Información Ambiental. (Contrato 026/02 suscrito entre el Centro de Investigación de las Telecomunicaciones - CINTEL y la Universidad del Cauca). Popayán. Octubre de 2004.
- [GIT04g] Grupo de Ingeniería Telemática., Universidad del Cauca. "Módulo de Gestión de Información Hídrica". Proyecto ARIADNA – Sistema Integrado de Información Ambiental. (Contrato 026/02 suscrito entre el Centro de Investigación de las Telecomunicaciones - CINTEL y la Universidad del Cauca). Popayán. Octubre de 2004.
- [GLV98] F. Gingras and L. V. S. Lakshmanan. nD-SQL: a multi-dimensional language for interoperability and Olap. Proc. VLDB, New York, 1998.
- [GMU95] A. Gupta and A. S. Mumick. Maintenance of Materialized Views: Problems, Techniques and Applications. Bulletin of Data Engineering, IEEE. June 1995. Vol.18 No.2 pages 3-18.
- [GTO03] The Geotools Project. 2003. <http://www.geotools.org/History>
- [GUI92] Guimet, J. Introducción conceptual a los Sistemas de Información Geográfica (SIG). Madrid : Estudio Gráfico; cop. 1992.
- [GWE98] J.R. Groff., P.N. Weinberg, "Guía Lan Times de SQL". McGraw-Hill. ISBN 84-487-1416-7. España.1998.
- [HDT99] Hommer-Dixon, T.F. 1999. Environment, Scarcity, and Violence. New Jersey: Princeton University Press.
- [HML93] J. Hammer and D: McLeod. An Approach to resolving semantic heterogeneity in a federation of autonomous database systems Int. Journal of Intelligent and Cooperative Information Systems 2(1):51-83 Marzo 1993.
- [IGI91] IGISE, (1991), Geographic data as a model of reality. Background Unit B in the Curriculum Development Toolkit, IGISE, Ottawa.
- [JBS97] J. Bosque S. Sistemas de Información Geográfica. Madrid: Ediciones Rialp, S.A., 2ª edición. 1997.
- [JWE97] Janicke, M. y H. Weidner, eds. 1997. National Environmental Policies. Berlin: Springer.

- [KSH96] V. Kashyap and A. Sheth. Semantic and Schematic Similarities between Data-base Objects: A Context-based Approach. The VLDB Journal , 5(4), 1996.
- [LSS96] Laks V. S. Lakshmanan, Fereidoon Sadri, Iyer N. Subramanian. SchemaSQL – A Language for Interoperability in Relational Multi-database Systems. Proc. 22nd VLDB, Mumbai, India, 1996, pp. 239-250.
- [LIT94] W. Litwin. Multidatabase Systems. Prentice Hall: Englewood Cliffs, N.Y., 1994.
- [MFW95] M.F. Worboys. "GIS: A Computing Perspective". Taylor&Francis, London, 1995.
- [MMN01] Ministerio de Medio Ambiente y los Recursos Naturales, República de Nicaragua. "Sistema de Información Ambiental Nacional". SINIA Nicaragua V1.2 2001 <http://www.marena.gob.ni/SIA02.htm>
- [MMP01] Ministerio de Medio Ambiente, República de Peru. "Sistema de Información de la Diversidad Biológica y Ambiental de la Amazonía Peruana". Perú 2001. <http://www.siamazonia.org.pe>
- [MOT02] Motz. R. Interoperabilidad de Sistemas de Información. Material Curso de Maestría en Ingeniería con Énfasis en Telemática. Universidad del Cauca. Colombia. 2002.
- [OCD97] OCDE, 1997a. Economic Globalization and the Environment. Paris: OECD.
- [PNU98] PNUMA, 1998. Informe del Comité Intersesional de Ministros de Medio Ambiente de América Latina y el Caribe. Ciudad de México: PNUMA.
- [RAP02] Raposo J., Alvarez M., Pan A., Viña A. Un Sistema Mediador en la Práctica: Base de Datos Virtual. II Congreso Iberoamericano de Telemática (CITA 2002). Mérida-Venezuela. Septiembre 2002.
- [REC+89] M. Rusinkiewicz, R. Elmasri, B. Czejdo, D. Georgakopoulos, G. Karabatis, A. Jamousi, K. Loa and Y. Li. Query processing in a heterogeneous multidatabase environment . In proceedings of the first annual symposium on Parallel and Distributed Processing, 1989
- [REE99] D. Reeve. Base de Datos, Estructuras y Modelos de Datos Espaciales. Master Internacional en Sistemas de Información Geográfica. Universidad de Girona. España. 1999.
- [REW02] Rodríguez .M. Espinoza .G. D. Wilk. Gestión ambiental en América Latina y el Caribe Evolución, tendencias y principales prácticas. Banco Interamericano de Desarrollo, Departamento de Desarrollo Sostenible, División de Medio Ambiente. Diciembre de 2002.
- [RMM99] Romero M. M., Lenguaje de Consultas para una Multibase de Datos. Tesis de Maestría. Ciencias con especialidad en Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas-Puebla. 1999.
- [RSM87] Robinson A.H., Sale R., Morrison J.L., Muehrcke P.C. Elementos de Cartografía, Omega, pp. 1987.
- [SLA90] A. P. Sheth and J. A. Larson. Federated Database Systems and managing distributed heterogeneous, and autonomous databases, ACM Computing Surveys, 22(3):183-226, 1990.

- [SSP94] S. Spaccapietra and C. Parent View Integration: A step forward in solving structural conflicts. IEEE Transactions on Knowledge and Data Engineering, 6(2), 1994.
- [SSU91] Silberschatz A., Stonebraker M., Ullman J.D. "Database Systems: Achievements and Opportunities", Comm. of the ACM, 34, 10, pp. 110-119.
- [WCE87] WCED (World Commission on Sustainable Development), 1987. Our Common Future. Oxford: Oxford University Press.
- [WIE92] G. Wiederhold. Mediators in the architecture of future information systems. IEEE Computers, Vol 25, N3, pp 38-49, March 1992.
- [WIE94b] G. Wiederhold. Interoperation, Mediation and Ontologies. Proc. Int. Symposium on 5th Generation Computer Systems (FGCS'94). Workshop on Heterogeneous Cooperative Knowledge Bases, Vol W3, pp 33-48, Tokyo, 1994.
- [WSA03] Web Services Architecture. 2003. [HTTP://www.w3c.org/TR/ws-arch](http://www.w3c.org/TR/ws-arch)
- [W3C02] W3C, Web Services Activity Home Page. www.w3c.org/2002/ws/
- [WSC03] Web Services Choreography. 2003. [HTTP://www.w3.org/TR/2003/WD-ws-chor-reqs-20030812/](http://www.w3.org/TR/2003/WD-ws-chor-reqs-20030812/)
- [WSO01] Web Services – Technical overview. 2001. [HTTP://dcb.sun.com/practices/webservices/overviews/overview_wsdl.jsp](http://dcb.sun.com/practices/webservices/overviews/overview_wsdl.jsp)
- [WWG01] C. Wyss, F. D. Wyss and V. Gucht. Augmenting SQL with Dynamic Restructuring to Support Interoperability in a Relational Federation. Proceedings of the 4th Workshop on Engineering Federated Information Systems EFIS 2001, October 9-10, 2001, Berlin, Germany.