



UNIVERSIDAD NACIONAL DEL SUR

TESIS DE MAGISTER EN CIENCIAS DE LA COMPUTACIÓN

**Modelos de Representación
de Información Geográfica**

María Mercedes Vitturini

BAHÍA BLANCA

ARGENTINA

2010



UNIVERSIDAD NACIONAL DEL SUR

TESIS DE MAGISTER EN CIENCIAS DE LA COMPUTACIÓN

**Modelos de Representación
de Información Geográfica**

María Mercedes Vitturini

BAHÍA BLANCA

ARGENTINA

2010

Prefacio

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Magister en Ciencias de la Computación, de la Universidad Nacional del Sur, y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otras. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Departamento de Ciencias e Ingeniería de la Computación, durante el período comprendido entre el 1 de noviembre de 2002 y el 24 de diciembre e 2010, bajo la dirección del Dr. Pablo R. Fillottrani, Profesor Asociado del Departamento de Ciencias e Ingeniería de la Computación y la Dra. Silvia M. Castro, Profesora Titular del Departamento de Ciencias e Ingeniería de la Computación.

María Mercedes Vitturini

`mvitturi@cs.uns.edu.ar`

DEPARTAMENTO DE CIENCIAS E INGENIERÍA DE LA COMPUTACIÓN

UNIVERSIDAD NACIONAL DEL SUR

Bahía Blanca, 24 de Diciembre de 2010.



UNIVERSIDAD NACIONAL DEL SUR
Secretaría General de Posgrado y Educación Continua

La presente tesis ha sido aprobada el .../.../..., mereciendo la calificación de(.....).

Agradecimientos

A mis directores de tesis, Dr. Pablo R. Fillottrani y Dra. Silvia M. Castro por su generosidad, capacidad y experiencia científica brindadas en un marco de confianza y afecto.

Al jurado, Dr. Marcelo A. Falappa, Dr. Daniel E. Riesco y Dra. Aurora Vizcaíno por la dedicación y valiosas devoluciones.

A las autoridades y directivos del Departamento de Ciencias e Ingeniería de la Computación y de la Dirección General de Sistemas de Información por alentar, comprender y permitir que trabaje en este proyecto.

A mi amiga y compañera de trabajo Mg. Karina M. Cenci por su ánimo, presencia y colaboración desinteresada leyendo y corrigiendo este trabajo de tesis.

Al Ing. Carlos J. Matrangolo y personal técnico del Centro Científico Tecnológico CONICET de Bahía Blanca que hicieron posible el desarrollo de la defensa mediado por video conferencia.

Al personal administrativo del Departamento de Ciencias e Ingeniería de la Computación por su disposición, colaboración y calidez humana.

A mi familia y amigos por su comprensión y compañía en todo momento.

A mis compañeros de trabajo del Departamento de Ciencias e Ingeniería de la Computación y de la Dirección General de Sistemas de Información que me brindan cotidianamente la oportunidad de crecer compartiendo la actividad profesional.

A mis padres y mi hermano que me acompañan y están siempre presentes.

A mis sobrinos Juan, Ramiro, Tomás y Violeta que alegran mi vida.

A todos GRACIAS!

Mercedes Vitturini

Resumen

Los primeros sistemas automatizados que usaron información geográfica datan de los '70. Desde esa época y hasta ahora se crean aplicaciones para satisfacer las necesidades de diversos ámbitos de uso y problemas variados incluyendo: meteorología, planeamiento urbano, transporte, servicios ambientales, turismo, distribución económica, etc. Entre sus usuarios se cuenta a organizaciones estatales y gubernamentales, educativas y de investigación, industriales y comerciales.

Representar información geográfica es un problema complejo, se necesita mantener referencia a una o más ubicaciones de la superficie de la Tierra y eventualmente se puede requerir de propiedades geométricas. La información geográfica que modela a los fenómenos del mundo real se organiza en dos abstracciones: el modelo discreto que considera al espacio y sus objetos, y el modelo continuo que imagina al espacio formado por posiciones junto con las propiedades que las describen. Originalmente cada solución resultó en un sistema de información que administra datos espaciales desarrollado en base a una tecnología, jurisdicción y empresa, y generalmente no interoperable con otros sistemas. En las últimas décadas, con la explosión de los sistemas distribuidos y las redes globales, ha crecido el interés por facilitar la interoperabilidad, reutilización e integración de datos espaciales. Para muchos sistemas, incorporar tecnologías con gestión de información geográfica resulta un valor agregado que no había sido considerado previamente, como es el caso de los servicios basados en ubicación.

La Web Semántica se relaciona con la noción de interoperabilidad. Se pretende disponer en la Web de conocimiento formal y explícito que pueda ser interpretado por agentes de software. En particular, se plantea la necesidad de una *Web Semántica Geoespacial*. La demanda es por mejorar la disponibilidad de información y servicios geográficos. Contar con la posibilidad de compartir datos y servicios implica disminuir costos y tiempos de puesta en funcionamiento, así como optimizar la confiabilidad de la información. Pa-

ra ello se necesitan estándares que provean el marco para la definición de fenómenos y servicios geográficos y unifiquen los modelos de representación. Compartir información no solo significa proveer acceso a los datos, también se debe brindar *conocimiento* para que los datos puedan ser “interpretados”. Una forma de representar conocimiento es por medio de ontologías. Las ontologías en Ciencias de la Computación sirven como medio para especificar formalmente el significado que se le asigna a los conceptos que modela un programa o sistema. La especificación formal y explícita permite que el conocimiento pueda ser compartido y reutilizado. En esta tesis se estudian los modelos de representación para información geográfica, los resultados en estandarización sobre representación de fenómenos geográficos, y las necesidades y posibles soluciones usando ontologías para describir semánticamente la información geográfica, encausados a mejorar la disponibilidad de entornos distribuidos heterogéneos.

Abstract

First automated systems working with geographic information date back to the '70s. Since then, several applications have evolved to satisfy the needs of different areas, like meteorology, urban planning, transport, environmental services, tourism, economic distribution, etc. Among its users are state and governmental organizations, educational and research institutions, industry and commerce.

Finding ways for representing geographic information is a complex problem. Geographic data have references to one or more locations of the Earth's surface and eventually may require to include geometric properties. Models to represent geographic information about real-world phenomena are organized in one of two abstractions: discrete model, which considers the space and its objects, and continuous model, which imagines the space as formed by positions with properties that describe them. First information systems with the capacity of managing spatial data were restricted to a certain technology, jurisdiction and business rules, and in general were not interoperable with other systems of the same kind. In the last few decades, due to the explosion of distributed systems and global networks, a strong interest in providing interoperability, reuse and integration of spatial data have emerged. Nowadays, for many systems incorporating technologies with geographic information management is an added value that was not previously considered, like is the case of location-based services.

The Semantic Web is related to the notion of interoperability. The desire is to have formal and explicit knowledge in the Web so that it can be interpreted by software agents. Particularly, there is a need of a *Geospatial Semantic Web*. The new demand is to improve the availability of geographic information and services. Providing data and services sharing means lower costs and start-up times beside optimize the reliability of the information. This involves standards which provide the framework for defining services and geographic phenomena and unify the models of representation. Sharing of information not only means

providing access to data, but also provide *knowledge* to make the data “interpretable”. One way to represent knowledge is through the use of ontologies. In Computer Science ontologies are a tool to formally specify the meaning assigned to concepts that use a program or system. This formal and explicit specification allows knowledge to be shared and reused. This thesis focuses on models to represent geographic information, standards of representation for geographic features and needs and potential solutions using ontologies for describe the semantic of geographic information aimed at improving the availability of distributed heterogeneous geographic environments.

Índice general

Índice general	VII
Índice de figuras	XIII
Índice de cuadros	XVII
1. Introducción	2
1.1. Contexto	2
1.2. Objetivos	5
1.3. Propuesta de investigación	6
1.4. Organización	8
1.5. Resumen de las contribuciones	9
2. Fundamentos de Bases de Datos Geográficas	11
2.1. Niveles de abstracción	12
2.2. Modelos conceptuales para sistemas geográficos	14
2.2.1. Modelo de datos geográfico	14
2.2.2. Modelos para esquemas de aplicación geográficos	16
2.3. Modelos lógicos para información geográfica	24
2.3.1. Modelo mosaico	25
2.3.2. Modelo vector	25

2.3.3.	Representación de geometría para colecciones de objetos	26
2.3.4.	Evaluación de los modelos	29
2.4.	Bases de Datos Espaciales	30
2.4.1.	Tipos de Datos Abstractos Espaciales	31
2.4.2.	Operaciones espaciales	31
2.5.	Resumen del capítulo	34
3.	Sistemas de Información Geográfica	37
3.1.	La tecnología SIG	38
3.1.1.	Alcance y servicios	38
3.1.2.	Organización	39
3.2.	Información geográfica	41
3.3.	Caracterización de las aplicaciones SIG	42
3.3.1.	Aspectos funcionales	42
3.3.2.	Restricciones	45
3.4.	Visualización de mapas	45
3.4.1.	Formas de representación	46
3.4.2.	Los mapas interactivos	47
3.5.	Arquitectura de los SIG	49
3.6.	Áreas de aplicación	50
3.7.	Aplicaciones	52
3.7.1.	Sismos y Terremotos en el mundo	52
3.7.2.	TOXMAP: Mapas sobre salud ambiental	53
3.7.3.	MAPS: Herramientas de Análisis y Mapas sobre Seguridad Pública	54
3.7.4.	PROSIGA: Proyecto Sistema de Información Geográfica Nacional de la República Argentina	55
3.8.	Resumen del capítulo	57

4. Integración de Información Geográfica	59
4.1. Motivación	60
4.2. Interoperabilidad e Integración	61
4.2.1. Heterogeneidad de información	62
4.2.2. Integración de bases de datos heterogéneas	65
4.3. Integración de Información Geográficos	67
4.3.1. Caracterización de la información geográfica	67
4.3.2. Heterogeneidad en información geográfica	68
4.4. Solución a los problemas anteriores	74
4.4.1. Vocabularios y tesauros	74
4.4.2. Limitaciones de los tesauros	77
4.5. Resumen del capítulo	79
5. Estándares Internacionales para Sistemas de Información Geográfica	81
5.1. Conceptos generales	82
5.2. Organismos internacionales de estandarización	83
5.2.1. Open Geospatial Consortium (OGC)	84
5.2.2. Organización Internacional para la Estandarización ISO/TC 211	85
5.3. Estándares para información geográfica	87
5.3.1. ISO/TC211 19101 Modelo de Referencia	88
5.3.2. ISO/TC211 19109 Modelo conceptual estándar para Esquemas de Aplicación	90
5.3.3. ISO/TC211 19107 Modelo conceptual estándar para Fenómenos Geográficos	94
5.3.4. Otras normas de la familia ISO 19000	99
5.4. GML: El Lenguaje de Mercado Geográfico	101
5.4.1. Características de GML	102

5.4.2.	Descripción de GML schema	103
5.4.3.	Ventajas de GML	105
5.5.	Resumen del capítulo	107
6.	Ontologías para Información Geográfica	109
6.1.	Ontologías y representación de conocimiento	110
6.1.1.	¿Por qué definir ontologías?	111
6.1.2.	Uso de ontologías en Ciencias de la Computación	112
6.2.	Formalización	113
6.2.1.	Definición	113
6.2.2.	Componentes de una ontología	114
6.2.3.	Definición formal	115
6.3.	Tipos de ontologías	116
6.3.1.	Clasificación según generalidad	117
6.3.2.	Clasificación según expresividad	117
6.4.	Interoperabilidad e Integración para Información Geográfica	120
6.4.1.	Recuperación de IG basada en catálogos	120
6.4.2.	Recuperación de IG basada en ontología	122
6.5.	Diseño de ontologías para información geográfica	124
6.5.1.	Creación de ontologías	124
6.5.2.	Cohesión entre una aplicación y una ontología de dominio	126
6.6.	Sistemas de Información Geográfica conducidos por ontologías	128
6.6.1.	El ciclo de vida del sistema	128
6.6.2.	Inclusión de ontologías en la etapa de desarrollo	129
6.6.3.	Vinculación de un SIG en producción con nuevas ontologías	130
6.7.	Resumen del capítulo	131

7. Conclusiones y Trabajos Futuros	133
7.1. Resumen de la investigación	133
7.2. Resultados obtenidos	135
7.3. Publicaciones	137
7.4. Trabajo Futuro	138
A. Ejmplo GML	141
A.1. Definición GML del Esquema de Aplicacion Mapa Urbano	141
B. Abreviaturas y Acrónimos	147
C. Glosario	149
Bibliografía	157

Índice de figuras

1.1. El proceso de modelado	4
1.2. Niveles de Abstracción y Modelos de Representación	7
1.3. Integración de IG	7
2.1. Abstracciones en SMBD	13
2.2. Modelos Conceptuales Geográficos	14
2.3. El espacio según el modelo basado en campos	15
2.4. El espacio según el modelo basado en objetos	15
2.5. Jerarquía de abstracción espacial en MADS	19
2.6. Vista parcial de un modelo conceptual para un problema diseñado con MADS	20
2.7. Esquema de aplicación en MADS con objetos y relaciones espaciales	21
2.8. Clases espaciales Geo-campos o continuas	22
2.9. Clases espaciales Geo-objetos o discretas	22
2.10. Generalización en OMT-G	23
2.11. Agregación espacial en OMT-G	23
2.12. Modelos de representación	25
2.13. Dos representaciones del modelo mosaico	25
3.1. Composición de un SIG	40
3.2. Representaciones geográficas en el mapa	46
3.3. Google Maps - Buscar camino	48

3.4. Earthquake Hazards Program - Mapas sobre los terremotos de Chile 2010	53
3.5. SIG - TOXMAP	53
3.6. PROSIGA - Datos provistos por la Secretaría de Energía de la Nación	56
4.1. Dos modelos de datos para un mismo problema inmobiliario	64
4.2. Mapas Temáticos de Europa - variables cualitativas y cuantitativas	68
4.3. Definición del Thesaurus GEMET para el concepto “Wetland”	75
5.1. El proceso de modelado	88
5.2. Proceso de Transformación: de la realidad a los datos	89
5.3. Fragmento del Modelo de Fenómeno General	90
5.4. Fragmento del Esquema Conceptual de Aplicación “Ríos de Argentina”	91
5.5. Modelo de Objetos de GFM para el modelado de “Ríos de Argentina”	91
5.6. Jerarquía de Tipo de Atributo	93
5.7. Jerarquía de Tipo de Asociación	93
5.8. ISO 19107 - Paquetes y sus clases geométricas	95
5.9. Jerarquía de clases de geométricas	97
5.10. Paquetes Topológicos	98
5.11. Diagrama de Clases Topológicas	99
5.12. Primitivas topológicas	99
5.13. Relación entre geometría y topología	100
5.14. Ejemplo de esquema de aplicación Mapa Urbano	105
6.1. Ejemplo de Modelado de Conceptos y Relaciones	112
6.2. Composición de una ontología	113
6.3. Conceptualización y ontología	114
6.4. El entorno Cliente-Proveedores en el proceso de búsqueda de IG	120
6.5. Formas de usar ontologías para incluir conocimiento	123

6.6. Un ejemplo de ontologías de dominio 125

6.7. Asociaciones semánticas entre Fenómenos, Tipos de Fenómenos y Ontologías de Dominio 125

6.8. Vistas múltiples de un fenómenos geográficos 126

6.9. Integración entre aplicaciones usando ontología híbrida 126

6.10. Cohesión: Ontología de Aplicación y Ontología de Dominio 128

A.1. Ejemplo de esquema de aplicación Mapa Urbano 141

Índice de cuadros

1.1. Ejemplos de heterogeneidad entre SIG	6
2.1. Ejemplos de Objetos Espaciales	16
2.2. Ejemplos de aplicación para los tipos espaciales en MADS	20
2.3. Operaciones de los tipos región y línea	32
4.1. Heterogeneidad semántica en ontologías	71
4.2. Definición de las Relaciones entre Términos de un Tesauro	76
4.3. Relaciones entre término ZONA INUNDABLE y otros elementos del tesauro .	77
5.1. Modelos Abstractos OGC	86
6.1. Ejemplos de Ontologías	119

Capítulo 1

Introducción

1.1. Contexto

Se reúne bajo el nombre de *Sistema de Información Geográfico* (SIG)¹ al conjunto de herramientas que sirven para capturar, almacenar, analizar, administrar y presentar datos con referencia a una o varias ubicaciones sobre la superficie terrestre. Los requisitos principales que se consideran para la definición de los servicios esperados de un SIG son [Lua04, Mit99a]:

1. un SIG es una *colección de herramientas computacionales para realizar análisis geográfico y simulaciones* [RSV02, BM98, DJ06].
2. un SIG está basado en un *conjunto de estructuras de datos y algoritmos para representar, recuperar y gestionar información geográfica* [WD04, RSV02].
3. un SIG es una herramienta *que colabora con la toma de decisiones en tareas relacionadas con aspectos geográficos* [RSV02, LGMR05].

En la tecnología SIG se combina conocimiento de diversas áreas de las ciencias como Cartografía, Análisis Estadístico y Bases de Datos, entre otras. Los SIG son una tecnología de propósito general que incluye servicios para analizar modelos, localizar eventos, medir distancia entre eventos, encontrar caminos y explorar las relaciones entre dos o más

¹En Inglés *Geographic Information System* (GIS).

objetos. Los campos de aplicación de esta tecnología son diversos, y entre ellos se pueden citar cartografía, sensado remoto, distribución de superficie, manejo de distribución política, análisis sociológico, distribución de recursos naturales, planeamiento urbano y localización.

En un SIG las piezas de información se organizan por *temas* [Cor00, ASR⁺05, LGMR05]. Un ejemplo de tema podría ser “Hidrografía de Argentina” con datos sobre ríos, lagos, lagunas; así como sus ubicaciones, caudales, cuencas, cauces, etc. Tradicionalmente, en el campo de la geografía, se utilizó la estrategia de organizar la información por temas. Utilizar esta misma abstracción en la representación digital de información geográfica (IG) favorece el modelado y los servicios de visualización. Desde la visión del modelado, permite dividir y concentrar las necesidades de representación a un conjunto acotado de fenómenos reales y sus vínculos. Desde la visualización, la opción de filtrado por tema da libertad a los usuarios para decidir dinámicamente qué temas mostrar y qué temas ocultar creando distintas representaciones visuales del mismo espacio geográfico. Por ejemplo, de la superposición de los temas “Hidrografía de Argentina” y “Caminos y rutas de acceso de Argentina” se puede responder a la consulta ¿cómo llegar desde Bahía Blanca al lago Nahuel Huapi?

La unidad de representación en IG se denomina *fenómeno geográfico* o simplemente *fenómeno*²[MFW90, SMB⁺97]. Cada fenómeno es único y distinguible; éste referencia a una locación, espacio, área o región, y se puede visualizar en un mapa. Los fenómenos geográficos se organizan en dos abstracciones: *objetos discretos*, por ejemplo, un monumento, una montaña o un lago, y *objetos continuos*, por ejemplo, elevación de la superficie terrestre o el nivel de lluvias. Paralelamente, también existen dos formas de representar digitalmente los datos geográficos: el *modelo mosaico* o raster y *el modelo vector*[LT92, SCG⁺97, SC03].

En el proceso de modelado de un tema se identifican los fenómenos de interés para un *contexto de aplicación*, se los organiza en patrones o *tipos de fenómenos* y se caracterizan por sus atributos y relaciones. Los atributos pueden ser atributos alfanuméricos y/o atributos espaciales. Finalmente, la instanciación de los fenómenos reales a los tipos de fenómenos y relaciones materializa la base de datos de la aplicación (figura 1.1). Los atributos espaciales tienen asociada una componente geométrica, que determina la forma y dimensión del fenómeno. También se identifican las relaciones espaciales geométricas y/o

²En Inglés *feature*.

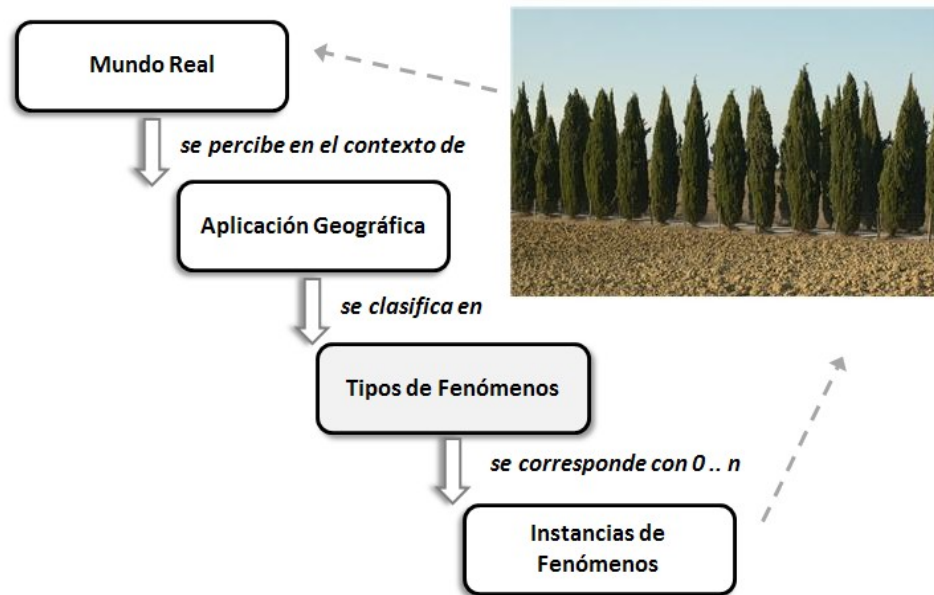


Figura 1.1: El proceso de modelado

topológicas de interés. Las propiedades métricas de objetos geográficos incluyen longitud y superficie (dependen de la dimensión del objeto) y relaciones métricas entre objetos como distancia y/o relaciones de orientación (“sobre”, “a la derecha de”, etc.). Las operaciones topológicas hacen referencia a propiedades de vecindad, adyacencia, inclusión y conectividad y permanecen invariantes ante cambios morfológicos, de escala o de proyección. Las representaciones usando los modelos vector o mosaico difieren en cuanto a la eficiencia de representación y la capacidad para resolver consultas espaciales. La elección de una u otra implementación depende principalmente del problema. En general los SIG proveen ambos modelos de representación.

Las calidades “interoperabilidad” e “integración” en varias oportunidades se utilizan indistintamente, aunque en un sentido estricto son distintas [GJM02]. En Ingeniería de Software, la calidad de *interoperabilidad* se relaciona con la capacidad de un sistema para integrar componentes de software que provienen de diferentes desarrolladores. Esto permite a un nuevo producto ingresar al mercado a competir con otros y ofrecer software integrable con otros módulos. Desde el modelo de datos, la calidad de *integración* se refiere a la posibilidad de compartir información, esto es, los datos y su significado. Cuando la información es producida y gestionada localmente, en general, resulta precisa en el ámbito

de su comunidad de uso. Sin embargo, cuando la misma ha sido producida por otros esto deja de ser cierto, y se requiere de conocimiento y de reglas de transformación para integrar información “heterogénea” o de fuentes no relacionadas.

En la última década está creciendo la necesidad de interoperabilidad e integración en distintos ámbitos de aplicación de IG: SIG, Infraestructuras de Datos Espaciales (IDE)³ y Sistemas Móviles Basados en Ubicación (LBMS)⁴. Interoperabilidad e integración son calidades que ganan importancia en el campo de los productos tecnológicos de información geográfica. La integración en las aplicaciones que trabajan con IG presenta una dificultad adicional: al momento de intercambiar información no solo aparecen diferencias a nivel de modelos del dominio, sino también de modelos de representación para datos geográficos heredados desde sistemas de geoprocésamiento ya desarrollados y que carecen de estandarización en lo que hace a la representación de la IG.

Los problemas de *heterogeneidad en IG* se clasifican en niveles o vistas que impactan de distinta forma en cuanto a las dificultades de integración. La vista de *heterogeneidad sintáctica* considera que distintas soluciones pueden ofrecer diferentes formatos y modelos de datos (vector y mosaico) y hasta distintos sistemas de coordenadas para representar fenómenos idénticos. Se considera como nivel de *heterogeneidad estructural* en IG cuando los modelos de distintas soluciones representan los mismos fenómenos con distintos atributos (estructuras distintas). El nivel de *heterogeneidad semántica* en IG se establece cuando distintas soluciones interpretan significados distintos para los mismos fenómenos del mundo a modelar. El cuadro 1.1 ejemplifica dichos niveles de heterogeneidad. Las dificultades de heterogeneidad sintáctica son más simples de resolver usando sistemas de transformación automáticos. Una solución parcial a los problemas de heterogeneidad estructural y semántica es contar con tesauros que normalicen los términos del dominio de aplicación. La integración semántica de IG ha motivado importantes investigaciones en el área de Ciencias de la Computación. Las soluciones a problemas de integración semántica generalmente se vinculan con la definición de ontologías que incluyan especificaciones formales para los modelos de representación de cada solución.

³En Inglés *Spatial Data Infrastructure* (SDI).

⁴En Inglés *Local Based Mobile System* (LBMS).

Ejemplos		
Heterogeneidad	Sistema S_1	Sistema S_2
<i>sintáctica</i>	S_1 representa zonas por densidad de población de Bahía Blanca bajo el modelo vector.	S_2 representa zonas por densidad de población de Bahía Blanca bajo el modelo mosaico.
<i>estructural</i>	S_1 representa zonas por densidad de población de Bahía Blanca, incluyendo detalle de los servicios públicos.	S_2 representa zonas por densidad de población de Bahía Blanca, incluyendo detalles sobre los tipos de construcción (edificios, casas, barrios cerrados, etc.).
<i>semántica</i>	S_1 representa densidad de población de Bahía Blanca considerando número de habitantes.	S_2 representa densidad de población de Bahía Blanca considerando unidades familiares.

Cuadro 1.1: Ejemplos de heterogeneidad entre SIG

1.2. Objetivos

El objetivo principal de este trabajo es “estudiar *los modelos de representación para IG* con el fin de *establecer las necesidades de descripción semántica* que permitan avanzar hacia una mejora en la disponibilidad de la IG en entornos heterogéneos distribuidos”.

Como parte de este objetivos se identifican las siguientes metas parciales que colaborarán con el resultado final:

1. Estudiar las particularidades de representar IG.
2. Estudiar los modelos de representación para IG existentes, organizándolos según su nivel de abstracción.
3. Identificar las necesidades y las dificultades en integración de IG.
4. Releva el estado actual del arte en estandarización de IG y las posibilidades de integración basada en estándares.

5. Relevar el estado actual del arte en definición de ontologías de IG y las posibilidades de integración semántica.

1.3. Propuesta de investigación

Como se mencionó en la sección anterior, el objetivo de esta tesis es el estudio de los modelos de representación de IG con el propósito de añadir conocimiento semántico a los mismos. Inicialmente, el tema de investigación plantea dos desafíos: existen diversas formas de representación para la IG y de las que no se tiene un conocimiento acabado y por otra parte, el dominio de aplicación y las necesidades de los SIG son desconocidos.

Como inicio de la investigación se abordarán estos dos temas. Para los modelos existentes, se utilizará la estrategia usada por las tecnologías de bases de datos y se identificarán las vistas abstractas para los datos geográficos: física, lógica y conceptual. Se hará un relevamiento de los modelos actuales, analizando sus características y restricciones [Ull88, RSV02, PSZ⁺98, RJB98, RBP⁺91]. En la figura 1.2 se presentan gráficamente las piezas en las que se organizan estos temas de estudio.

A partir del conocimiento básico sobre representación de IG, se hará un relevamiento de las características y requerimientos de los SIG. Para ellos se estudiarán las tecnologías existentes, arquitecturas, representación de abstracciones geográficas, los SIG como una herramienta de propósito general para problemas con datos espaciales, ámbitos de aplicación y algunos ejemplos de uso [Zei99, Cla86, BM98].

Las tecnologías SIG comenzaron a utilizarse en década del '70. Las características especiales de la IG, como mantener referencias con una o más locaciones de un espacio subyacente y eventualmente requerir mantener propiedades geométricas, hicieron que originalmente, cada propuesta de implementación eligiera soluciones diferentes y propias. Esto impacta al momento de buscar compartir datos o servicios entre sistemas. En la década del '90, e impulsados por las necesidades de nuevas tecnologías emergentes (LBMS, IDE), organizaciones internacionales comienzan a trabajar fuertemente en definir estándares para sistemas geográficos [OGC94, ISO94, CEN07]. Se hará un relevamiento de las necesidades de integración, resultados en lo relacionado con recomendaciones y estándares para gestión de IG y limitaciones. Finalmente, dada la tendencia actual, para enriquecer semánticamente la información y analizar el impacto en cuanto a capacidades de integración de información, se estudiará el uso de las ontologías aplicado al ámbito de aplicaciones

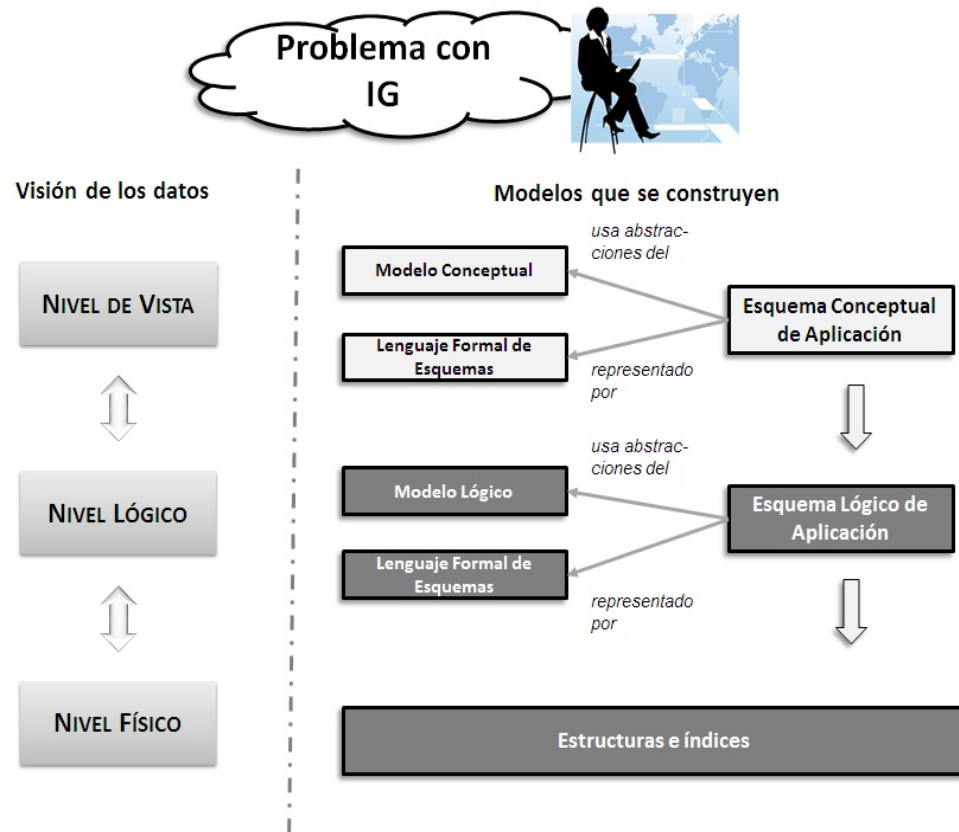


Figura 1.2: Niveles de Abstracción y Modelos de Representación

con IG [Gru93, Gua98, WVV⁺01, Usc98]. En la figura 1.3 se esquematiza cómo se organiza la definición de estándares y ontologías para SIG en relación con las posibilidades de integración.

1.4. Organización

La estructura de esta tesis es la siguiente:

En el *Capítulo 2* se describen los modelos de representación para la IG. La IG es un tipo particular de información con exigencias comparables a las de cualquier otra información que requiera de persistencia. En el área de bases de datos se usa la estrategia de identificar distintos niveles de abstracción para los modelos de datos. Así se identifica el *nivel conceptual* o de mayor abstracción. Este nivel provee los constructores para identificar los elementos del dominio que se desea modelar y cómo se relacionan. El *nivel*

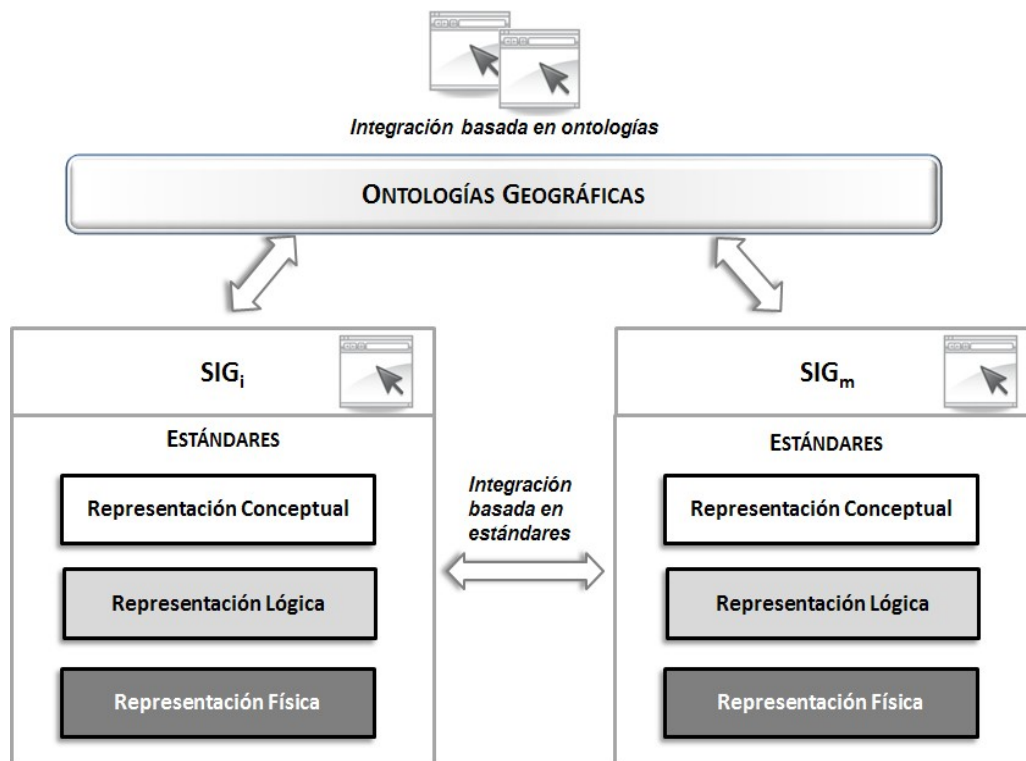


Figura 1.3: Integración de IG

lógico o discreto es un nivel de abstracción intermedia. Éste provee los constructores para representar los esquemas de aplicación en la computadora. Finalmente el *nivel físico* o menos abstracto es el que se ocupa de las estructuras de bajo nivel para administrar y optimizar el acceso a los datos almacenados. En este capítulo se caracterizan los requerimientos específicos de la IG en los distintos niveles y se recopilan y describen los modelos de datos para IG existentes.

En el *Capítulo 3* se hace un estudio de las caracterizaciones y requerimientos de los SIG y se identifican sus componentes. En los '70 cuando surgen las primeras aplicaciones SIG, bases de datos, lenguajes de programación y paradigmas de desarrollo no estaban preparados para representar información con requerimientos espaciales. Así cada proveedor definió sus propios modelos y estrategias de representación. Las áreas de aplicación para los SIG son amplias y siguen creciendo. Asimismo, son diversas las necesidades de información y consulta. Entre los dominios de aplicación se incluyen economía, gestión ambiental, turismo, comunicaciones, demografía y organización política entre varios otros. Éstas y otras áreas pueden tomar ventajas del conocimiento de la información vinculada

al espacio de referencia. Por esta razón los SIG se han hecho populares en el ámbito privado, público y de investigación y educación. Al final de este capítulo se incluyen a modo de ejemplo algunas aplicaciones de variados dominios de interés.

En el *Capítulo 4* se presentan los conceptos de *integración e interoperabilidad* relacionados con aplicaciones que manejan IG. Se clasifican y estudian los niveles de heterogeneidad en la información en general y en particular los relacionados con IG. Varias fuentes de datos SIG heterogéneas acumulan IG con datos espaciales en diversos formatos. El costo que significa para un SIG contar con IG y el hecho de que cierta información de base, como por ejemplo la distribución política de una región exista, hacen que haya crecido el interés en clientes y proveedores por compartir datos y servicios. Para compartir IG se trabaja en definir normas y estándares relacionados con la representación de información en los distintos niveles de abstracción.

En el *Capítulo 5* se recopila el trabajo en estandarización de los dos organismos internacionales que trabajan en normalización geoespacial: el consorcio “Open Geoespacial Consortium” (OGC) y el Organismo Internacional de Estandarización (ISO) desde su comité Técnico ISO/TC 211. Ambas organizaciones vienen trabajando desde el año 1990, y lo siguen haciendo en la actualidad, en lo que respecta a la definición de estándares para favorecer la geodisponibilidad de la información y de los servicios geográficos. Actualmente trabajan cooperativamente, ISO/TC 211 definiendo las normas y estándares de representación y OGC especificando estándares implementables que respeten las normas ISO. Los resultados obtenidos hasta el momento son aceptados en el dominio de IG y han incrementado las posibilidades de compartir información y servicios.

En el *Capítulo 6* se estudian a las ontologías como medio para enriquecer semánticamente a las aplicaciones geográficas. Las ontologías buscan unificar la representación para símbolos y términos de forma tal que dicha interpretación sea única y además pueda ser “leída” por agentes automáticos. Existen diversas ontologías y entre ellas difieren en el nivel de detalle que expresan. Más aún, una buena práctica es organizar los términos en distintas ontologías según el nivel de detalle. Esto permite contar con ontologías generales y específicas. Una o más ontologías definidas para un dominio de aplicación común a varios SIG podrían ser las mediadoras para la integración de información. En este capítulo se va estudiar el estado actual del arte en el uso de ontologías para SIG.

En el *Capítulo 7* se cierra este trabajo de investigación. Se presentan los principales logros alcanzados, conclusiones y trabajos futuros.

Se incluyen además Apéndices con el desarrollo de un ejemplo en GML, las abreviaturas y acrónimos usados y un glosario de definiciones. Finalmente se ubican las referencias bibliográficas.

1.5. Resumen de las contribuciones

Como resultado de este trabajo de investigación se esperan realizar las siguientes contribuciones:

1. Recopilar y organizar el estado actual del arte en modelos de representación para IG. La información geográfica es naturalmente compleja y variada. Para administrar su complejidad se han definido diversos modelos con diferentes grados de representación. Sin embargo, el material existente sobre este tópico es disperso y en general poco relacionado. Se espera contribuir en su organización.
2. Hacer un análisis exhaustivo de los problemas de integración de IG y las soluciones actuales. Las exigencias por integrar información y servicios geográficos han tomado impulso como resultado de las tecnologías emergentes para IG. Las características particulares de los fenómenos geográficos y las heterogeneidades de representación exigen analizar los nuevos desafíos de integración que se presentan.
3. Recopilar los estándares en IG e identificar los aportes de los mismos con relación a interoperabilidad e integración de IG. Las primeras tecnologías SIG datan de los '70. Originalmente cada proveedor determinó sus propios modelos y estructuras de representación. La falta de unificación en los modelos es la primera dificultad a sortear para avanzar en las calidades de interoperabilidad e integración. Desde la década del '90 organizaciones internacionales vienen trabajando en definir estándares para datos y servicios geográficos. En este trabajo se recopilarán los principales resultados obtenidos.
4. Recopilar y organizar el estado actual del arte en ontologías para IG y soluciones propuestas. Sin la definición de estándares de representación la comunicación entre clientes y proveedores heterogéneos es casi imposible. Sin embargo, esta única solución no es suficiente si se desea avanzar en la automatización del proceso de integración. Cada modelo de datos de un problema tiene conocimiento semántico

implícito que no se puede expresar en el diccionario de datos. Un medio emergente para representar conocimiento semántico son las ontologías. En este trabajo se van a investigar las posibilidades de uso de ontologías para modelar conocimiento semántico sobre IG.

5. Identificar problemas abiertos y trabajos futuros. Finalmente, como resultado de este trabajo se espera identificar problemas abiertos a la investigación que generen nuevos desafíos y líneas de investigación.

Capítulo 2

Fundamentos de Bases de Datos Geográficas

Se denominan *Sistemas de Información Geográfica* (SIG) a los sistemas usados para recolectar, analizar y presentar información que describe propiedades físicas y lógicas del mundo de la geografía. En un sentido amplio, el término SIG se aplica al conjunto de procedimientos manuales o asistidos por la tecnología que sirven para gestionar datos con referencias geográficas [SMB⁺97] o datos espaciales¹. Los datos geográficos están vinculados con una locación en la superficie terrestre.

En sus orígenes, el manejo de IG se limitó a procesos manuales no interactivos con representación en el plano bajo la forma de *mapas*. Con el desarrollo de tecnologías para la recolección y digitalización de datos geográficos surgió la necesidad de proveer algún tipo de aplicación que permitiera el manejo y análisis interactivo de los mismos. Esto dio lugar a aplicaciones que se conocen como *aplicaciones geográficas* o *SIG*. Un SIG [RSV02] es más que una herramienta cartográfica para producir mapas, es un tipo de software de aplicación que incluye, entre otras, capacidades para manipular datos geográficos, provee herramientas para análisis espacial y ofrece mecanismos de visualización e interacción que permiten que usuarios expertos del dominio analicen la información.

Debido al incremento en el volumen de los datos geográficos, una de las tareas de un SIG es también administrar eficientemente grandes bases de datos con IG. Las bases de

¹En esta tesis se refiere a los términos dato espacial y geográfico indistintamente. En un sentido estricto, *dato espacial* es más general. Refiere datos con referencia a un plano, que puede ser la Tierra u otro.

datos espaciales se ocupan específicamente de la gestión de este tipo de información con referencia a un plano. En representación de datos, y usando el principio de abstracción, se definen distintos modelos de datos para manejar representaciones de datos con diversos grados de detalle. Los modelos de datos utilizados por los sistemas de información (SI) para aplicaciones comerciales no están preparados para representar y administrar IG. Ni los modelos conceptuales de datos tradicionales como *Modelo Entidad–Relación* (MER) [Che76] o los modelos orientados a objetos como el *Modelo de Clases de UML* [RJB98]; ni los modelos lógicos como el *relacional* [Cod70] cuentan con las estructuras adecuadas para representar entidades y relaciones que se ajusten naturalmente a problemas geográficos. La representación de IG en una computadora requiere de nuevas abstracciones como punto, curva o superficie y colecciones de éstos. Las operaciones provistas por los modelos lógicos tradicionales también resultan incompletas para resolver consultas como distancia, dirección o intersección. Finalmente las representaciones físicas típicas tampoco son adecuadas para manipular de forma eficiente datos geográficos almacenados.

En este capítulo se recopilan y se describen los requerimientos para representar IG en los distintos niveles de abstracción conceptual, lógico y físico bajo la premisa de cubrir las necesidades de diversas aplicaciones SIG. Se presentan los principales modelos existentes para cada uno de los niveles, sus componentes, capacidades y limitaciones. Se presentan los sistemas de manejo o gestión para Bases de Datos Espaciales como el componente de la arquitectura de un SIG responsable de la representación, recuperación y consulta de IG.

2.1. Niveles de abstracción

Para proveer manejo eficiente de grandes volúmenes de información almacenada las bases de datos se apoyan en estructuras complejas. Tradicionalmente, los *Sistemas de Manejo de Bases de Datos (SMBD)*² [UII88, EN00, GMPQ⁺04, SKS10] han utilizado el recurso de definir *niveles* o *vistas abstractas* sobre los datos para lograr cierta independencia entre modelos. Esto permite ocultar detalles de implementación y simplificar las interfaces a los usuarios de bases de datos. La figura 2.1 muestra los tres modelos de abstracción tradicionalmente usados.

²En Inglés *Database Management System* (DBMS).

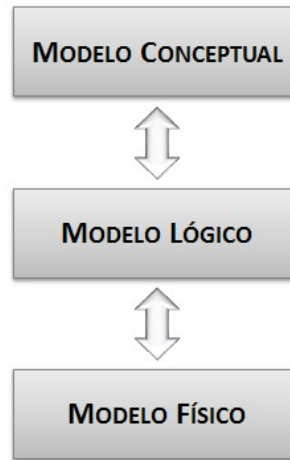


Figura 2.1: Abstracciones en SDBD

El *modelo conceptual* o abstracto se utiliza para identificar componentes del mundo real. El modelo conceptual no es directamente representable en la computadora porque ignora las limitaciones computacionales para expresar tamaño, continuidad o precisión. El *modelo lógico* o discreto provee los constructores para trasladar el modelo conceptual a las estructuras usadas a nivel de desarrollo. El *modelo físico*, por su parte, define cómo se almacenan los datos. Los modelos del nivel físico incluyen estructuras de datos complejas y de bajo nivel.

Los modelos de datos para SI tradicionales son insuficientes para modelar IG. Modelos conceptuales históricos como el MER carecen de constructores para describir ciertas propiedades espaciales típicas de las entidades geográficas como forma o relaciones y/o restricciones espaciales: distancia, inclusión, área y adyacencia entre otras. De la misma forma, las transformaciones al modelo lógico son incompletas en los mismos aspectos. El modelo relacional tradicional ampliamente utilizado para SI provee un conjunto de tipos de datos básicos (cadena, fecha, número) insuficiente. Estos problemas de representación obviamente tampoco han sido considerados en el modelo físico.

A continuación a modo de ejemplo plantean se los requerimientos para un problema que requiere de entidades geográficas y relaciones espaciales.

Ejemplo 2.1.1 *Plantear el diseño para un problema que modele la organización administrativa de una región. Se requiere representar a las entidades PROVINCIAS junto con*

sus propiedades nombre, extensión, número de habitantes y forma (representando la ocupación geográfica). Se definen además las entidades CIUDADES, con atributos nombre y ubicación sobre la provincia y PAISES con atributos nombre, sistema de gobierno y forma. Entre ellas se establecen relaciones espaciales de inclusión, distancia y área ocupada.

Dificultades de modelar el problema con los modelos tradicionales.

- No contar con tipos de datos para representar la geometría de las provincias.
- No existen constructores para relaciones espaciales como Bahía Blanca “está incluida” en la provincia de Buenos Aires, ó Buenos Aires “limita con” La Pampa.
- Responder a consultas como “¿qué distancia separa a Bahía Blanca de Mar del Plata?” no es simple.

En las próximas secciones se presentan los principales modelos desarrollados para representar IG. El proceso de modelado se basa en fundamentos matemáticos rigurosos fuera del alcance de esta tesis. A modo de referencia las entidades se localizan sobre un *espacio* de interés \mathfrak{R}^d y la función *distancia euclidiana* al que se denomina *Espacio Euclídeo*. Los *puntos* son elementos del espacio, asociados a un par de coordenadas cartesianas x e y . Las entidades geográficas a través del proceso denominado *proyección al mapa* son mapeadas a una representación plana.

En lo que sigue siempre que se hable de “entidad espacial” se está refiriendo indistintamente a una entidad geográfica siguiendo el tema de interés de esta tesis. En un sentido estricto, el término “espacial” implica un alcance más general, una instancia espacial está vinculada a una ubicación en un plano, que puede ser la Tierra o cualquier otro, como por ejemplo el plano de diseño de un circuito electrónico. Los datos geográficos están específicamente vinculados con una locación en la superficie terrestre.

2.2. Modelos conceptuales para sistemas geográficos

En la representación de problemas geográficos se distinguen **dos planos conceptuales** de abstracción [MGR99, RSV02, Lua04]: un modelo conceptual para el “espacio geográfico” o **modelo de datos geográfico** y un modelo conceptual para especificar el “esquema de datos” o **modelo de esquema de aplicación geográfica** (figura 2.2). A continuación se presentan las alternativas para cada uno de ellos.

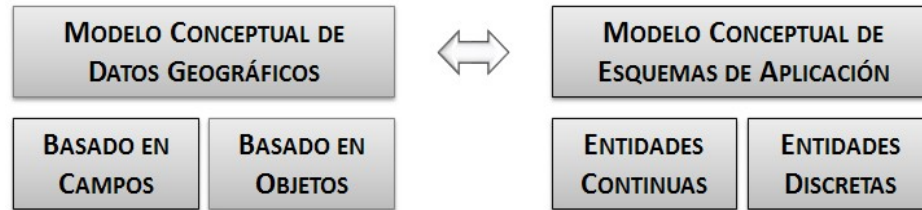


Figura 2.2: Modelos Conceptuales Geográficos

2.2.1. Modelo de datos geográfico

La interpretación del espacio geográfico depende de la semántica del problema. Por ejemplo, dado el territorio de la provincia de Buenos Aires, en un problema administrativo, la provincia se particiona en municipios; desde una visión geológica, el espacio se organiza en áreas geológicas; en control del tráfico, el foco se ubica en la red de caminos. Cada problema define una interpretación del espacio y la colección de entidades necesarias para describirlo. La organización de las entidades sobre un tópico particular en un SIG se denomina *tema*.

Las tecnologías SIG reconocen dos formas diferentes de representación para el espacio subyacente [SCG⁺97]: el *modelo espacial basado en campos*³ y el *modelo espacial basado en objetos*⁴. En la aproximación espacial basada en campos cada punto del espacio está asociado con uno o varios valores de atributos definidos como funciones continuas sobre x e y . Por ejemplo, altitud sobre el nivel del mar, temperatura, etc. Las mediciones sobre los distintos fenómenos se reúnen como valores de atributos variando con la ubicación en el plano. La figura 2.3 muestra la representación en un mapa para el modelo espacial basado en campos. La interpretación del espacio como un campo continuo es lo que contrasta con el modelo basado en objetos.

En la aproximación basada en objetos el espacio está poblado de *entidades discretas* con identificación. Cada entidad tiene asociada un *componente espacial*, que se corresponde con la forma y ubicación del objeto en el espacio subyacente, y un conjunto de propiedades descriptivas. En la figura 2.4 se muestra el mapa de la provincia de Buenos Aires bajo esta representación; se pueden observar las entidades discretas RUTAS representadas como “líneas” y CIUDADES que están simbolizadas con “puntos”.

³En Inglés *Field-based model*.

⁴En Inglés *Object-based model*.

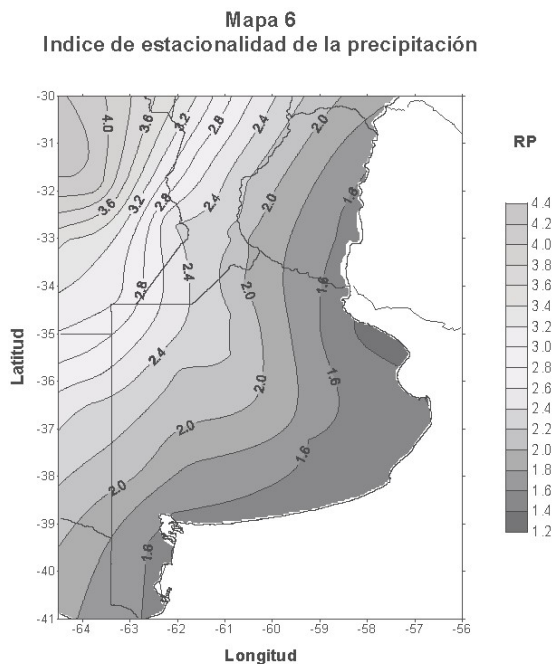


Figura 2.3: El espacio según el modelo basado en campos

Los modelos basados en objetos usan abstracciones para representar las distintas clases de objetos geográficos. La elección de alguna abstracción depende de las dimensiones del objeto geográfico *punto*(0–dimensión), *curva*(1–dimensión), *superficie*(2–dimensión) ó *volúmen*(3–dimensión). Los puntos son usados para representar entidades cuya forma no es de interés o es despreciable, pero sí se quiere conocer su ubicación. Los objetos de tipo líneas comúnmente se utilizan para representar redes. El tipo geométrico generalmente se conoce como *polilínea*. Una polilínea se define como un conjunto finito de *arcos* o *segmentos*, de modo que cada extremo es compartido exactamente por dos segmentos a excepción de los puntos extremos, que pertenecen a un único segmento. Los objetos de dimensión dos se usan para representar áreas y su tipo geométrico es el *polígono*. Un polígono es una región del plano limitada por una polilínea cerrada.

El cuadro 2.1 muestra algunos de los usos clásicos de estas abstracciones. La elección del tipo de abstracción no es fija y depende del uso que se dará a la colección de entidades. Distintos factores pueden influenciar en la decisión, uno especialmente importante es la *escala* de interés. Por ejemplo, un aeropuerto puede representarse como un punto, si interesan representar líneas aéreas, o como un área, si el foco es la organización interna

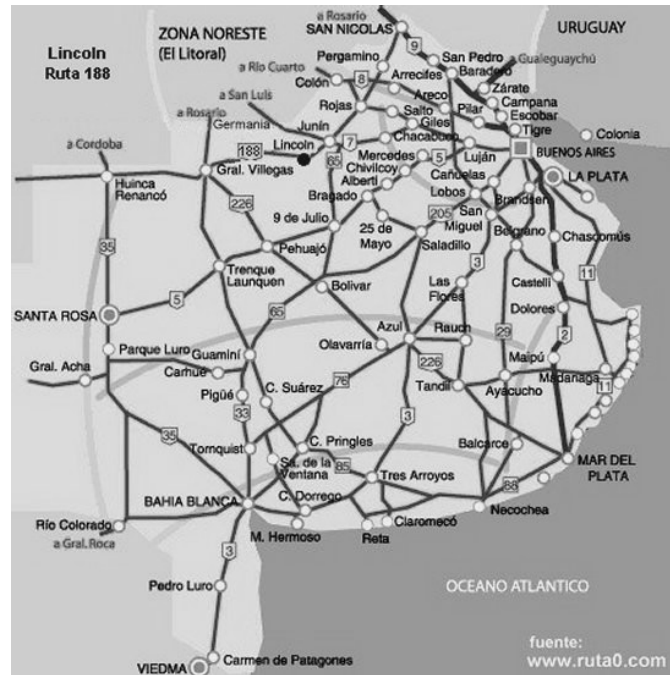


Figura 2.4: El espacio según el modelo basado en objetos

del aeropuerto.

2.2.2. Modelos para esquemas de aplicación geográficos

El término *esquema de aplicación* en el área de bases de datos se utiliza para referir al diseño de la estructura estática que mantiene la información, mientras que se denomina *instancia de base de datos* al conjunto de valores ajustados a un esquema en un instante de tiempo. Las instancias son dinámicas y se espera que el contenido de la base de datos evolucione constantemente en el tiempo; los esquemas son estáticos y se desea que permanezcan invariantes.

El modelo conceptual para especificar bases de datos espaciales debe permitir a los diseñadores definir esquemas de aplicación para SIG con una variedad de necesidades. En SI tradicionales este rol es cubierto por el MER [Che76] o los Diagramas de Clases del Dominio de UML [RJB98], pero resultan insuficientes para aplicaciones espaciales. Un modelo conceptual para representar problemas del área de la Geografía debe ser capaz de representar:

Dimensión	Nombre	Aplicación
0-Dimensión	<i>punto</i>	Ubican entidades cuya forma no es considerada relevante o la superficie que ocupan es muy chica con relación al espacio. Ejemplos: ciudades, centros comerciales, etc.
1-Dimensión	<i>línea</i>	Se usan para representar entidades con forma de redes o grafos. Ejemplos: ríos, caminos, etc.
2-Dimensión	<i>superficie</i>	Representan entidades con área. Los polígonos son el principal tipo geométrico para representar tales objetos. Ejemplos: regiones, secciones, etc.

Cuadro 2.1: Ejemplos de Objetos Espaciales

- Abstracciones de alto nivel para modelos espaciales basadas en objetos o en campos.
- Entidades tradicionales y georeferenciadas.
- Tipos de atributos usados por aplicaciones geográficas incluyendo abstracciones geométricas.
- Relaciones espaciales y restricciones de integridad espacial.

El modelo conceptual de esquemas de aplicación provee las abstracciones para representar a los elementos del mundo real. En orientación a objetos a esta abstracción se la denomina *objeto*, mientras que en el MER se la identifica como *entidad*. En los modelos conceptuales para IG recibe el nombre de *fenómeno* y es lo que identifica a un objeto geográfico del mundo real distinguible de otros. Por ejemplo, el parque nacional “Los Arrayanes” es un fenómeno. Los fenómenos se caracterizan por atributos que pueden ser *atributos espaciales* y/o *atributos descriptivos* o *alfanuméricos*.

Ejemplo 2.2.1 Fenómeno

Parque Nacional Los Arrayanes = ((nombre, “Los Arrayanes”), (año-fundación, “1971”), (ubicación, instancia de *GM_Surface*⁵), (descripción, “Unidad boscosa

⁵Instancia de *GM_Surface* sobre el extremo norte del Lago Nahuel Huapi.

compuesta por arrayanes. El Parque Nacional Los Arrayanes conserva una muestra de la eco-región bosque patagónico, cuyo clima es templado a frío y húmedo, con nevadas y lluvias invernales. Se caracteriza por un paisaje de montaña de relieve abrupto y escarpado, con valles glaciarios”)).

En la literatura se pueden encontrar diversas propuestas para modelar esquemas con IG [SCG⁺97, KPS95, OPM97, MFW90]. Los beneficios de utilizar un enfoque de modelado conceptual que apoye el diseño de la base de datos son conocidos y ampliamente aceptados: diseñadores y usuarios pueden expresar e intercambiar su conocimiento acerca de la aplicación utilizando conceptos que les son familiares; independizándose así de cuestiones propias de la herramienta de software de implementación. En esta sección se incluye la presentación de dos de los modelos existentes para esquemas de aplicación geográficos: *Modeling Application Data with Spatio-temporal features* (MADS) [PSZ99, PSZ⁺98, PSZ06b, PSZ06a] y OMT–G [BLD99, KABL01]. Ambos proveen soporte visual para representar información espacial y temporal, lo que los hace más claros e intuitivos para modelar. Se eligieron estos dos modelos por ser representativos de los tipos de modelos más usados en diseño conceptual de SI. Mientras que MADS se asemeja al MER tradicional, OMT–G se presenta como una extensión del Lenguaje de Modelado Orientado a Objetos OMT⁶ [RBP⁺91], que sustentó la definición del Lenguaje Unificado de Modelado UML [RJB98].

MADS: Modelado de aplicaciones de datos que incluyen fenómenos espacio temporales

MADS es un modelo conceptual creado para problemas que manejan datos espaciales y/o temporales diseñado respetando las calidades de simplicidad y poder expresivo. Una característica saliente de MADS es que se apoyó en el principio de ortogonalidad. Maneja los aspectos multidimensionales descomponiéndolos en vistas independientes, considerando por separado las diferentes dimensiones a modelar: estructuras de datos, espacio, tiempo y representación. Incluye una sintaxis visual y adopta el paradigma objeto–relacional y las definiciones normalizadas por Object Database Management Group (ODMG)⁷.

En MADS la espacialidad y la temporalidad se asocian con tipos de objetos, atributos, relaciones y agregados. El sistema de tipos cuenta con los tipos espaciales básicos: punto,

⁶En Inglés *Object-Modeling Technique*.

⁷<http://www.odbms.org/odmg/>

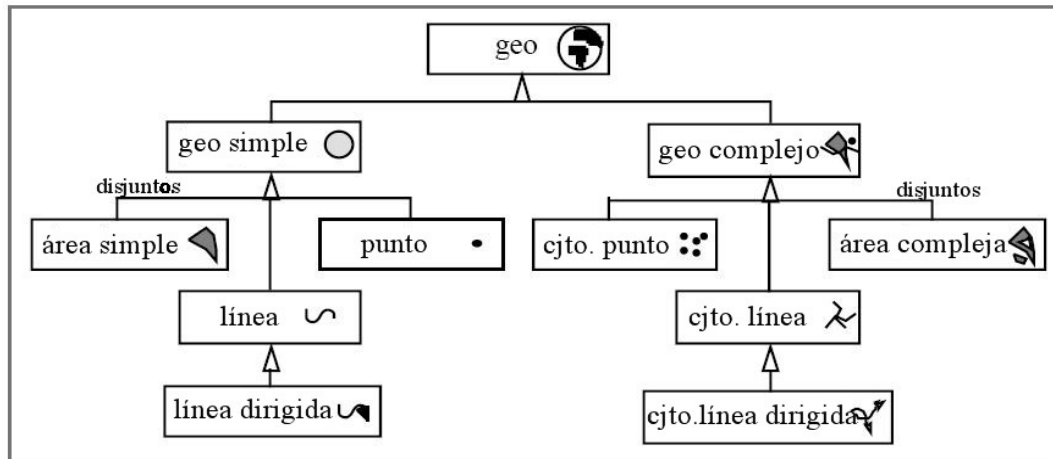


Figura 2.5: Jerarquía de abstracción espacial en MADS

línea, área y especializaciones de éstas. También incluye tipos espaciales para representar conjuntos homogéneos de objetos básicos: conjunto de puntos, conjunto de líneas, conjunto de líneas orientadas y conjuntos de áreas. Estos tipos espaciales conforman la jerarquía que se muestra en la figura 2.5. La raíz se corresponde con el tipo espacial “geo” del que desprenden “simple geo” y “complex geo”. También se puede observar que para cada tipo espacial la espacialidad se visualiza con íconos que expresan la información de manera no ambigua, visual y sintética.

Cada tipo espacial tiene asociado un conjunto de métodos que permiten definir y manipular instancias del tipo. Por ejemplo, el método *dimensión* está definido para tipo geo, mientras que *longitud* se define para línea. De manera similar MADS define una jerarquía de tipos para tipos de datos temporales [PSZ06b].

Ejemplo 2.2.2 Problema de Aplicación

Diseñar el modelo de datos para el problema de representar recorridos de las líneas de colectivo urbano.

El cuadro 2.2 muestra ejemplos de entidades y sus tipos espaciales según MADS para el problema del ejemplo.

Los objetos pueden tener espacialidad estática, si se les asigna alguno de los tipos espaciales hoja a su tipo de objeto, o indeterminada si se le asigna el tipo espacial geo.

Tipo de Objeto	Tipo Espacial	Descripción
PARADA	punto	Representa a fenómenos del tipo parada de colectivo.
PARADASLINEA	conjunto de puntos	Representa a fenómenos formados por todas las paradas de una línea de colectivos.
REDVIAL	conjunto de líneas	Representa a los fenómenos de tipo red de caminos.
CUADRA	línea dirigida	Mantiene a los fenómenos Cuadra. La dirección señala el sentido de numeración.
PARCELA	área simple	Mantiene fenómenos que representan a porciones pequeñas de terreno. Cada Parcela tiene un propietario
MANZANA	conjunto de áreas	Contiene fenómenos manzana como la composición de varias Parcelas.

Cuadro 2.2: Ejemplos de aplicación para los tipos espaciales en MADS

Para los objetos asociados a un tipo espacial genérico, el tipo espacial de cada instancia se define en tiempo de creación. La posición de un objeto también se puede definir en forma absoluta fijando las coordenadas del objeto, o relativa, definiendo la posición con respecto a otra ubicación conocida, por ejemplo, X está a $60km$ de Y , sobre la ruta Z .

Además MADS define *relaciones espaciales*, las que pueden ser topológicas, de orientación, métricas o agregados espaciales. Las relaciones espaciales están restringidas por la espacialidad de los objetos. Sin embargo, es importante contar con medios para definir explícitamente relaciones espaciales en el modelo conceptual, por ejemplo, para especificar que dos secciones de un mismo río pertenecen a la relación “*está adyacente a*”. Esto enriquece el esquema, permite nombrar a las relaciones, agregarles atributos y métodos y darles semántica adicional. Una relación espacial vincula al menos dos objetos espaciales.

Ejemplo 2.2.3 Vista de un modelo conceptual MADS

La figura 2.6 muestra una vista parcial del diseño en MADS para un problema. Del diagrama se puede interpretar que se consideraron los tipos de fenómeno PAIS, con un



Figura 2.6: Vista parcial de un modelo conceptual para un problema diseñado con MADS

atributo espacial de tipo conjunto de áreas, CIUDAD cuyo atributo espacial es de tipo punto y LAGO, con espacialidad de tipo área simple. Entre los tipos de objeto PAIS y CIUDAD se definió la relación espacial “contiene” por la cual se impone la restricción que una instancia de CIUDAD debe estar incluida en el espacio ocupado por la instancia de PAIS a la que pertenece. Existe otra relación topológica entre CIUDAD y LAGO que permite expresar si un fenómeno ciudad está en relación de “adyacencia” a un lago.

La figura 2.7 muestra un modelo conceptual completo para un problema diseñado usando MADS [PSZ06b]. Incluye atributos y relaciones espaciales y temporales. La misma figura contiene las referencias simbólicas. Se puede observar que el modelo es claramente interpretable por distintos usuarios: fenómenos que maneja, sus representaciones y las relaciones de asociación y espaciales definidas entre ellos.

En general se puede decir que MADS favorece el diseño de estructuras para una representación del espacio basada en objetos. Sin embargo también provee mecanismos para representar vistas continuas del espacio basadas en campos.

OMT–G: Lenguaje de Modelado Orientado a Objetos para fenómenos geográficos

OMT–G es un modelo conceptual para diseñar esquemas de aplicación que incluyan elementos geográficos. Provee las primitivas para modelar geometría y topología de datos geográficos, soporta estructuras topológicas agregadas bajo el modelo “todo-parte”, estructuras de red, múltiples vistas de objetos, y relaciones espaciales.

Incluye los conceptos básicos de clases, relaciones y restricciones de integridad espacial. Distingue entre *clases convencionales* o no espaciales y *clases espaciales* o *georeferenciadas* que pueden representar a datos geográficos continuos o discretos. Las clases convencionales describen a un conjunto de objetos con propiedades, comportamiento, relaciones

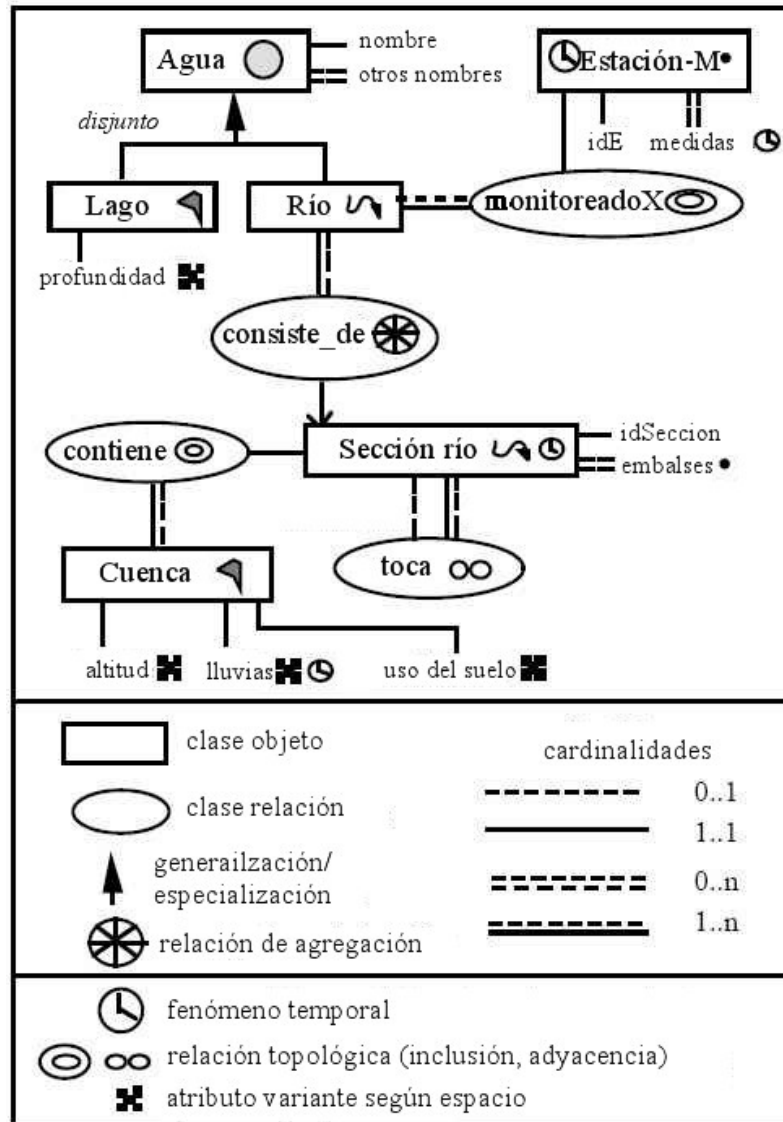


Figura 2.7: Esquema de aplicación en MADS con objetos y relaciones espaciales

y semántica similares, y que mantienen alguna relación con clases espaciales. Las clases georeferenciadas se especializan de los tipos de clases *geo-campos* y *geo-objetos*. Las clases *geo-campos* se usan para representar a los fenómenos continuos, mientras que las clases *geo-objetos* representan a objetos geográficos individuales particulares. Notar que esta discriminación permite modelar las dos vistas del espacio basado en campos y en objetos. Una clase georeferenciada se representa por un rectángulo subdividido en cuatro partes. El rectángulo superior izquierdo es utilizado para ubicar en forma simbólica la geometría de la clase. El resto de las secciones organiza los atributos geográficos, atributos descriptivos



Figura 2.8: Clases espaciales Geo-campos o continuas

y operaciones.

OMT-G propone un conjunto fijo de tipos geométricos y usa representación simbólica para distinguir las clases georeferenciadas de geo-objetos y de geo-campos (Figuras 2.9 y 2.8). El uso de estos tipos y su iconografía simplifica el modelado. Cuenta con cinco clases descendientes de geo-campos: *isolínea*, *polígonos adyacentes*, *mosaico*, *muestreo* y *red triangular irregular*. Existen dos clases descendientes de geo-objetos: *geo-objetos con geometría* y *geo-objetos con geometría y topología*. Estas especializaciones junto con la agregación espacial (la primitiva *todo-parte*) permiten definir reglas de integridad espacial. Las clases de geo-objetos con geometría se utilizan para los objetos que sólo tienen propiedades geométricas que se especializan en las clases *Punto*, *Línea*, y *Polígono*. Las clases de geo-objetos geometría y topología representan a los objetos que tienen, además de geometría, propiedades de conectividad topológicas y se usan para representar estructuras de red. Incluyen a las clases *Nodo* y *Arco* con o sin dirección.

OMT-G considera tres tipos de relaciones entre sus clases: *asociaciones simples*, *relaciones topológicas de red* y *relaciones espaciales*. La discriminación de estas clases de relaciones permite definir explícitamente el tipo de interacción que existe entre objetos. Las asociaciones simples muestran relaciones entre objetos de diferentes clases convencionales o georeferenciadas. Las relaciones espaciales representan relaciones topológicas y métricas: *tangente*, *dentro*, *atraviesa*, *superpone*, *disjunto*, *adyacente*, *coincide*, *contiene*, etc. Para cada una de estas relaciones se definen las restricciones de integridad espacial, y es posible definir nuevas relaciones espaciales con sus restricciones. Las relaciones de red se definen entre objetos que están conectados bajo alguno de los modelos *arco – nodo* ó *arco – arco*.

OMT-G además incluye la definición semántica para relaciones espaciales “generalización” y “agregación”. Las asociaciones de generalización y especialización se aplican

Geo-Objects con geometría



Geo-objects con geometría y topología

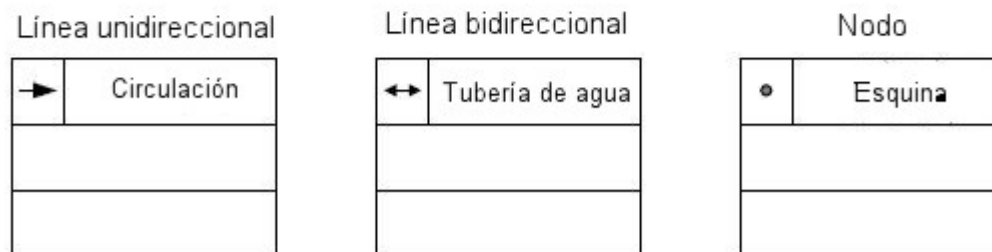


Figura 2.9: Clases espaciales Geo-objetos o discretas

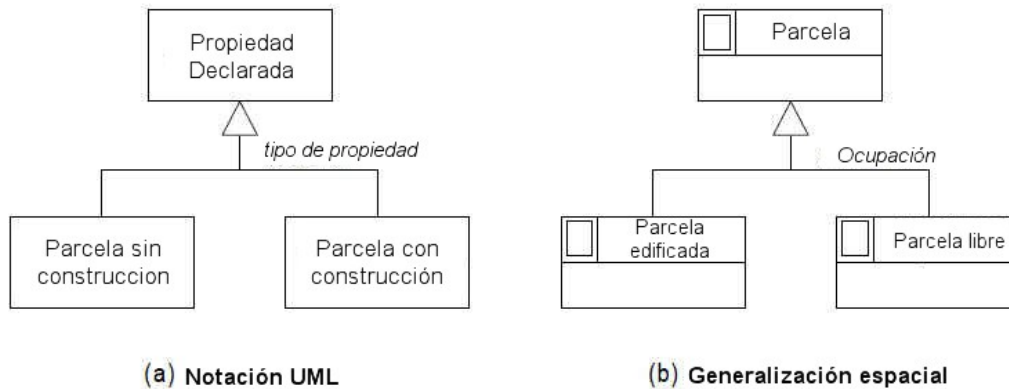


Figura 2.10: Generalización en OMT-G

sobre clases georeferenciadas y clases convencionales, siguiendo las definiciones y notación propuesta por UML (figura 2.10). La agregación es una forma especial de asociación entre objetos que define la composición (figura 2.11); se puede definir entre clases convencionales, georeferenciadas o combinaciones de ambas. Cuando la agregación está definida entre clases georeferenciadas, se denomina “agregación espacial”. En la agregación espacial la relación topológica *todo-parte* es explícita. El uso de agregación espacial impone restricciones de integridad.

Ejemplo 2.2.4 En la figura 2.11 se muestra un ejemplo de agregación espacial entre las clases espaciales MANZANA y LOTE, ambas con geometría polígono. Esta agregación implica las siguientes restricciones de integridad espacial:

1. El área de una instancia de MANZANA está limitada al área ocupada por las instancias de LOTE que la componen.
2. No existe superposición entre instancias de LOTE.
3. No existe ningún objeto de LOTE fuera de un objeto MANZANA.

Justamente, una característica de OMT-G es que transforma las relaciones topológicas y espaciales en restricciones de integridad. Las restricciones de integridad se clasifican

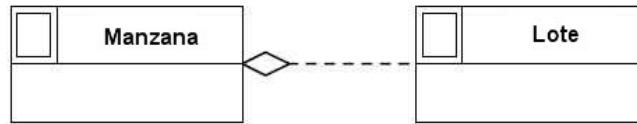


Figura 2.11: Agregación espacial en OMT-G

así en reglas de integridad topológicas, semánticas y/o restricciones definidas por el usuario. Las reglas de integridad topológicas están restringidas por las propiedades geométricas y las relaciones espaciales de las entidades espaciales [EF91]. Por ejemplo, una restricción de integridad topológica podría ser: “la PARADA(*punto*) de una línea de colectivos debe estar ubicada *sobre* la RUTA(*conjunto de líneas*) de recorrido”. Las restricciones semánticas tienen que ver con el significado de los fenómenos y son propias de la aplicación. Un ejemplo de restricción semántica es el siguiente: “no es posible ubicar una CONSTRUCCIÓN de edificio sobre un RÍO”. Las restricciones definidas por el usuario se conocen como “reglas de negocio”. Por ejemplo: “por cuestiones legales no se permiten centros comerciales cercanos a menos de 200mts de un HOSPITAL”.

Una característica saliente de OMT–G es que se diseñó no solo resaltando las características de modelado, sino también la capacidad de definir restricciones de integridad a nivel conceptual. OMT–G provee las primitivas necesarias para definir datos espaciales, relaciones espaciales, y permite definir explícitamente restricciones topológicas, semánticas y reglas de negocio. Siendo un modelo orientado a objetos, además permite que ciertas restricciones espaciales sean encapsuladas como métodos asociados a clases espaciales.

2.3. Modelos lógicos para información geográfica

Como ya se dijo, definir niveles de abstracción es una práctica común en Ciencias de la Computación para resolver un problema enfocándose en un conjunto reducido de conceptos y ocultando los detalles que no son de interés. En el campo de modelado de información por medio de SMBD, estos niveles están bien definidos y son ampliamente utilizados (figura 2.1) [Ull88, SKS10, GMPQ⁺04, EN07]. Cada nivel representa una “vista” diferente de la misma información, donde un nivel superior se fundamenta en el nivel inferior.

El *nivel lógico* está debajo del nivel conceptual y describe el modelo de datos en términos de un conjunto de estructuras relativamente simples, enfocándose en objetos de almacenamiento de alto nivel. Seguramente, las estructuras simples del nivel lógico implican estructuras de nivel físico complejas. En la implementación de un problema se usan las estructuras lógicas, sin necesidad de manejarse con la complejidad del nivel físico, lo que simplifica considerablemente el diseño de la aplicación.

Con relación a los modelos de datos para información espacial de interés en esta tesis, en la sección anterior se identificaron los objetos espaciales y su componente geométrica al mayor nivel de abstracción. A nivel lógico se definen primitivas como puntos, arcos o superficies y se considera sus formas de representaciones. Se necesita representar conjuntos infinitos de puntos del espacio Euclídeo en la computadora. Existen dos modos principales de resolver ese problema [RSV02, Mol95]:

- *Modelo mosaico*⁸, aproxima el espacio continuo a uno discreto (figura 2.12(a)).
- *Modelo vector*, construye estructuras de datos que incluyen arcos y puntos (figura 2.12(b)). De manera similar trabaja la representación “Medio Plano” [BT88].

Ejemplo 2.3.1 *Supongamos que se requiere mantener información sobre el área ocupada por una ciudad C. La geometría del área que ocupa la ciudad es un polígono. Bajo la representación mosaico, se identifica al conjunto de celdas que aproximan el área ocupada por el interior de la ciudad, mientras que en el modelo vector se representaría con la lista de puntos que definen el contorno del polígono que aproxima al área.*

2.3.1. Modelo mosaico

El modelo mosaico define una descomposición celular del plano como una grilla o arreglo de celdas disjuntas. A cada celda se le asocia un conjunto de valores de atributos. La partición del espacio en celdas define el plano discreto, llamado *malla*, el que puede hacerse en un azulejado de celdas fijas o regulares denominado *raster* o en mosaicos variables o irregulares (figura 2.13 (a-b)). En este modelo cada celda está asociada con una posición en el arreglo y podría asociarse con un *pixel*.

⁸En Inglés *tessellation model*.

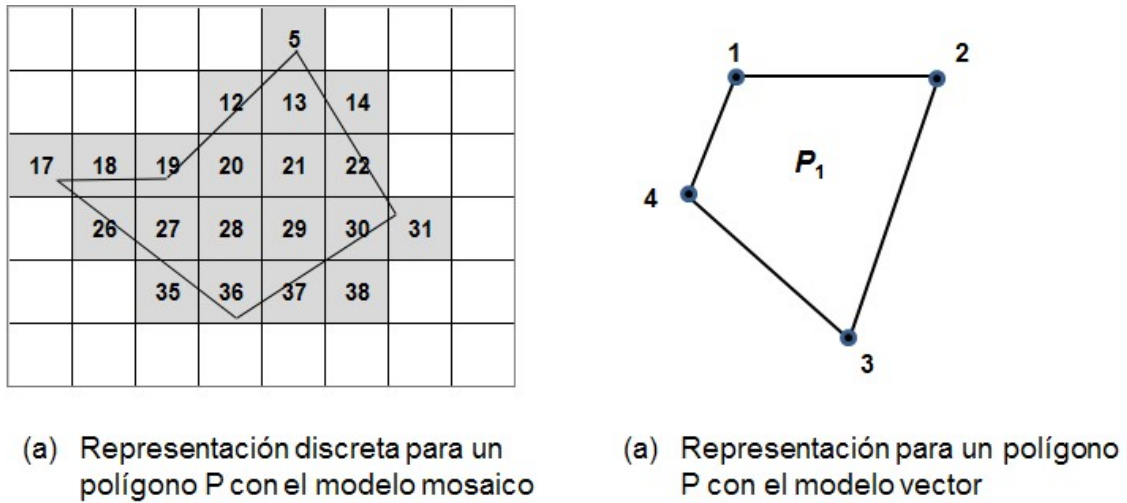


Figura 2.12: Modelos de representación

Si bien en la práctica existe una asociación natural entre el modelo mosaico y la representación del espacio basado en campos, el modelo mosaico también se puede utilizar para el modelo del espacio basado en objetos, donde los objetos abstractos como polilínea, polígono o región se representan por la enumeración finita de celdas. En el modelo mosaico se aproxima cada objeto espacial con un conjunto finito de celdas. A mayor resolución de grilla, mejor representación pero mayor cantidad de celdas a almacenar para representar un objeto. Por ejemplo, para una imagen A4, con 12pts/mm se requieren de 9 millones de pixels. Un problema claro es que a mayor resolución, más ineficiente será el tiempo de respuesta de las operaciones.

2.3.2. Modelo vector

En el modelo vector los objetos son construidos a partir de las primitivas *punto* y *arco*. El punto se almacena a partir del par de coordenadas con respecto al espacio de referencias y el arco se modela como un segmento definido por sus puntos extremos. Los objetos más complejos requieren de estructuras auxiliares de representación: listas, arreglos y/o conjuntos. En general, la representación en memoria del modelo vector requiere de menos recursos que el modelo mosaico. Existen variantes en cuanto a la representación del modelo vector. Una representación simple sería:

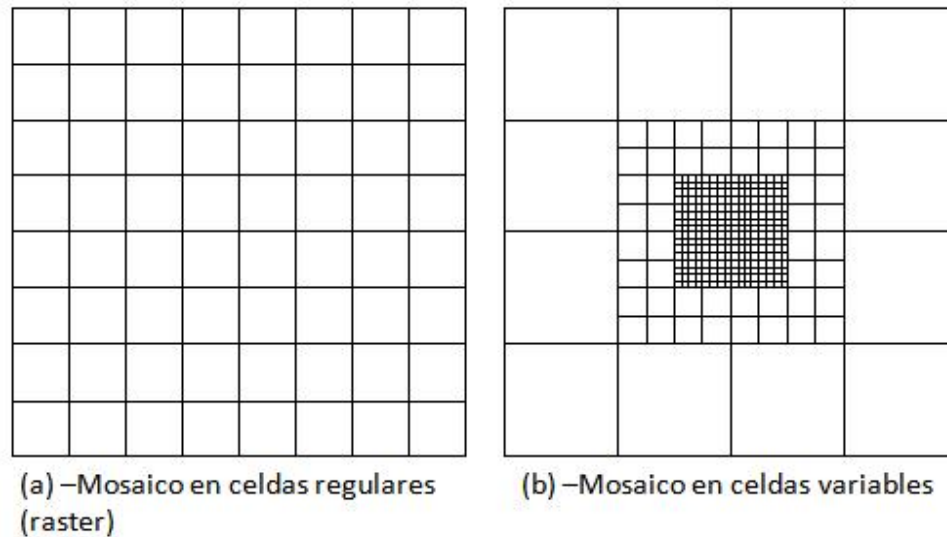


Figura 2.13: Dos representaciones del modelo mosaico

- Representar polilíneas como una lista de puntos $\langle p_1, \dots, p_n \rangle$, donde cada p_i es un vértice. Cada par (p_i, p_{i+1}) , con $i < n$, representa un segmento de la polilínea.
- Representar polígonos también como una lista de puntos con la restricción polilínea *cerrada*, y el par (p_n, p_1) como un arco más del polígono.
- Representar región como un conjunto de polígonos.

Para el modelo espacial basado en campos la presentación generalmente se basa en Modelos de Elevación Digital (DEM's). Los DEM's son útiles para representar cualquier fenómeno natural que es una función continua del espacio 2D. Los DEM's toman colecciones de datos de muestra, a partir de los cuales los otros puntos son obtenidos por interpolación. Un método de interpolación usado es el de Redes Irregulares Triangulares (TIN's).

2.3.3. Representación de geometría para colecciones de objetos

Las siguientes representaciones están enfocadas al modelo vector y espacio basado en objetos. Al considerar colecciones y no objetos individuales interesan las *relaciones* entre objetos de la misma colección. Los modelos más comunes usados para colecciones

de objetos son [RSV02]: *spaghetti*, *red*, y *topológico*. Entre ellos fundamentalmente difieren en la expresión de las *relaciones topológicas* entre los objetos componentes. Las relaciones topológicas entre los objetos espaciales son aquellas que permanecen invariantes bajo transformaciones topológicas (*adyacencia*, *superposición*, *disjunto* e *inclusión*). Esto es, se preservan aún cuando los objetos espaciales son trasladados, rotados o escalados en el plano. La representación explícita de tales relaciones en el modelo de datos espacial provee mayor conocimiento y es útil para la evaluación de consultas espaciales.

En el modelo de representación *spaghetti* [LT92, WD04], la geometría de cualquier objeto espacial que pertenece a la colección se describe en forma independiente de los otros objetos. En este modelo no se guarda la topología y las relaciones topológicas se calculan por demanda. Las principales características del modelo son que es simple y que habilita la representación heterogénea para poder mezclar puntos, polilíneas y regiones sin restricciones. Además, debido a que los objetos se guardan independientemente, provee al usuario final una manera simple de ingresar nuevos objetos a la colección. Por otro lado, las desventajas del modelo son principalmente la pérdida de información explícita sobre las relaciones topológicas entre objetos espaciales y la redundancia que implica la representación. Por ejemplo, los límites entre dos regiones adyacentes se mantienen duplicados y eventualmente podrían no coincidir.

El modelo de representación *red*⁹ [LT92, WD04] fue diseñado para modelar redes y grafos. Es especialmente adecuado para aplicaciones cuyo modelo conceptual subyacente es del tipo red, como por ejemplo las redes de transporte. El modelo almacena las relaciones topológicas entre puntos y polilíneas. El conjunto de tipos geométricos a considerar incluye:

- punto: $[x : real, y : real]$
- nodo: $[punto, < arco >]$
- arco: $[nodo - inicial, nodo - final, < punto >]$
- polígono: $< punto >$
- región: $\{polígono\}$

⁹En Inglés *Network*.

Introduce los conceptos *nodo* y *arco*. Un nodo es un punto distinguido que conecta una lista de arcos. Un arco es una polilínea que comienza en un nodo y finaliza en un nodo. Con esta representación es posible *navegar* a través de la red; cuando se alcanza un nodo se elige el arco por donde seguir. Los nodos permiten definir eficientemente el test de conectividad y cálculos de red, como camino más corto. Dependiendo de la implementación, las redes pueden ser planas o no. La principal ventaja de este modelo es su descripción intrínseca de topologías de red, con la noción de conectividad. Esto lo hace apropiado para los requerimientos de representación y consultas de cierto tipo de aplicaciones basadas en redes.

El modelo *topológico* [LT92, WD04] es similar al modelo de red, excepto que la red es planar. Este tipo de red induce a la subdivisión planar en polígonos adyacentes, algunos de los cuales pueden no corresponder a objetos geográficos. Los objetos de interés de este modelo son:

- punto: $[x : real, y : real]$
- nodo: $[punto, < arco >]$
- arco: $[nodo-inicial, nodo-final, polígono-izq, polígono-der, <punto>]$
- polígono: $< arco >$
- región: $\{polígono\}$

Como en el modelo de red, un nodo se representa por un punto y la lista de arcos que se inician o terminan en dicho nodo. Si la lista está vacía corresponde a un punto. Los puntos sueltos se usan para identificar ubicaciones específicas, como torres, capitales, etc. Los arcos, además de los puntos extremos y vértices, tienen referencia al polígono izquierdo y derecho que los tienen como límite común.

Un polígono se representa por una lista de arcos, cada arco siendo compartido con polígonos vecinos. Existe cierta redundancia por razones de eficiencia al acceder a los objetos. Por ejemplo, los polígonos pueden ser accedidos a través de la estructura polígono o de sus arcos. Sin embargo, no hay redundancia al almacenar la geometría, cada punto/línea se guarda una única vez.

Una ventaja del modelo topológico es que es eficiente para el cálculo de consultas topológicas. Por ejemplo, la consulta “*encontrar los polígonos adyacentes a un polígono*”

P ” es directa. Otra ventaja está relacionada con la consistencia de la actualización. Como los objetos están compartidos, la consistencia de operaciones de actualización es más fácil de garantizar. Esta aproximación tiene algunas desventajas. Mostrar una región requiere recuperar los polígonos adyacentes y filtrar la información que no interese. Segundo, la complejidad de la estructura subyacente puede resultar en pérdida de eficiencia para algunas operaciones, por ejemplo, mostrar una vista de un mapa. En el modelo topológico, las operaciones de recorrido son mucho más lentas que en otros modelos. Y finalmente, agregar un objeto requiere el cálculo previo de parte del grafo planar.

2.3.4. Evaluación de los modelos

La selección de un modelo conceptual y un modelo lógico son independientes. Cualquiera de los dos modelos lógicos vistos sirve para representar tanto los modelos conceptuales basados en objetos como en campos. En [Lua04] se enumeran algunos aspectos que pueden considerarse al momento de tomar una decisión:

- *Eficiencia de la representación.* El modelo vector es más eficiente en el uso del espacio de almacenamiento que el modelo mosaico, que dependiendo de la precisión requerida, puede implicar grandes volúmenes de datos o celdas. Esto además implica que las operaciones sobre los datos consumen más tiempo.
- *Eficiencia de cálculo.* El modelo vector implica el manejo de estructuras de datos más complejas que hace que ciertas operaciones espaciales como por ejemplo “cubrimiento” entre fenómenos sean más ineficientes que en el modelo mosaico. El modelo mosaico, por su parte, usa una estructura de datos simple como el arreglo y las que operaciones como filtrado son simples de implementar.
- *Poder expresivo.* El modelo vector permite representaciones explícitas de las relaciones topológicas entre objetos geográficos, por lo que resulta más apropiado en problemas que requieren de análisis topológico como “identificar caminos”.
- *Manejo de transformaciones.* Las operaciones de transformación de coordenadas como rotación, traslación y cambio de escala, en general, son más simples en el modelo vector. En la representación mosaico son más ineficientes y pueden resultar en distorsión y/o pérdida de información.

- *Análisis de espacio*. El análisis del comportamiento en cada ubicación del espacio es más simple en el modelo vector.
- *Producción de datos*. El modelo mosaico se asocia a la recuperación de datos que provienen de satélites y existen aplicaciones donde esta manera de recuperar información es esencial. El modelo vector se asocia a aplicaciones donde los datos provienen de diseños del tipo Diseño Asistido por Computadora (CAD)¹⁰.

2.4. Bases de Datos Espaciales

Un *Sistema de Manejo de Bases de Datos Espacial* (SMBDE)¹¹ provee la administración efectiva y eficiente de datos que están relacionados con un espacio de interés. Este espacio puede referirse al mundo físico (geografía, planificación urbana, astronomía), a partes de organismos vivos (anatomía humana), a diseños de ingeniería (circuitos integrados de gran escala, diseño de automóviles o estructuras moleculares para drogas farmacológicas), etc. Un SMBDE es una clase particular de *Sistemas de Manejo de Base de Datos (DBMS)* (SMBD)¹² que además administra información con referencia a un espacio. Los requerimientos y técnicas necesarias para manipular objetos geográficos, que tienen identidad, extensión, ubicación y relaciones particulares son diferentes de los necesarios para manejar imágenes capturadas. Por su parte, los servicios ofrecidos por los sistemas de bases de datos tradicionales tampoco son suficientes. Según [Gut95] un sistema de bases de datos espacial se puede definir como un software con las siguientes características.

1. Un SMBDE es un SMBD.
2. Cuenta con un modelo de datos, tipos de datos espaciales y el lenguaje de consulta para acceder a los datos.
3. Provee la implementación de datos espaciales, capacidades para indexado espacial y algoritmos eficientes para recuperar información.

Una forma de extender un SMDB Relacional a un SMBDE sería enriqueciendo el sistema de tipos de forma que incluya datos espaciales. La inclusión de estos tipos de datos

¹⁰En Inglés *Computer-Aided Design* (CAD).

¹¹En Inglés *Spatial Database Management System*(SDBMS).

¹²En Inglés *Database Management System* (DBMS).

requiere definir su estructura junto con las operaciones sobre ellos. Por ejemplo, si se define el tipo de datos *polígono* se podría incluir la operación *áreaPolígono*. Las combinaciones de tipos de datos y sus operaciones se definen como *tipos de datos abstractos espaciales* (TDAE). Definir el sistema de TDAE adecuado es una tarea desafiante y, en general, se puede decir que el resultado final es un compromiso entre simplicidad de la definición y riqueza de la representación.

Por otra parte, los Sistemas de Manejo de Bases de Datos Relacionales Orientados (SMBDROO)a Objetos¹³ combinan dos tecnologías: sistemas de bases de datos y lenguajes de programación orientados a objetos. Esta mezcla trae ventajas tanto desde el punto de vista de diseño como desde el punto de vista de desarrollo: poder de modelado, extensibilidad, reutilización de código y facilidad de mantenimiento.

2.4.1. Tipos de Datos Abstractos Espaciales

La definición de *tipos de datos abstractos* (TDA) es una estrategia de los SMBDROO para solucionar la falta de poder de modelado de las bases de datos relacionales para ciertos problemas. Un TDA es una vista de los objetos que define y el conjunto de operaciones sobre ellos. La idea es ocultar la estructura del tipo de dato y extender el lenguaje de consulta. El primer desafío consiste en definir los tipos geométricos. En general, existe un compromiso entre el poder de modelado capturado por la definición y las restricciones impuestas sobre la representación geométrica elegida. Esta relación está fuera del estudio de esta tesis. Sin perder generalidad se enumeran los siguientes objetos espaciales:

- *punto*(0-dimensión).
- *polilínea*(1-dimensión). Lista de pares de segmentos conectados, quedan fuera las polilíneas complejas.
- *región*(2-dimensión). Una instancia de este tipo es cualquier conjunto de polígonos no superpuestos.

Para completar la definición de un TDA Espacial (TDAE) el próximo paso es definir las operaciones necesarias para expresar las consultas.

¹³En Inglés *Object Relational Database Management System* (ORDBMS).

Operaciones para <i>región</i>	
Operación	Descripción
$PuntoEnRegion : region \times punto \rightarrow bool$	Verdadero si el punto pertenece a la región.
$Overlaps : region \times region \rightarrow bool$	Verdadero si las regiones se intersectan.
$Clipping : region \times rectangulo \rightarrow region$	Calcula la intersección entre una región y un rectángulo. El resultado es una región eventualmente vacía.
$Interseccion : region \times region \rightarrow region$	Es un caso general de la anterior. Calcula la región resultante de la intersección de dos regiones.
$Meets : region \times region \rightarrow bool$	Verdadero si las regiones de entrada son adyacentes.
$Area : region \rightarrow real$	Devuelve el área ocupada por la región.
$UnionRegion : \{region\} \rightarrow region$	Dado un conjunto de regiones, se retorna una nueva región con la unión de ellas.
Operaciones para <i>línea</i>	
Operación	Descripción
$PuntoEnLinea : linea \times punto \rightarrow bool$	Verdadero si el punto pertenece a la línea.
$Longitud : linea \rightarrow real$	Retorna la longitud de la línea.
$OverlapsLR : linea \times region \rightarrow bool$	Verdadero si la línea intersecta a la región.

Cuadro 2.3: Operaciones de los tipos **región** y **línea**

2.4.2. Operaciones espaciales

Encontrar todas las operaciones de un TDAE es un problema que no tiene una caracterización absoluta. Se espera que un SMBDE cuente con un conjunto general de operaciones, capaces de proveer un rango amplio de funciones. El cuadro 2.4.1 presenta algunos ejemplos de las operaciones espaciales más comunes para los tipos de datos región y línea y sus interfaces. Sin embargo, los requerimientos funcionales de las aplicaciones SIG son amplios y exceden a las funcionalidades que están bajo la responsabilidad del SMBDE. Por ejemplo, para el análisis de mercado es necesario mostrar datos a distintos niveles de resolución. Los sistemas de rutas necesitan operaciones de búsqueda de caminos

y las aplicaciones que manejan modelado de territorio necesitan interpolación, entre otros ejemplos.

Para ilustrar la complejidad del problema y la gran variedad de posibilidades que existen se confeccionó la siguiente lista que organiza las operaciones según sus argumentos. Se asume que al lo menos uno de los argumentos es espacial.

- *Operaciones unarias con resultado booleano.* Operaciones testean sobre alguna propiedad de un objeto espacial, por ejemplo, *existe(f)*.
- *Operaciones unarias con resultado escalar.* Las operaciones de este tipo más comunes incluyen el cálculo sobre alguna medida escalar, por ejemplo, *longitud(f)*, *area(f)*, entre otras.
- *Operaciones unarias con resultados espaciales.* Una variedad de operaciones caen dentro de este tipo.
 - *Transformaciones geométricas.* En este grupo se incluyen rotación, traslación, cambio de escala, entre otras.
 - *Operaciones de transformación de dimensión.* A partir de un objeto en d -dimensional generan otro de una dimensión superior o inferior, por ejemplo *limite(f)*.
 - *Operaciones para extracción de objetos.* Retornan un objeto espacial que satisface la operación. Por ejemplo, *copiar(f)*.
- *Operaciones binarias con resultados booleanos.* También denominadas *predicados espaciales*, son la base de los lenguajes de consulta espacial. Se distinguen *selección espacial* y *join espacial*.
 - *Predicados topológicos.* Son invariantes con respecto a transformación topológica, tal como *intersecta(f₁, f₂)*, *limita(f₁, f₂)*, *cubre(f₁, f₂)*.
 - *Predicados de dirección*, tales como *alNorte(f₁, f₂)*.
 - *Predicados métricos*, por ejemplo, testear si la distancia entre dos objetos es menor o igual a un número de unidades.
- *Operaciones binarias con resultados espaciales.* Las operaciones de uso más frecuente para consultas espaciales son las *operaciones de conjunto* que aplican sobre

objetos representados como conjuntos, $unión(f_1, f_2)$, $intersección(f_1, f_2)$ y $diferencia(f_1, f_2)$. Sirven para crear un nuevo objeto espacial, por ejemplo, la intersección puede usarse en la consulta “*Filtrar las líneas de ferrocarril que pertenecen a la provincia de Buenos Aires*”.

- *Operaciones binarias con resultados escalares.* La operación típica es $distancia(f_1, f_2)$.
- *Operaciones n -arias con resultados espaciales.* Por ejemplo *seleccionar- n* .

Semántica de las Operaciones

Una operación puede tener distintos significados según la dimensión de los objetos considerados. Por ejemplo, la operación intersección no necesariamente preserva la dimensión de sus operandos. Una solución es acotar el problema definiendo diferentes operaciones de intersección de manera que cada una preserve la dimensión de sus operandos (intersección regularizada). Aún con la intersección regularizada, si los polígonos no son convexos, se puede generar más de un polígono.

Tipos Fuertes versus Tipos Débiles

El resultado de operaciones espaciales debería satisfacer alguno de los tipos existentes. La razón detrás de esta exigencia es que existen situaciones en las que el resultado de una operación es reutilizado por la siguiente operación. Hay numerosos ejemplos que muestran los límites en las operaciones con tipos simples: la intersección de una línea con una región no necesariamente retorna una línea, la diferencia de dos polígonos sin agujeros puede resultar en un polígono con agujeros. Una solución es redefinir la región como un conjunto de polígonos. La definición de tipos más abarcativos puede ser apropiada en algunos casos, por lo menos incorporando a la definición la noción de conjunto. Esta aproximación se denomina tipado débil. Sin embargo, trabajar con tipos genéricos, en muchos casos hace que se pierda el poder de los tipados fuertes. Por ejemplo, redefinir la línea como conjunto de líneas, hace que la unión de líneas acepte cualquier conjunto de líneas.

2.5. Resumen del capítulo

La característica saliente que distingue a los SIG de otros tipos de SI es que incluyen datos con referencias geográficas o datos espaciales. Eventualmente además pueden incluir algún atributo geométrico para representar el espacio ocupado. Actualmente, las aplicaciones SIG delegan la responsabilidad sobre la gestión de datos al SMBDE.

El diseño apropiado de la base de datos es fundamental para satisfacer los requerimientos de los usuarios. Los modelos de datos capturan la naturaleza de los datos y sus relaciones. Tradicionalmente los SMBD utilizan modelos de datos de distinto nivel de abstracción para separar la complejidad de la implementación de los requerimientos de datos del problema. En el nivel superior se ubica a los modelos conceptuales, que sirven como herramienta de diseño y comunicación entre desarrolladores y expertos del dominio. En el diseño de modelos de datos para SIG se consideran dos abstracciones conceptuales: el modelo de datos geográfico y el modelo para el esquema de aplicación. La interpretación del espacio varía entre los modelos basado en campos y basado en objetos. Adicionalmente, el modelo conceptual también describe el diseño del esquema de aplicación y debe proveer las abstracciones para representar los distintos tipos de fenómenos geográficos junto con sus relaciones espaciales, de una manera gráfica y sin ambigüedad.

El próximo nivel de representación lo constituyen los modelos lógicos, que se encargan de trasladar el modelo conceptual a estructuras internas. En el nivel lógico se ubican los constructores que se utilizan en el desarrollo de la aplicación, y se siguen ocultando detalles propios sobre la implementación de los mismos. Relacionados con la representación lógica existen dos alternativas: el modelo mosaico y el modelo vector. Entre ellos difieren en la concepción del plano, los requerimientos de almacenamiento y la capacidad para resolver consultas espaciales. El modelo vector considera la existencia de objetos en el plano. Para la representación de estos objetos existen distintas alternativas como son los modelos spaghetti, red y topológico. La selección del modelo conceptual y su representación a nivel lógico es ortogonal.

Capítulo 3

Sistemas de Información Geográfica

Bajo la sigla SIG se reúne al conjunto de herramientas que sirven para capturar, almacenar, analizar, administrar y presentar datos con referencia a una o varias ubicaciones sobre la superficie terrestre. La tecnología SIG combina conocimiento de diversas áreas de las ciencias como son cartografía, análisis estadístico y bases de datos, entre otras. Los SIG son una tecnología de propósito general que incluye servicios para analizar modelos, localizar eventos, medir distancia entre eventos, encontrar caminos y explorar las relaciones entre dos o más objetos. Esto permite a usuarios expertos del dominio combinar capas de información cruda de manera que se genere un nuevo conocimiento sobre un espacio geográfico. Los campos de aplicación para esta tecnología son diversos y entre ellos se pueden citar: cartografía, sensado remoto, distribución de superficie, organización política, análisis sociológico, distribución de recursos naturales, planeamiento urbano y localización.

Los datos de un SIG se corresponden con fenómenos del mundo real (por ejemplo: rutas, edificios, superficies agrícolas, etc.) representados digitalmente. Como se vio en el capítulo anterior, las entidades reales se organizan en dos abstracciones: objetos discretos (por ejemplo: un monumento o un lago) y campos continuos (por ejemplo: elevación de la superficie terrestre o nivel de lluvias). Tradicionalmente, también existen dos formas de almacenar los datos en un SIG para ambas clases de abstracciones: mosaico y vector. Los SIG que se centran en el manejo de datos en formato vectorial son populares. Sin embargo, los SIG raster son utilizados en problemas que requieren la generación de capas continuas. Ambos modelos difieren en cuanto a la eficiencia de representación y a la capacidad y eficacia para resolver determinado tipo de consultas espaciales.

En este capítulo se describe la tecnología SIG, sus componentes y servicios como un tipo de herramienta de propósito general, y se presentan algunos ejemplos de aplicaciones desarrolladas en base a esta tecnología. En la actualidad los SIG están teniendo un fuerte impulso con los llamados *Servicios Basados en la Localización*(LBS)¹ debido a la masificación de la tecnología de los *Sistemas de Posicionamiento Global*(GPS)² integrada en dispositivos móviles. Cada instancia desarrollada en base a tecnología SIG es un SI que administra datos espaciales con un propósito para el cual fue desarrollado. Un sistema desarrollado con tecnología SIG para una aplicación, jurisdicción, empresa o propósito no siempre es interoperable o compatible con otra instancia SIG desarrollado para otra aplicación, jurisdicción, empresa o propósito. En los últimos años ha crecido el interés en la integración de IG, buscando facilitar la interoperabilidad y la reutilización de datos entre aplicaciones SIG.

3.1. La tecnología SIG

Hasta hace unas décadas atrás, manipular, sintetizar y representar IG estaba limitado al uso de mapas de papel, siguiendo algún procedimiento esencialmente manual y no interactivo. En los '70 y con el avance de la tecnología, sumado a la demanda de manipulación y análisis interactivo de IG surge un nuevo tipo de herramienta denominada SIG. En un sentido general, el término SIG se aplica a sistemas que integran, almacenan, editan, analizan, comparten y muestran IG para colaborar en la toma de decisiones. Los SIG son herramientas que permiten a los usuarios crear consultas interactivas con búsquedas orientadas, analizar información espacial, editar datos, mapas y presentar los resultados de aplicar tales operaciones [Cla86].

3.1.1. Alcance y servicios

En la introducción del capítulo 1 (sección 1.1) se presentó una definición para tecnologías SIG propuesta por [Lua04, Mit99a]. De ella se consideran los siguientes tipos de servicios esperados:

1. Soporte de al menos un modelo de datos geográfico.

¹En Inglés *Location Based Services* (LBS).

²En Inglés *Global Position System* (GPS).

2. Herramientas para análisis y simulación.
3. Herramientas alternativas de visualización.

De lo anterior se puede concluir que un SIG es un tipo de sistema con facilidades para modelar, capturar, almacenar, manipular, guardar, consultar, recuperar, analizar y visualizar información que tiene algún componente geográfico. Además se espera que provea las herramientas para analizar modelos, localizar eventos, medir distancia entre dichos eventos, encontrar la mejor manera de llegar a un destino y explorar cómo los problemas se relacionan unos con otros.

Cada una de las piezas de información se organiza en una *capa*³ o *tema* y los usuarios deciden dinámicamente qué capas mostrar y qué capas ocultar. Por ejemplo, una capa podría representar los lagos de una zona geográfica y otra los caminos y rutas de la misma región. Superponiendo ambas capas se puede responder a la pregunta “¿es cierto que para todo lago existe un camino que lo comunica?”.

Combinar capas de información y ofrecer una nueva representación visual del espacio geográfico es un recurso flexible. Qué capas de información definir y cómo y cuáles combinar depende del objetivo de investigación. Un software tipo SIG debe permitir realizar el mapeo entre dónde ocurren los hechos y los atributos que poseen, organizados en una estructura de capas. Superponer información y seleccionar aquella que es relevante en una visión particular, fortalece los cuatro modos de preguntar de cualquier proyecto de investigación: exploración, explicación, predicción y planificación.

3.1.2. Organización

Una aplicación SIG de calidad resulta de la combinación del conocimiento de los expertos, la IG en la forma de datos, hardware, software y métodos de análisis (figura 3.1), organizados para automatizar, gestionar y distribuir información y proporcionar una representación geográfica.

Las personas juegan un papel activo tanto en la etapa de diseño como en la de uso de los servicios de una aplicación SIG. Durante el diseño y la construcción se sugiere la conformación de equipos de trabajo interdisciplinario con expertos del dominio, expertos

³En Inglés *layer*.



Figura 3.1: Composición de un SIG

del área de geociencias y desarrolladores. Entre los actores de un SIG se caracterizan los siguientes según su perfil:

- *Lectores de mapas.* Suelen ser los consumidores finales. Estas personas usan los mapas creados con un propósito específico.
- *Diseñadores de mapas.* Organizan la información por temas de interés y combinan datos para diseñar mapas de acuerdo a un propósito.
- *Analista geográfico.* Encargado de resolver problemas geográficos, como dispersión de contaminación, encontrar la ubicación para una industria o buscar recorridos.
- *Operador de datos.* Ingresa los datos geográficos que nutren al sistema. Se vale de técnicas de edición, conversión y generación de datos.
- *Diseñador de esquemas* o diseñador de base de datos. Construye los modelos conceptuales y lógicos apropiados de acuerdo a las necesidades de un problema.
- *Desarrollador.* Personaliza un software SIG para resolver una necesidad específica.

Una fábrica de IG cuenta con diversas formas de capturar IG y diversas fuentes orígenes de datos. Algunos de los métodos usados para generar los datos de un SIG son la

digitalización de datos (impresos o filmados), mapas digitales, fotografías aéreas, datos obtenidos de mediciones topográficas y/o sensores remotos. En caso de usar representaciones vectoriales, puede ocurrir que la misma se genere a partir de un proceso de vectorizado de datos analógicos. Dependiendo del método usado para capturar datos, se deben considerar el coeficiente de exactitud del método y si al proceso de captura se le debe adicionar las etapas de edición y procesado posterior para eliminar los errores (limpieza de datos). Actualmente, la mayoría de los datos digitales provienen de fotografías aéreas que capturan datos en dos y tres dimensiones. Otra fuente de información muy utilizada últimamente son las coordenadas de posición tomadas a través de tecnología GPS.

Entre las razones por las que se extendió el desarrollo de aplicaciones SIG se considera la versatilidad que ofrecen como fuente de investigación y análisis [Mit99b]. La tecnología SIG suele incluir instrumentos para análisis, así como también opciones alternativas del tipo “*plug-in*” de herramientas de análisis geográfico. Como se vio en el capítulo anterior, las capacidades de análisis geográfico están restringidas y condicionadas por el modelo de datos subyacente. Por ejemplo, un SIG basado en el modelo topológico podrá reconocer y analizar relaciones topológicas espaciales, como “identificar la/s parcela/s catastrales que carecen de un servicio”, esto es, no lo incluyen dentro sus límites. Las aplicaciones representadas bajo el modelo red son más adecuadas si se quiere hacer análisis sobre “mejor camino” a seguir para trasladarse desde un punto origen a un destino. Otros ejemplos de funciones de análisis más complejos incluyen conocimiento científico y se aplican para análisis y predicción en meteorología, epidemiología, riesgos de catástrofe, etc.

3.2. Información geográfica

En el ámbito de uso de la tecnología SIG se utilizan diversos términos relacionados con IG. En esta sección se presentan definiciones de tales conceptos que serán usados en el resto de esta tesis.

Se denomina *información geográfica* (IG) a los datos espaciales georreferenciados, esto es, vinculados con uno o varios puntos sobre la superficie terrestre. La IG posee una posición implícita (la población de una sección censal, una referencia catastral), o explícita (coordenadas obtenidas a partir de datos capturados mediante GPS, etc.). En general

se dice que un SIG administra IG. La IG se caracteriza por su volumen y estructura intrínsecamente compleja. Para representar IG se requiere de modelos conceptuales, lógicos y físicos (secciones 2.2 y 2.3) con constructores específicos.

Los datos que mantiene un SIG son abstracciones que se corresponden con objetos del mundo real (por ejemplo rutas, edificios, superficies agrícolas, etc.) representados en forma digital. Cada uno de estos objetos que conforma la IG recibe el nombre de *fenómeno geográfico*. Como se vio en la sección 2.2.2 cada fenómeno es único y distinguible, está referenciado a una locación, espacio, área o región y se puede visualizar en un mapa.

La abstracción que representa al conjunto de fenómenos geográficos de una misma clase se denomina *tipo de fenómeno*. Un tipo de fenómeno caracteriza los atributos y relaciones de fenómenos reales. Los atributos pueden ser *atributos alfanuméricos* y/o *atributos espaciales*. Los atributos espaciales tienen asociada una componente *geométrica*, que determina la forma y dimensión del fenómeno (sección 2.2.1). Un modelo para información geográfica también incluye la definición de relaciones espaciales en fenómenos, que pueden ser *geométricas* y/o *topológicas*. Entre las propiedades métricas de objetos geográficos se incluyen longitud y superficie (dependen de la dimensión del objeto) y relaciones métricas entre objetos como distancia y/o relaciones de orientación (“arriba de”, “a la derecha de”, etc.). La topología hace referencia a propiedades de vecindad, adyacencia, inclusión y conectividad, que permanecen invariables ante cambios morfológicos, de escala o de proyección.

Si bien el tipo de análisis sobre IG es variado, no existe un modelo lógico que asegure proveer el conjunto completo de operaciones de análisis. Generalmente, la tecnología SIG posee un conjunto de primitivas que cubren ciertas capacidades de análisis y ofrecen la posibilidad de desarrollo de funcionalidades específicas. Desde la perspectiva de un SIG las relaciones entre fenómenos geográficos se organizan en alguna de las siguientes categorías: topológicas, de dirección, métricas o generales. Por relaciones generales se refiere a vínculos que pueden existir entre objetos del sistema como por ejemplo la relación “es dueño” entre personas y parcelas.

3.3. Caracterización de las aplicaciones SIG

3.3.1. Aspectos funcionales

Presentados la definición y el alcance de una aplicación del tipo SIG y bajo conocimiento de las características particulares que comprende manejar datos geográficos digitalizados, se propone la siguiente organización de los componentes funcionales de un SIG [BM98]:

- *Ingreso de datos y verificación.* Esta funcionalidad se relaciona con aspectos de captura de datos geográficos, verificación de su correctitud y conversión a forma digital.
- *Almacenamiento y gestión de datos.* Abarca la estructura y organización de la IG en términos de la manera en la que la percibe el usuario (es decir el *modelo conceptual*) y sobre la forma en que se maneja en la computadora (*modelo lógico* y *modelo físico*)
- *Transformación y análisis de datos.* Esta funcionalidad comprende el proceso de edición de información con el fin de mantenerla actualizada y libre de errores. El análisis de datos es una de las tareas principales de cualquier SIG y comprende la aplicación de métodos de análisis de información para alcanzar las respuestas para las consultas recibidas.
- *Presentación de Resultados.* La funcionalidad de producir mapas y/o materiales similares es una de las características de los SIG que los diferencia de los sistemas de propósito general.

El ingreso y verificación de datos es una de las fuentes principales de costo y tiempo de una aplicación SIG. Para la obtención de datos se necesita definir un procedimiento de extracción, transformación y carga (ETL)⁴ adecuado. Existen métodos y herramientas para mover datos desde múltiples fuentes, “limpiarlos”, reformatearlos y cargarlos en el SIG. Para obtener información geográfica, los métodos de captura son variados. Originalmente los datos se recuperaron a partir de mapas existentes. En la actualidad, se han ampliado las posibilidades de obtener datos digitales de fuentes diversas, incluyendo: fotografías aéreas, tecnologías de posicionamiento (GPS [KH96], EGNOS, GALILEO [LB00],

⁴En Inglés *Extract-Transform-Load* (ETL).

DGPS [GHEK00]), imágenes de sensores remotos (IKONOS [ME02], QBIRD) o nuevos mapas creados. Así también existe la posibilidad de contar con datos espaciales digitales ya producidos por instituciones, compañías privadas o servicios de la armada [Raj10]. Organismos e instituciones de Estados Unidos y Europa trabajan para definir estándares de exactitud y precisión de datos geográficos [FGD07, CEN07].

Los servicios referidos a la gestión de datos, esto es, fundamentalmente almacenado y recuperación, suelen ser trasladados a un SMBDE. Como vimos, los datos que gestionan los SMBD en aplicaciones del tipo procesamiento de transacciones en línea⁵ están limitados a grandes volúmenes de datos simples (números, caracteres, fechas, etc). La IG requiere de nuevos tipos de datos como son *punto*, *curvas*, *superficie* y colecciones de éstos como *colección de puntos*, *colección de líneas*, *colección de superficie* (sección 2.3). Los SMBDE extienden los SMBD con capacidades para gestionar datos espaciales. Algunos ejemplos de SMBDE son Oracle Spatial & Oracle Locator [ORA] y PostGIS [POS10]. Oracle Database 11g incluye servicios 3-D y Web para gestionar datos geoespaciales bajo los modelos vector y mosaico, representación de topología y modelos de red, cubriendo así las necesidades de diversas aplicaciones SIG. Además trabaja con formatos estándares lo que le permite poder interactuar con servicios de diferentes proveedores. PostGIS por su parte incorpora a PostgreSQL el concepto base de datos objeto-relacional para objetos geográficos. PostGIS “globaliza” al servidor PostgreSQL, ya que puede ser usado como el servidor de base de datos espacial de diversos SIG. PostGIS cumple con las especificaciones OpenGIS “Simple Features Specification for SQL” [BCU⁺99] certificando como estándar.

Los datos que son procesados por un SIG describen fenómenos en términos de “qué”, “cuándo” y “dónde” [Raj10]. La magnitud de la componente ubicación puede medirse en unidades que van desde los centímetros a miles de kilómetros. La habilidad de transformar datos de una escala a otra y de un sistema de proyección a otro es una parte importante del procesamiento de datos. Involucra no solo el cambio de tamaño, sino cómo tratar los problemas de precisión, eficiencia y consistencia. Otros tipos de transformaciones de datos ocurren cuando se realizan cambios de modelo mosaico a modelo vector y a la inversa. Estas transformaciones se relacionan con la funcionalidad de ingreso de datos. Finalmente, otras tareas de procesamiento tienen que ver con el objetivo principal de un SIG que consiste en transformar “datos” en “información”. La tecnología SIG ofrece

⁵En Inglés *On-Line Transaction Processing* (OLTP).

métodos para elaborar y analizar los datos geográficos y generar nuevos temas o salidas en mapas. Las aplicaciones SIG funcionan como herramientas de soporte de decisión. El proceso de interpretación involucra medidas objetivas y juicio subjetivo de expertos, para realizar asociaciones, identificar hechos y patrones.

Las habilidades de presentación de la tecnología SIG permiten comunicar hechos a la audiencia general. Incluye mapas, resúmenes, cuadros, reportes, animación, siendo continuos los avances en este área. Una calidad importante cuando se requiere de técnicas de análisis es la eficiencia. Para ello es importante la selección del modelo de datos adecuado para representar los fenómenos de cada problema particular, de forma que permita la recuperación de datos y algoritmos eficientes para procesamiento de consultas.

3.3.2. Restricciones

Es amplio el rango de problemas para los que una propuesta de solución es el desarrollo de una aplicación SIG. Las diferentes aplicaciones, más allá de la diversidad del dominio de aplicación, comparten las siguientes características comunes [Zei99]:

- En general un SIG necesita estar integrado con otras aplicaciones. Es importante que los datos del SIG se almacenen y organicen de forma tal que permitan el acceso distribuido.
- Es especialmente importante considerar una arquitectura de datos abierta⁶ que facilite la integración de datos geográficos con otros datos como pueden ser: datos obtenidos en tiempo real, imágenes o bases de datos corporativas generadas por proveedores externos.
- Los mapas siguen siendo la manera más usada para visualizar IG. En el uso de SIG las aplicaciones con acceso dinámico a mapas vía Internet actualmente son una restricción importante a considerar.
- Se debe elegir la estructura de datos apropiada que permita obtener la clase de análisis que se desea realizar. Por ejemplo, es diferente el modelado de una superficie continua si se usa un modelo mosaico o si se opta por un modelo del tipo vector y también es muy diferente el tipo de consultas que se podrán realizar al sistema.

⁶En Inglés *open data architecture*.

3.4. Visualización de mapas

El mapa es la representación visual principal de un sistema geográfico. En el mapa se vuelca el conocimiento sobre dónde se ubican ciertos fenómenos, cómo se puede llegar a ellos, su referencia geográfica con relación a otros elementos del mapa, etc. Esta información se puede obtener en un mapa tradicional; sin embargo, la característica diferente de un SIG es que permite sesiones interactivas, destacando los aspectos de interés y ocultando información no relevante.

Por definición, un mapa es una proyección específica que transfiere puntos o locaciones de la Tierra escaladas en una superficie o mapa de acuerdo con una transformación sistemática y usando alguna grilla latitud/longitud. Las imágenes y mapas son inherentemente bi-dimensionales o planos, aunque con el uso de colorimetría y sombreado pueden presentar una apariencia tri-dimensional.

3.4.1. Formas de representación

Desde hace años que existen convenciones para describir *accidentes geográficos* e identificar objetos en un mapa. Algunos ejemplos ampliamente conocidos son: el uso de línea doble para representar rutas, etiquetas para identificar lugares y el color azul para representar cuerpos de agua. Los accidentes geográficos representan elementos naturales (ríos, vegetación o montañas), elementos creados por el hombre (rutas, redes de comunicación o edificios) o pueden incluir divisiones administrativas (países, provincias, parcelas etc). Existen tres maneras principales de representación de un área geográfica sobre un mapa:

- como un conjunto de elementos discretos (figura 3.2(a)).
- como una imagen o grilla de datos de muestra (figura 3.2(b)).
- como una descripción de la superficie (figura 3.2(c)).

Cuando se habla de fenómenos geográficos con representación discreta, se refiere a rasgos que tienen distinta forma y que se pueden representar por puntos, líneas y/o polígonos. Los puntos representan rasgos geográficos muy pequeños como para ser modelados usando líneas o áreas. Entre los objetos que suelen tomar esta representación se pueden citar localidades, edificios o ubicaciones que no tienen área, como puede ser un pico montañoso.

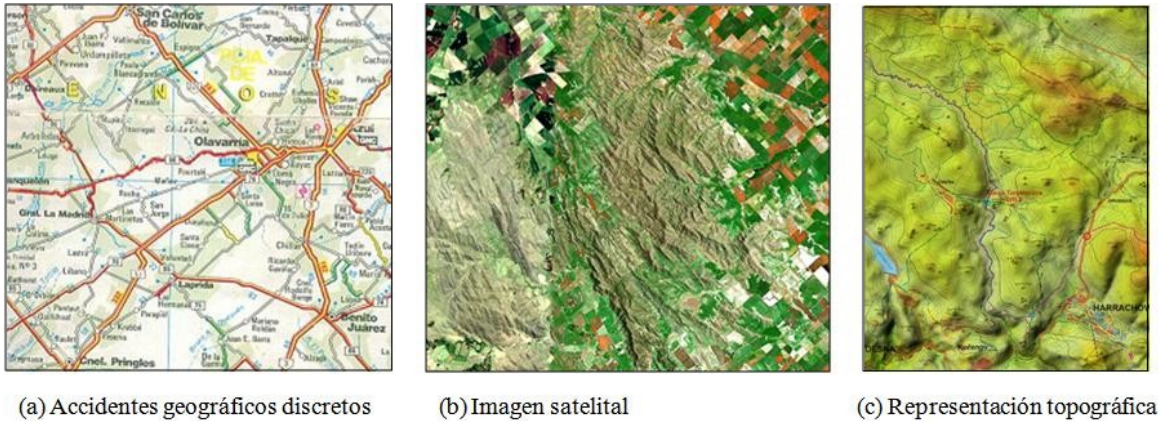


Figura 3.2: Representaciones geográficas en el mapa

Hay que tener presente que la calidad de pequeño depende de la escala que se utilice. Las líneas representan accidentes geográficos demasiado delgados como para ser modeladas usando áreas. En general incluyen a caminos, ríos, líneas de comunicación, y en algunos casos particulares, contornos. Por último, los polígonos son figuras cerradas que representan la forma y la ubicación de un fenómeno geográfico, como pueden ser estados, países, parcelas o zonas.

Un mapa como grilla de datos se obtiene en forma de fotografías aéreas o imágenes satelitales. Similarmente, representan fenómenos continuos muestreados como son temperatura, precipitaciones o elevaciones. El mapa se representa como una matriz comprimida bi-dimensional de celdas, donde cada celda mantiene sus propios atributos.

La superficie terrestre es continua. Algunos aspectos de dicha superficie pueden incluir elevaciones, declinaciones, picos y corrientes de agua. Las líneas que comparten igual elevación se representan dibujándolas en el mismo color. Así para describir la forma de la tierra se puede crear una superficie que se muestra usando algún rango de colores de manera de caracterizar la iluminación del sol, elevación, pendiente y apariencia. En general, este tipo de mapas modela la topografía de un lugar, pero la misma idea puede utilizarse para representar otros atributos como densidad de población de una superficie.

En resumen, se puede decir que al hablar sobre IG existen distintas características o propiedades geográficas que necesitan de diferentes abstracciones para su representación [Mit99b]. Cada abstracción ofrece sus propias cualidades y restricciones, es así, que dependiendo del dominio de aplicación y de las necesidades de consultas se deberá optar

por la representación que mejor responda a las calidades de performance, usabilidad y requerimientos de almacenamiento.

3.4.2. Los mapas interactivos

Los mapas tienen uno o más propósitos particulares. El primer objetivo es simplemente registrar dónde se ubica algo. A partir de allí dan a sus usuarios la posibilidad de ubicar una *posición relativa*. El tercer propósito sería el de interrelacionar la información de distintos mapas para alguna tarea de análisis o investigación específica que conduzca a diferentes acciones. Así, por ejemplo, para el objetivo “encontrar el mejor lugar donde instalar una fábrica”, se podrían integrar mapas con información topográfica, demográfica y de distribución de la energía, entre otros. Los mapas interactivos, son capaces de [Zei99] compartir y visualizar conocimiento del mundo de distintas maneras:

- Identificar con qué se corresponde una posición. Se puede obtener el nombre y otros atributos descriptivos.
- Identificar distribuciones, relaciones y tendencias. Por ejemplo, un especialista en demografía estaría en condiciones de definir guías de políticas públicas a partir de comparar mapas urbanos. Un especialista en epidemiología puede vincular los lugares en donde se desarrolla una epidemia con factores ambientales para encontrar las posibles causas.
- Integrar datos de diversas fuentes en un sistema de referencias geográficas compartido. Así, por ejemplo, se pueden combinar imágenes satelitales sobre climatología con mapas de distribución de siembra para mejorar la productividad.
- Resolver problemas espaciales. Por ejemplo, se pueden combinar varios niveles de datos para encontrar la mejor ubicación para un depósito de desechos.
- Encontrar el mejor camino. Por ejemplo, obtener el mejor recorrido para distribuir una carga, o definir el camino óptimo para un transporte público.
- Simular comportamiento ante posibles eventos futuros. Por ejemplo, simular el impacto de una emanación tóxica y desarrollar escenarios de evacuación.

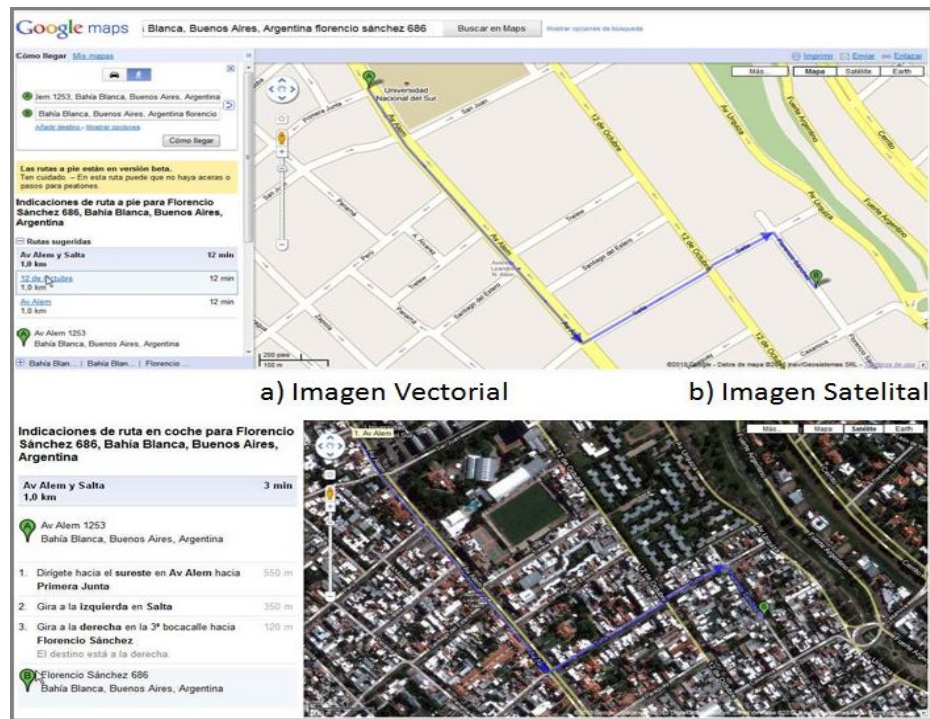


Figura 3.3: Google Maps - Buscar camino

Uno de los ejemplos más populares de mapas interactivos lo constituyen los Mapas Google⁷ [Goo05]. Mapas Google es un servidor de aplicaciones de mapas vía web libre. Cuenta con tres alternativas de representación: vista mapa (representación vectorial), imágenes satelitales e imágenes 3D (representaciones mosaico y 3-D). La representación satelital da opción de superponer la capa “calles y nombres”. Provee herramientas para aproximar/alejar y cambiar orientación. Cuenta con servicios para localizar una dirección, “pinchar” marcas en el mapa y buscar un camino para llegar desde un sitio a otro/s, bajo las alternativas desplazarse “a pie”, “en auto” ó “transporte público”. Los mapas generados son fácilmente integrables a sitios web. En la figura 3.3 se puede observar un ejemplo de uso de la aplicación Google Maps para encontrar el camino para llegar desde la Universidad Nacional del Sur a un destino dado. Incluye el detalle de calles a tomar y el tiempo estimado del recorrido.

⁷En Inglés *Google Maps*.

3.5. Arquitectura de los SIG

Dada una aplicación SIG típica se identifican las siguientes cuatro unidades funcionales principales [SMB⁺97]:

- **Unidad para ingreso de datos.** El origen de datos de un SIG es diverso. Típicamente, se vincula con sensores, cámaras y/o sistemas de posicionamiento global. Los datos obtenidos son discretizados para ser representados en la computadora. En algunos casos se combinan con procedimientos manuales. El proceso de discretización puede imponer errores o distorsiones. La unidad de ingreso de datos debe contar con alguna política para medir la calidad de los datos: procedimientos de validación y cuantificación de errores. Una propuesta consiste en definir y chequear restricciones de integridad y topológicas. Un ejemplo de restricción topológica podría ser “Carmen de Patagones está contenida dentro de los límites de la provincia de Buenos Aires”.
- **Modelo de datos.** Los modelos de datos proveen las abstracciones para referir a los fenómenos. Se fundamentan en la identificación de conceptos tal como se refieren en el mundo real y definir para ellos una representación lógica. Los datos que administra un SIG generalmente se agrupan en temas. Cada tema presenta valores sobre un tópico particular, por ejemplo “Rutas y caminos”. En el modelado de un tema se descompone y fragmenta el espacio como un conjunto de objetos o modelo vector, o como un conjunto de valores asociados a una ubicación o modelo de campos (sección 2.2). Las operaciones sobre el modelo de datos generalmente se delegan a algún SMBDE.
- **Unidad de Consultas.** Los datos geográficos son analizados y consultados por varias operaciones. Algunas de las consultas sobre datos involucran búsquedas espaciales y superposiciones. Las operaciones sobre tipos de dato vector incluyen operaciones geométricas, operaciones topológicas y operaciones métricas. La disponibilidad de operaciones está condicionada y restringida por el modelo de datos subyacente. Por ejemplo, un SIG que contiene un nivel de “rutas y caminos” de la provincia de Buenos Aires, representado por el modelo red, podrá calcular caminos para llegar a Mar del Plata partiendo desde Capital Federal.

- **Unidad de Visualización.** Un SIG generalmente cuenta con capacidades para presentar resultados visuales en la forma de mapa. La representación visual de las salidas de operaciones espaciales puede incluir: imágenes gráficas con objetos, imágenes discretizadas bajo el modelo de campos, composición o superposición de mapas y producción cartográfica.

3.6. Áreas de aplicación

Los SIG son ampliamente usados y su aplicación abarca una gran variedad de tipos de problemas. Entre los usuarios de los sistemas SIG se incluyen a grandes empresas, organismos estatales o gubernamentales, y organizaciones internacionales. En el ámbito privado, las empresas están interesadas en crear y mantener bases de información sobre comunicaciones, tendidos, desagües, etc. Por su lado los organismos estatales y gubernamentales están interesados en los SIG y su aplicación en proyectos de planeamiento urbano, transporte, servicios ambientales, turismo e información censal. También existe interés en utilizar tecnología SIG por miembros internacionales como la Organización para la Agricultura y la Alimentación de las Naciones Unidas (FAO)⁸ [SWB⁺], y una variedad de organizaciones comerciales que abarcan desde grandes cadenas de supermercados a productores de baja escala. A modo descriptivo, los distintos tipos de problemas [Cor00] para los que está indicado el uso de SIG se pueden clasificar en las siguientes categorías:

- *Análisis demográfico:* comprende el estudio sobre la distribución de la población de un país o región. Su conocimiento es importante para muchas actividades relacionadas con análisis de captación de interés, como puede ser para el desarrollo de pequeñas y medianas empresas, el planeamiento público (transporte, salud, servicios) y la distribución la tecnología de comunicación. En tales actividades, el objetivo es encontrar regiones apropiadas para cierto desarrollo y el impacto del mismo.
- *Gestión ambiental:* varían desde el monitoreo de polución hasta estudios de distribución de la fauna silvestre. La mayoría de los proyectos en este área son menos comerciales y, a menudo, se valen de grandes volúmenes de datos sensados remotamente. Un área que está en crecimiento es el estudio de impacto ambiental, que necesita que se considere en su análisis el contexto geográfico. Agricultura de alta

⁸En Inglés *Food and Agriculture Organization of the United Nations.*

precisión, planes de fertilización y sembrado son otro tipo de tareas que se pueden organizar usando los servicios de un SIG.

- *Administración de servicios públicos*: se refieren al mantenimiento y planificación eficiente del uso, por ejemplo, de redes de gas o comunicación. En este caso un SIG se puede utilizar para estudiar la distribución eficiente de la energía, el balance de carga de redes eléctricas, etc. Las organizaciones se valen de datos propios para crear sus bases de información, y estos datos son combinados con datos de infraestructura y demográficos.
- *Análisis de redes de comunicación*: está relacionado con el diseño eficiente de sistemas de transporte y rutas de distribución, y son de interés para compañías de distribución o mensajería a través de la búsqueda de caminos óptimos. Aunque éste área es una aplicación propia de los SIG, en ciertos casos se deben incluir paquetes específicos debido a la complejidad de las funciones de análisis. Para el desarrollo de este tipo aplicaciones, además de los datos del sistema de conexión, también se necesita de información demográfica.
- *Distribución territorial*: está relacionada con tareas de planificación y desarrollo. Si bien estas aplicaciones en cierto modo se interceptan con el análisis demográfico y la gestión ambiental, se mantienen como un ítem separado dado que los departamentos de planeamiento pueden necesitar que el SIG mantenga atributos específicos e incluso aspectos temporales.

Esta clasificación no es rígida ni exclusiva y muchas aplicaciones SIG ocupan más de una categoría. De este breve resumen es claro que los SIG son ampliamente aplicables y de allí el interés de las empresas de desarrollo de software en invertir sus esfuerzos en la investigación y desarrollo de los mismos.

Otras aplicaciones que siguen ganando interés son las aplicaciones espaciales que incorporan aspectos temporales. La combinación de espacio y tiempo permite observar el comportamiento de objetos en movimiento y predecir o informar caminos, como por ejemplo movimientos de tormentas, desplazamiento de vehículos, personas o cualquier otro objeto en movimiento.

3.7. Aplicaciones

A continuación se presentan algunos ejemplos de aplicaciones del tipo SIG. Todos ellos corresponden a aplicaciones gubernamentales de uso público.

3.7.1. Sismos y Terremotos en el mundo

El “Earthquake Hazards Program”⁹ [USG07] forma parte de la multi-agencia National Earthquake Hazards Reduction Program (NEHRP). Cuenta con un sitio web proporcionado por los United States Geological Survey (USGS) con varias aplicaciones que incluyen componentes SIG. Una de las áreas de investigación en desarrollo de USGS es este programa que se ocupa de colaborar en la predicción para reducir peligros a causas de tsunamis, sismos y erupciones volcánicas. En el sitio se puede encontrar una aplicación en línea con información en tiempo real sobre los movimientos terrestres en el mundo. La información a visualizar incluye información histórica y hasta los movimientos terrestres registrados en la última hora. Provee mapas con vistas vectoriales y satelitales. Permite consultar información mundial, posee facilidades de zoom y detalles descriptivos de los eventos. Visualiza en mapas los movimientos sistémicos. En la figura 3.4 se ven dos mapas con información de los movimientos terrestres ocurridos en Chile en febrero de 2010. La imagen 3.4(a) corresponde a un mapa con información continua que muestra la escala de intensidad de los movimientos, la imagen 3.4 (b) identifica los distintos puntos de movimientos, donde el color y tamaño de los puntos sirve como referencia sobre la intensidad.

Con relación a las características del software, se trata de un desarrollo que trabaja con información en tiempo real. Como resultado de este desarrollo se creó una nueva herramienta, Mecanismos focales 3D (3DFM), para ver dimensionalmente los símbolos que simulan focos de terremoto. Esta herramienta opera bajo Environmental Systems Research Institute (ESRI) GIS software ArcScene 9.x [ESR08]. El programa funciona con las entradas provistas por Advanced National Seismic System (ANSS) que se ocupa de obtener datos precisos y en línea de eventos sísmicos y sus efectos. Para ello se emplean las últimas tecnologías y métodos de monitoreo con foco en regiones de riesgo alto a moderado y redistribución automática de los eventos detectados, alarmas de tsunamis y erupciones volcánicas. ANSS cuenta con una red de 7000 sensores que proveen datos en línea con detalle, profundidad y valores muestreados.

⁹<http://earthquake.usgs.gov/>

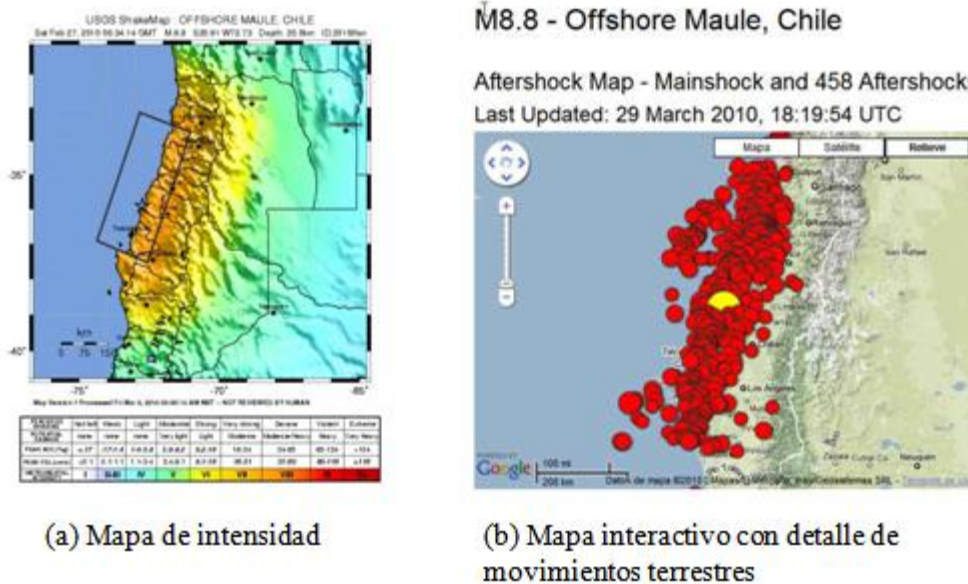


Figura 3.4: Earthquake Hazards Program - Mapas sobre los terremotos de Chile 2010

3.7.2. TOXMAP: Mapas sobre salud ambiental

TOXMAP: Environmental Health e-Maps [oM09] es una aplicación SIG que provee la Division of Specialized Information Services¹⁰ de la US National Library of Medicine (NLM)¹¹. Es una aplicación SIG que combina una capa de “división administrativa de Estados Unidos” con capas de contaminación ambiental generadas a partir de datos provistos por la US Environmental Protection Agency (EPA) sobre componentes tóxicos y manejo de desechos industriales.

El origen de datos de contaminación es la base de datos pública “Toxics Release Inventory (TRI)” que se genera a partir del inventario de liberación de químicos tóxicos y manejo de desperdicios reportados anualmente por los grupos industriales e instalaciones federales según lo establece la ley “the Emergency Planning and Community Right to Know Act (EPCRA)”¹². Esta base de información mantiene datos sobre cantidades y distribución de tóxicos químicos que son liberados al aire, al agua y/o que son enterrados cada año, así como el tratamiento, ubicación y cantidades de desperdicios químicos.

¹⁰<http://sis.nlm.nih.gov>

¹¹<http://www.nlm.nih.gov>

¹²<http://www.epa.gov/regulations/laws/epcra.html>

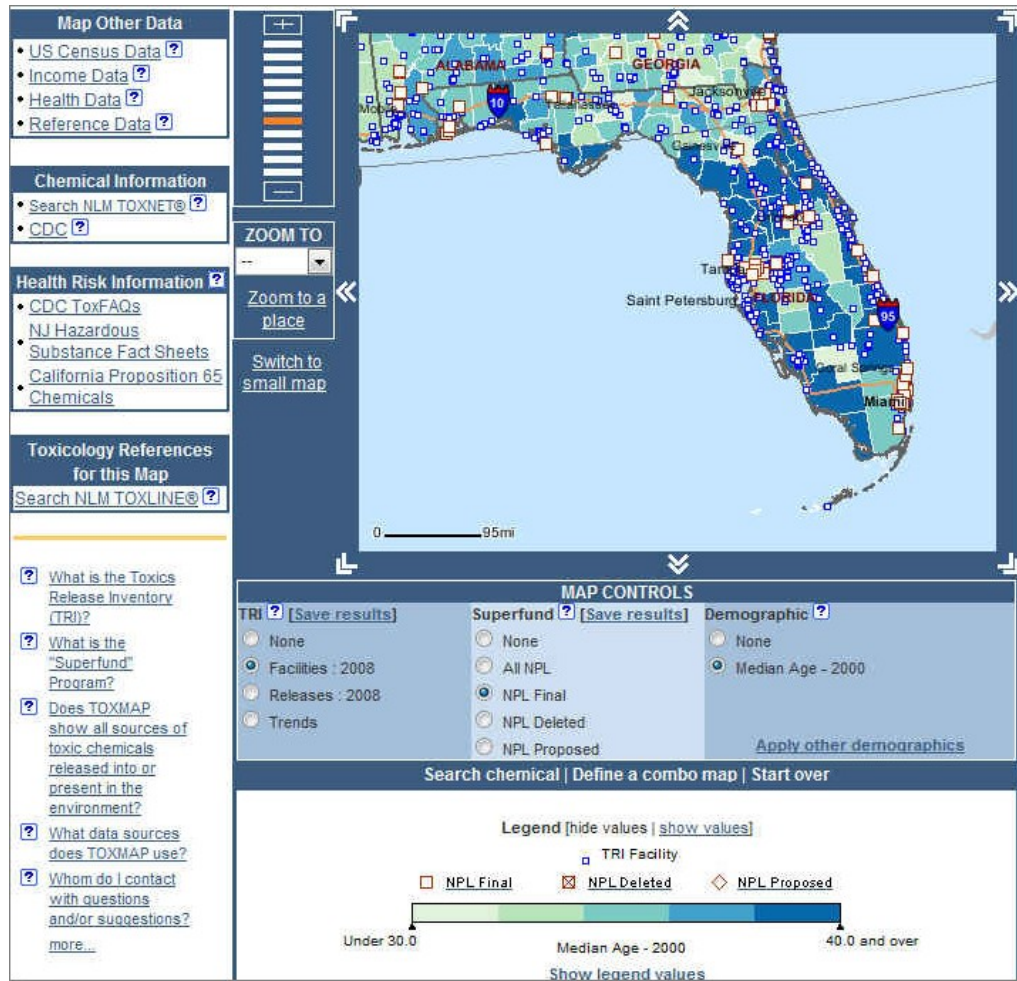


Figura 3.5: SIG - TOXMAP

La aplicación TOXMAP: Environmental Health e-Maps recupera y transforma la información que le provee la agencia de protección ambiental siguiendo procesos estrictos de control de calidad sobre los datos generados. Provee facilidades para crear en línea mapas interactivos que combinan las capas de organización administrativa de Estados Unidos, con las capas de emanación de tóxicos, medidas paliativas y mortalidad de las personas. La aplicación es de uso público y está disponible vía Web. Permite confeccionar mapas, aproximar regiones, ver información de contaminación y combinarlas con medidas de planeamiento y mortalidad. Es posible ajustar la capas de mortalidad definiendo filtros por sexo, raza y causal (cáncer u otros) (figura 3.5).

3.7.3. MAPS: Herramientas de Análisis y Mapas sobre Seguridad Pública

Mapping and Analysis for Public Safety (MAPS) [oJ09] del National Institute of Justice (NIJ). El departamento de justicia de Estados Unidos participa del proyecto MAPS, un programa que apoya la investigación y ayuda a los organismos a usar tecnología SIG para mejorar la seguridad pública. El programa incluye: usar mapas y datos espaciales para realizar análisis del delito, integrar el trabajo de investigadores para evaluar y definir programas y políticas y elaborar herramientas para construir mapas, agregar datos y realizar análisis espacial. En el marco del proyecto se crearon tres aplicaciones geoespaciales de escritorio para la aplicación cotidiana en organismos de seguridad pública:

- **CrimeStat III:** es un programa con estadísticas espaciales que se utiliza para analizar los lugares de incidencia de delito. Incluye características para identificar y analizar los lugares de mayor riesgo, trasladar incidentes a mapas, resaltar visualmente las zonas de concentración del delito, seguir la pista del comportamiento de los delincuentes, hacer seguimiento de la residencia de delincuentes seriales y conocer las rutas de persecución más ágiles.
- **FacilityCop software mapas:** es un software para análisis y registro de incidentes de delito en cárceles o prisiones. Las instalaciones carcelarias pueden utilizarlo para generar reportes tabulares, cuadros y gráficos.
- **School Crime Operations Package (School COP):** provee herramientas visuales para personal de escuelas que se puede utilizar para analizar mapas de incidentes de delito de la zona de influencia. Permite trabajar con información por escuela o por distrito escolar.

Este grupo de herramientas fue diseñado para ayudar a mejorar la seguridad, investigar los delitos y delincuencia, y definir iniciativas de prevención.

3.7.4. PROSIGA: Proyecto Sistema de Información Geográfica Nacional de la República Argentina

El *Proyecto Sistema de Información Geográfica Nacional de la República Argentina (PROSIGA)*¹³ [Nac06] integra esfuerzos de distintos organismos de la República Argentina que trabajan en IG y promueve la vinculación de información que es generada por los responsables oficiales. El objetivo es conformar un “mapa global del país” con información generada por productores oficiales, en formato digital, de acceso público y disponible a través de Internet. Se espera contar con información de utilidad para la toma de decisiones a nivel local, regional y nacional. Incluye información de dominio público relativa a diversos temas: relieve, hidrografía, composición de los suelos, bosques, producción agropecuaria, datos de los catastros urbanos y rurales, y equipamientos, entre otros.

El proyecto tiene por objetivos proveer disponibilidad en forma integrada de IG de la República Argentina, dar intervención directa a múltiples actores gubernamentales, vincular a los actores mediante una estructura nodal de intercambio de datos y brindar acceso a la comunidad a través de una herramienta que integra IG de distintas fuentes. Características de la propuesta:

- Provee acceso remoto en línea vía Web a IG.
- Garantiza datos actualizados por los distintos organismos responsables de proveer los datos.
- Ofrece acceso a información a escalas nacional, regional y catastral.
- Promueve el trabajo respetando normas y estándares que aseguren la integración de la información.

Entre las aplicaciones posibles se pueden citar: planificación y desarrollo, seguridad, forestación, gestión de la energía, catastro urbano y rural, agricultura, pesca, industrialización, ganadería, educación, etc. Un aspecto importante que favoreció el avance del proyecto fue la adopción de las tecnologías IDE y las normas y estándares ISO/TC 211, ISO y OGC.

¹³<http://www.sig.gov.ar/>

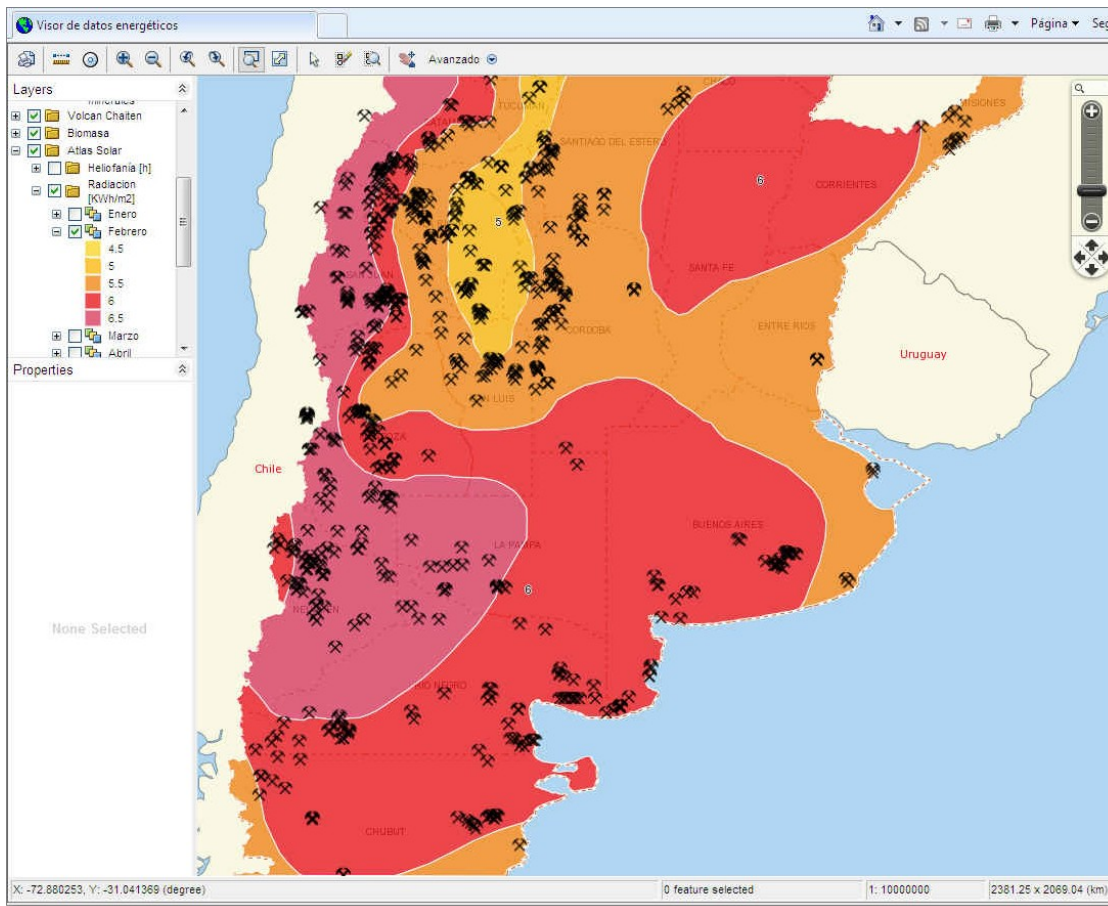


Figura 3.6: PROSIGA - Datos provistos por la Secretaría de Energía de la Nación

La operación de la aplicación es en línea. Es de dominio público, aunque requiere de la registración de usuarios. El ambiente posee múltiples capas. Posee facilidades para integrar nuevas capas y facilidades para definir zonas. El proyecto está en etapa de desarrollo. Algunas interacciones no pueden superponerse. La figura 3.6 muestra un ejemplo de mapa generado a partir de los datos y temas provistos por la Secretaría de Energía de la Nación. En el ejemplo se superpuso la capa de radiación solar medida en el mes de Noviembre (modelo continuo), con la capa de distribución de minas y canteras (modelo discreto). La aplicación interactiva provee diversas capas predefinidas, selección de capas por superposición, posibilidades de activar y desactivar capas, aproximar y alejar, y definir consultas.

Organismos que integran el proyecto PROSIGA

Actualmente veinte organismos y municipios forman parte del proyecto, ellos son: la Secretaría de Energía de la Nación (SE), el Gobierno de la Ciudad Autónoma de Buenos Aires (GCBA), la Secretaría de Agricultura Ganadería Pesca y Alimentos (SAGPyA), el Instituto Geográfico Nacional (IGN), la Municipalidad de Malvinas Argentinas (Pcia. de Bs As), el Instituto Nacional de Estadísticas y Censos (INDEC), el Ente Nacional Regulador del Gas (ENARGAS), la Municipalidad de Luján (Pcia. de Bs As), la Municipalidad de Junín (Pcia. de Bs As), la Municipalidad de Viedma (Pcia. de Río Negro), la Universidad Provincial de la Punta (Pcia. de San Luis), el ETISIG Chaco (Equipo de Trabajo Interdisciplinario en Sistemas de IG de la Provincia del Chaco), el Servicio Nacional de Sanidad y Calidad Agroalimentaria (SENASA), la Universidad Nacional de Catamarca, la Municipalidad de Rosario, la Secretaría de Ambiente y Desarrollo Sustentable de la Nación, el Centro Argentino de Cartografía, la Administración de Parques Nacionales, la Municipalidad de La Plata y la Universidad Católica de Salta.

3.8. Resumen del capítulo

Los SIG son una tecnología de “propósito general” que provee servicios para administrar datos geográficos, localizar eventos, encontrar caminos, explorar relaciones, y visualizar y analizar información a través de su uso interactivo. Basándose en esta tecnología se construyen diversas aplicaciones SIG para satisfacer los requerimientos de algún problema particular. Los datos geográficos que nutren las aplicaciones SIG están vinculados con uno o varios puntos de la superficie terrestre. La IG que representa a fenómenos reales se organiza en dos abstracciones: modelos discretos con objetos y campos continuos. Los SIG cuentan con los modelos de datos para poder representar en la computadora ambas abstracciones.

La estrategia de diseño ampliamente utilizada en geografía y que simplifica el modelo conceptual de una aplicación SIG es la de agrupar los hechos o fenómenos relativos a un tópico en “capas” o “temas”. Así, por ejemplo, para una región particular (país, provincia, distrito, ciudad) se podrían definir la capa *rutasy caminos*, para contener los elementos que representan a las vías de comunicación de dicha región, y la capa *espejos de agua* que se encargue de representar a lagos, ríos, lagunas, arroyos y diques. La organización por

temas permite trabajar superponiendo o quitando capas según se requiera. Superponer ambas capas permite tener conocimiento sobre qué caminos tomar para llegar a una laguna.

Un componente importante de la tecnología SIG lo constituye la unidad de visualización. En información geográfica la representación tradicional de fenómenos es en el mapa. El uso de mapas digitales agrega las posibilidades de superponer/quitar temas, enfocar/alejar una zona, optar por visualizaciones discretas o imágenes satelitales. Esto permite a usuarios expertos del dominio combinar capas de información cruda de manera que se genere un nuevo conocimiento sobre determinado espacio geográfico. Otro tipo de herramientas que acompaña a la visualización es la de consultas espaciales. Las consultas espaciales pueden ser métricas, geométricas o topológicas.

Los campos de aplicación para esta tecnología son diversos; entre ellos se pueden citar cartografía, sensado remoto, distribución de superficie, manejo de distribución política, análisis sociológico, distribución de recursos naturales, planeamiento urbano y localización. Así también el número de usuarios de aplicaciones SIG es amplio, incluyendo empresas privadas, organismos estatales y usuarios particulares. Entre los mayores costos (económicos y en tiempo) de una aplicación SIG está el de generar los datos. Esto sumado a los avances en tecnología de la comunicación que permiten tener acceso en línea a información distribuida ha aumentado el interés por poder compartir IG.

Capítulo 4

Integración de Información Geográfica

A través de los años varios problemas se han enfrentado con la necesidad de compartir y reutilizar el conocimiento adquirido sobre un dominio. Compartir conocimiento incluye transferir el saber de una persona a otra, de una organización a otra, de un grupo a una persona, entre otros modelos de colaboración. La comunicación distingue las entidades *remitente* y *receptor*. Si ocurre que el receptor y emisor son entes arbitrarios que no comparten el mismo lenguaje, terminología y/o modelo mental, se hace difícil o imposible descifrar el mensaje. La solución a este problema consiste en estructurar el mensaje con información suficiente para que el receptor pueda entenderlo.

La Web Semántica se relaciona con la noción de interoperabilidad. Se busca disponer en la Web de conocimiento especificado sin ambigüedades y que pueda ser interpretado por agentes automáticos. Esto permitirá usar la información más inteligentemente que simplemente para mostrarla, e incluso se podrían compartir y reusar datos y/o servicios de fuentes diversas. En particular, se plantea la necesidad disponer de una Web Semántica Geoespacial [Ege02], basada en un marco que comprenda diversas ontologías espaciales temáticas junto con una forma canonizada para especificar consultas geoespaciales. La demanda es por el acceso completo a la IG, que es heterogénea y está distribuida. Compartir IG en forma eficiente requiere encontrar una solución a los problemas técnicos relacionados con la heterogeneidad en la forma de representación. Las dificultades que surgen al reunir sistemas distribuidos heterogéneos se conocen como *problemas de interoperabilidad*. Para compartir información no solo se requiere proveer acceso a los datos, sino también

brindar información para que los datos puedan ser procesados e interpretados por sistemas remotos.

4.1. Motivación

Como se explicó en el sección 3.6 del capítulo anterior, los usos de SIG abarcan una amplia variedad de áreas de interés. Los propietarios incluyen a empresas privadas, organismos nacionales e internacionales y de educación e investigación. La IG en general es voluminosa, con estructura compleja y puede estar distribuida y clasificada por temas en diversos servidores. Uno de los mayores costos de implementación de una aplicación SIG es la adquisición de datos geográficos cuando no se dispone de ellos. En muchos casos, la IG de algún tópico específico, como por ejemplo “rutas y caminos”, ya existe para algún otro desarrollo. Contar con la posibilidad de compartir datos y servicios implica disminuir costos y tiempos de puesta en funcionamiento, así como optimizar la confiabilidad de la información. Los distintos ámbitos de uso de aplicaciones SIG comparten los siguientes requisitos:

- Necesitan estar integrados con otras aplicaciones. Es importante que los datos se almacenen y organicen de forma que permitan el acceso distribuido.
- Deben optar por la representación de datos apropiada ya que de ello depende el tipo de consultas que se podrán realizar al sistema.

Por muchos años, cada desarrollo SIG definió sus formatos de representación y almacenamiento para los datos geográficos, así como sus modelos conceptuales. A la hora de compartir se requieren establecer los formatos con los cuales se brindará la información (o se espera recibirla). El receptor posteriormente deberá convertirla al formato utilizado por su sistema. Los diversos formatos de IG existentes implican problemas de interoperabilidad entre diferentes SIG y aún entre aplicaciones desarrollados con la misma herramienta. Los costos y la posibilidad de automatizar la integración de datos se verán beneficiados si los formatos de los datos geográficos siguen algún estándar. Entre las propuestas de estándares para modelar e intercambiar datos y servicios geográficos se pueden citar por ejemplo, los propuestos por el consorcio Open Geospatial Consortium (OGC) ¹ [OGC94],

¹<http://www.opengeospatial.org>

el comité Técnico de normalización sobre Geomática e Información Geográfica ISO/TC 211² [ISO94], Spatial Data Transfer Standard (SDTS)³, y otras propuestas como la integración semántica guiada por ontologías para asegurar la integración semántica en el ámbito conceptual.

4.2. Interoperabilidad e Integración

Los sistemas y aplicaciones con datos geográficos y/o cartográficos surgen en la década del '70. Los primeros desarrollos se basaron en productos comerciales propietarios y con fuentes de datos propias, las que raramente se obtenían en forma digital. Los datos no se compartían entre organizaciones o empresas fundamentalmente porque era imposible. Cada SIG se podía considerar una “isla de información” [Bis98]. En las últimas décadas se viene trabajando fuertemente para revertir esta situación buscando contar con SIG “interoperables”.

Desde la perspectiva de Ingeniería de Software, la *calidad de interoperabilidad* se relaciona con sistemas con capacidades para integrar componentes de software que provienen de diferentes proveedores. Esto permite a un nuevo producto de software ingresar al mercado a competir con otros productos, rompiendo la dependencia cliente—proveedor, así como también ofrecer software integrable con otros módulos. En el ámbito de sistemas geoespaciales se requiere que SIG, bases de datos espaciales, servicios de mapas y otras aplicaciones sean capaces de integrarse de forma transparente.

Desde el punto de vista de los datos, el término *integración de información* se refiere a la necesidad de compartir datos y más específicamente información, esto es, datos y su significado. Si la información es producida y gestionada localmente, en general resulta no ambigua para su comunidad de uso. Sin embargo, cuando la misma ha sido producida por otros esto deja de ser cierto, y se requiere de conocimiento y reglas de transformación para integrar información “heterogénea”.

Los principales organismos trabajando en estandarización para asegurar interoperabilidad e integración en el ámbito de SIG son OGC e ISO/TC 211. La organización internacional OGC se encarga de promover y garantizar interoperabilidad para los componentes

²<http://www.isotc211.org/>

³<http://mcmcweb.er.usgs.gov/sdts/>

de software geoespacial. Se trata de un consorcio de compañías privadas, agencias gubernamentales y universidades que acuerdan en la especificación de interfaces de código abierto aplicables a sistemas de geoprocetamiento. Las *Especificaciones OpenGIS* proveen soluciones interoperativas. El comité ISO/TC 211 define normativas de representación para la IG digital y transferencia de datos y servicios que se ocupan de los métodos, herramientas y servicios para la gestión de datos, adquisición, procesamiento, análisis, acceso, presentación y transferencia de IG en formato digital entre diferentes usuarios, sistemas y localizaciones. El próximo capítulo de este trabajo (sección 5.2) presenta los principales resultados en estandarización para interoperabilidad e integración geoespacial.

En esta tesis nos hemos concentrado en identificar los problemas y los principales resultados obtenidos en el área integración de IG. Quedan fuera de este trabajo la investigación y el análisis de problemas y resultados relacionados con la interoperabilidad de servicios geográficos. En las secciones siguientes de este capítulo se presentan los principales problemas en integración geográfica y se los clasifica en distintos niveles de heterogeneidad. En los próximos capítulos se recopilan los principales esfuerzos y resultados alcanzados en integración de IG.

4.2.1. Heterogeneidad de información

La integración de datos, más allá del tipo de sistemas a integrar, presenta una serie de problemas que, en general, son complejos de resolver. El concepto de integración en sí mismo involucra decisiones a tomar en forma correcta para lograr un resultado consistente. Las principales fuentes de información son almacenamientos de datos persistentes o bases de datos. La integración de los diversos modelos de datos que utilizan las distintas soluciones de un mismo problema es clave para compartir información. Usando el recurso de abstracción se identifican dos vistas para el problema de heterogeneidad en bases de datos: *heterogeneidad estructural* y *heterogeneidad semántica* [KS96]. Se dice que dos soluciones presentan heterogeneidad estructural si almacenan sus datos bajo modelos de representación diferentes. Este tipo de heterogeneidad se relaciona con cuestiones “técnicas” o de un nivel de abstracción bajo y resultan más simples de generalizar. Desde otro punto de vista, la heterogeneidad semántica considera el “significado” de cada elemento de información. Los conflictos semánticos ocurren cuando en contextos diferentes no se utilizan las mismas interpretaciones. En general, los problemas de heterogeneidad semántica

son más difíciles de comprender, generalizar y solucionar. Es necesario definir reglas de integración semántica para encontrar las correspondencias o “mapeos” entre los modelos.

Ejemplo 4.2.1 Sean dos sistemas de información inmobiliarios S y T interesados en compartir información, ambos desarrollados bajo el modelo relacional. S considera dos esquemas de relación: PROPIEDADES con los atributos ubicación, precio e idAgente y AGENTES con los atributos idAgente, nombre, domicilio, ciudad, provincia y comisión. En esta organización existen las relaciones “propiedades” y “agentes”, vinculadas por el identificador de agente. La solución T por su parte considera un único esquema OFERTAS con los atributos zona, precioOferta, agente, contacto y una única relación “ofertas” (figura 4.1). A partir de estas dos propuestas se puede realizar el siguiente análisis sobre heterogeneidad:

- *Heterogeneidad estructural: no existe. Ambas soluciones utilizan como modelo de representación al modelo Relacional.*
- *Heterogeneidad semántica: claramente el modelo conceptual elegido por ambas soluciones es distinto. Sin embargo, se pueden plantear las siguientes reglas de correspondencia semántica:*
 - *el atributo “ubicación” de S.propiedades se corresponde con el atributo “zona” de T.ofertas.*
 - *el atributo “precio” de S.propiedades se corresponde con el atributo “precioOferta” de T.ofertas.*
 - *el atributo “nombre” de S.agentes se corresponde con el atributo “agente” de T.ofertas.*
 - *los atributos “domicilio+ciudad” en S.agentes se corresponden con el atributo “contacto” de T.ofertas.*

Para definir correspondencia semántica se requiere decidir si dos elementos s de S y t de T refieren a un mismo concepto, lo que no es fácil por varias razones:

- Existen pocas fuentes de información de dónde se pueda recuperar la semántica de los elementos. En general, el conocimiento semántico se debe capturar desde los

diseñadores del modelo, que no siempre están disponibles; desde la documentación, también no siempre disponible y posiblemente desactualizada; o desde el “reconocimiento” de los datos asociados.

- Aunque existen pistas tanto en los *esquemas* como en los *datos*, éstas pueden llegar a ser falsas. Dos elementos que comparten un mismo nombre de atributo pueden no referirse al mismo concepto real. Por ejemplo, un atributo *contacto* en un modelo podría referir a un nombre de persona y en otro a un teléfono. La inversa también es cierta: existen conceptos idénticos que se modelan con distintos nombres de atributo según el modelo. En el ejemplo de la inmobiliaria, los atributos *ubicación* y *zona* referencian al mismo concepto del mundo real.
- Decidir si un elemento s de S corresponde con un elemento t de T requiere que se examinen todos los otros elementos de T para asegurar que no hay otro elemento con el que hacer una correspondencia mejor.
- La correspondencia entre elementos es subjetiva y depende del entorno de aplicación. Por ejemplo, el atributo *propiedad-descripción* podría corresponder con *propiedad-estilo* en dos aplicaciones inmobiliarias, pero seguramente refieren a cosas distintas si el atributo *propiedad-estilo* está en el contexto de un modelo de datos de un sistema de turismo.
- Encontrar correspondencias entre elementos no se limita únicamente a definir equivalencias entre ellos. Dos elementos podrían estar vinculados por otros tipos de relaciones semánticas como por ejemplo: el concepto A “es una especialización de” B (*calleVecinal* es un tipo de *camino*), el concepto A “es uno de los componentes de” B (*parcela* es una parte de *manzana*), además de otras asociaciones diversas que se quieran conocer dentro del dominio de aplicación (*río* “desemboca en” *río* ó *río* “desemboca en” *mar*).

En el ejemplo 4.2.1 se muestra cómo se pueden establecer correspondencias semánticas entre estructuras o esquemas. La correspondencia entre *instancias de datos* enfrenta dificultades similares. Se trata de determinar si dos *elementos de datos* se corresponden con la misma información. Aplicado al ejemplo inmobiliario, se refiere a cómo identificar si ambos sistemas comparten una o más propiedades.

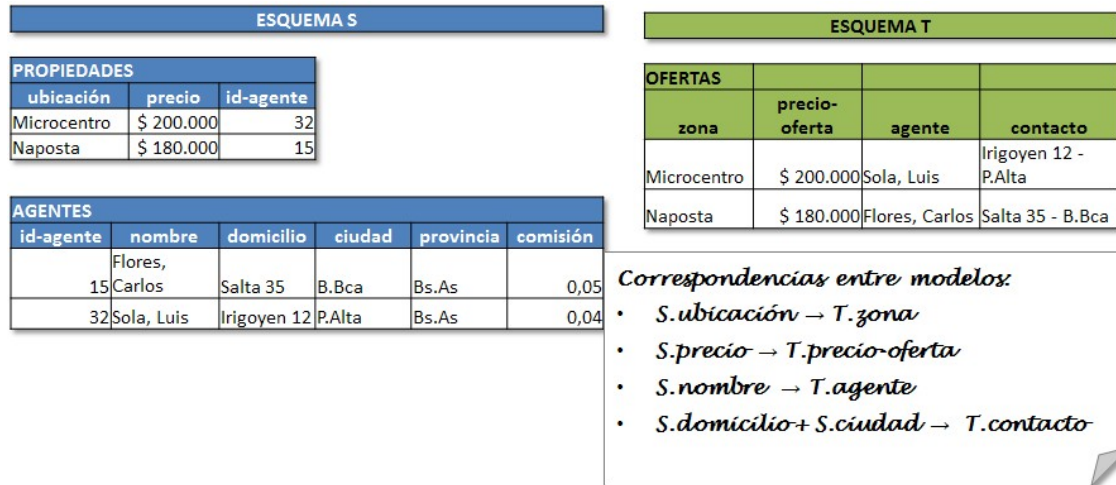


Figura 4.1: Dos modelos de datos para un mismo problema inmobiliario

En general se reconoce que encontrar correspondencias en forma manual es muy costoso. Existen algunos intentos de sistemas de integración que utilizan técnicas automáticas o semi-automáticas para encontrar correspondencias [MKS⁺05, MDVK, DDH01, DH05]. Investigaciones sobre este problema han tomado relevancia debido a la necesidad de administrar volúmenes muy importantes de datos distribuidos y heterogéneos.

4.2.2. Integración de bases de datos heterogéneas

Se definen como *sistemas de bases de datos distribuidos heterogéneos* a los entornos con varios sitios o nodos (servidores de bases de datos), con diferentes esquemas e instancias de base de datos y software de base [SKS10]. En un entorno distribuido no se requiere que todos los sitios conozcan de la existencia del resto y es posible que sólo estén dispuestos a compartir parcialmente sus datos. Las diferencias de esquemas entre bases causan problemas en el procesamiento de las consultas; estas diferencias, como se explicó anteriormente, se relacionan con la heterogeneidad semántica. Las diferencias en el modelo de representación (por ejemplo, relacional y orientado a objetos) se ubican dentro de la heterogeneidad sintáctica.

Sobre heterogeneidad semántica en bases de datos, las técnicas de correspondencia entre modelos se organizan en dos grupos que dan origen a distintos sistemas de integración: *soluciones basadas en reglas* y *soluciones basadas en aprendizaje*. Obviamente a

partir de estas soluciones de integración existen sistemas que combinan ambas propuestas. Las soluciones basadas en reglas automáticas o semi-automáticas emplean reglas de correspondencia entre esquemas en forma similar a lo que se haría en forma manual. En general, exploran metadatos de los esquemas: nombre de los atributos, tipos de dominio, estructuras, número de subelementos y restricciones de integridad. Algunos sistemas de integración basados en reglas son: TranScm [MZ98], DIKE [PU98], MOMIS (ARTEMIS) [CdA99] y CUPID [MBR01]. Estos sistemas calculan mapeos entre esquemas basándose en similitudes encontradas. Como ventajas de los sistemas basados en reglas se puede decir que son eficientes y funcionan en forma apropiada para cierto tipo de dominios de aplicación. Las principales desventajas están en que no explotan efectivamente las instancias de datos, existen combinaciones de información entre esquemas que son complicadas de comparar y a pesar de que encuentran coincidencias, es difícil garantizar que éstas son correctas. Estos sistemas además no tienen capacidad para aprender del conocimiento ya adquirido.

Las soluciones basadas en aprendizaje usan técnicas de inteligencia artificial y exploran información tanto de los esquemas como de los datos en sí mismos. Para encontrar coincidencias analizan la especificación de los atributos (tipos de datos, escalas, restricciones), información estadística (valores mínimos, máximos, promedios y varianzas) y la naturaleza jerárquica de los datos. Además, estos sistemas son capaces de aprender a partir de “reglas de coincidencia” encontradas y predecir satisfactoriamente futuras correspondencias. Algunos desarrollos en este sentido son: Semint [LC00], LSD [DDH01], iMAPS [DLD⁺04].

También existen trabajos que explotan la noción de correspondencia entre *colecciones de esquemas*. Este escenario es típico si se exploran modelos de varias fuentes de datos. Los resultados en este sentido demuestran que buscar coincidencias entre un número mayor de fuentes de información favorece encontrar mejores coincidencias que entre dos fuentes específicas [MKS⁺05, MDVK].

El problema de encontrar correspondencias varía si se consideran sólo relaciones 1:1 del tipo “dirección es equivalente a domicilio” a si se pretende además considerar relaciones más complejas, como “nombre” en el esquema E_1 se corresponde con la concatenación de *primerNombre* y *segundoNombre* y *apellido* del esquema E_2 .

Los trabajos más recientes han desarrollado técnicas para encontrar mapeos complejos basándose en ontologías de dominio [XE03]. Usan una combinación de búsquedas y

técnicas de aprendizaje y técnicas de *data mining*. Por ejemplo, para encontrar mejores mapeos entre dos esquemas, primero se hace un mapeo de ellos a una ontología de dominio y se construyen las coincidencias basándose en las relaciones heredadas del dominio. Se destaca la importancia de contar con *conocimiento del dominio* para obtener mapeos complejos. Este conocimiento colabora en la construcción de correspondencias múltiples, eliminando tempranamente los caminos incorrectos, evaluando sólo los caminos posibles en un tiempo de ejecución aceptable. Sobre ontologías de dominio nos ocuparemos en el capítulo 6 de esta tesis.

Otro enfoque para integración es establecer correspondencia entre *datos*. Incluye decidir si dos objetos de distintas fuentes refieren a la misma entidad real [BM03, AMT04]. El ejemplo más popular de correspondencia de datos se aplica al dominio de material de investigación y la búsqueda de correspondencia entre publicaciones, autores e instituciones. La problemática de correspondencias de datos se conoce bajo distintos nombres: limpieza de datos, eliminación de duplicados, consolidación de información y correspondencia entre objetos, entre otros, y constituye un área de investigación en Inteligencia Artificial, Minería de Datos y Recuperación de Información en la Web. Los sistemas de integración de datos también incluyen soluciones basadas en reglas y basadas en aprendizaje. Las primeras propuestas emplearon reglas definidas de forma manual [HS95], las que posibilitan que se adquieran las próximas reglas de mapeo [SB02]. Otras soluciones se enfocan en técnicas de mapeo para cadenas [ME96]. Otras se ocupan del problema de escalabilidad para bases de datos muy grandes. Los métodos más recientes usan recuperación de información [AMT04] y técnicas de teoría de la información.

4.3. Integración de Información Geográficos

4.3.1. Caracterización de la información geográfica

La IG distingue los componentes: *espacial* y *temático*⁴. El componente espacial mantiene la referencia a un espacio, que para el caso de interés sobre IG es la superficie terrestre. El componente temático tiene que ver con la descripción de las entidades físicas o abstractas de dicho espacio geográfico. La integración semántica en IG se relaciona con el componente temático.

⁴Opcionalmente puede existir un *componente temporal*.

Las variables temáticas o atributos se clasifican en *cualitativas* o *cuantitativas* según su dominio de valores. Las variables cuantitativas tienen por rango valores numéricos, que pueden ser continuos, como es el caso de temperatura o precipitación; o discretos como cuando se refiere a población o cantidad de especies. El valor de un atributo cuantitativo a su vez puede ser una medida directa como es el caso de población o temperatura, o puede ser calculado como por ejemplo **ingreso-per-cápita**. Otra característica de las variables temáticas cuantitativas es que opcionalmente se pueden agrupar en intervalos. Así, por ejemplo, la temperatura se podría clasificar según un rango de valores en “caluroso”, “templado” y “frío”. Cada rango estará asociado con intervalos de valores. De esta manera, una variable cuantitativa clasificada tiene un rango que consiste en conjuntos numéricos de intervalos disjuntos. Los aspectos semánticos de una variable temática cuantitativa tienen que ver justamente con el tema. Por ejemplo, para un atributo temperatura: las unidades de medida, grados Celsius o Fahrenheit; cómo se tomaron las muestras, si el conjunto de datos representan temperaturas mínimas y máximas; si el valor mínimo representa el de todo un año o un promedio de valores mínimos diarios. Problemas semánticos adicionales ocurren cuando las variables cuantificadas se organizan por rango donde, por ejemplo, los intervalos son distintos o se define una cantidad de subrangos diferentes.

Las variables temáticas cualitativas o variables nominales, tienen asociado un conjunto de valores nominales discretos y cada valor típicamente está vinculado a un término. Es frecuente el uso de IG cualitativa relacionada con áreas como “geomorfología”, “uso del terreno” y “tipos de cultivos” entre otros. Las variables temáticas cualitativas relacionan una ubicación con un concepto temático que representa una cualidad de la variable. Algunos ejemplos de valores para la variable cualitativo TIPO DE REGIÓN son “zona de pinos”, “área de aluviones”, “territorio de yacimientos”. Desde el punto de vista conceptual, el valor de un tema cualitativo es un concepto para el término o términos que representa. Cada uno de los valores cualitativos corresponde a un tema geográfico, que es más específico que el tema principal del mapa, por ejemplo, el valor “ganado vacuno” en el mapa sobre producciones agrarias. También podría ser el tema principal en otro mapa, donde los subtipos diferentes de razas vacunas serían los valores cualitativos. Este ejemplo ilustra cómo valores cualitativos de fuentes de datos diferentes podrían estar relacionados y hasta conformar una taxonomía de temas. En la figura 4.2 se pueden observar dos visualizaciones de mapas con atributos temáticos cuantitativos y cualitativos respectivamente. En general, está establecido que la semántica para variables temáticas cuantitativas es más simple que para variables temáticas cualitativas ya que el significado de los valores

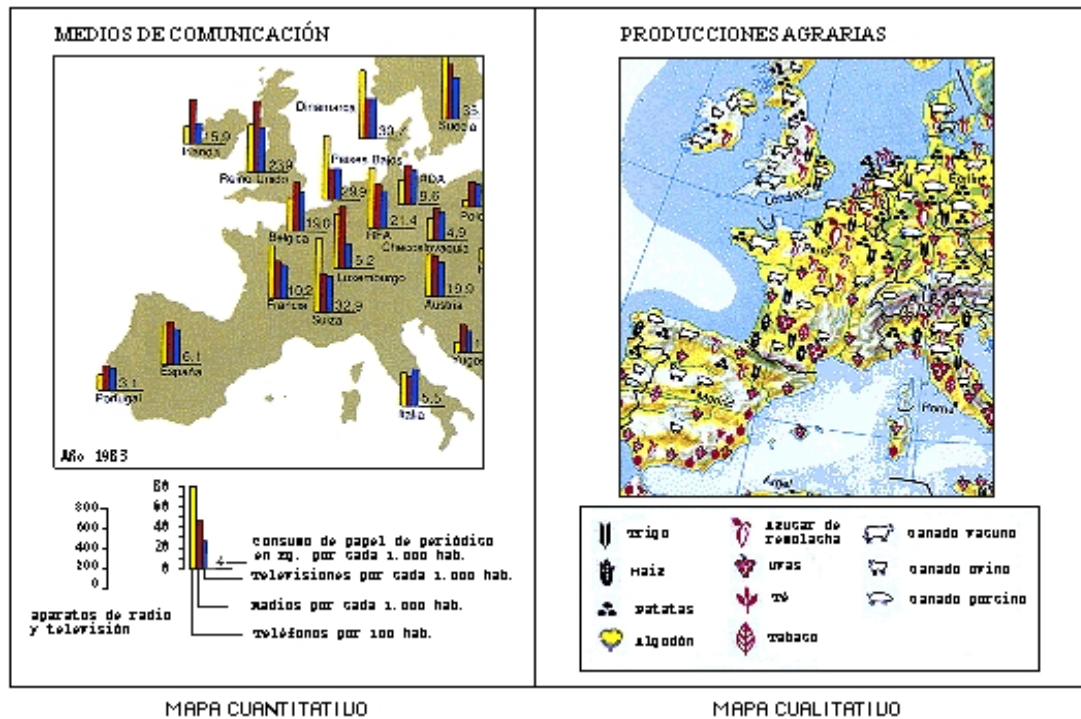


Figura 4.2: Mapas Temáticos de Europa - variables cualitativas y cuantitativas

cuantitativos es menos propenso a ambigüedades.

4.3.2. Heterogeneidad en información geográfica

Según [Bis98] “la comunidad de IG identifica fenómenos del mundo real, les asocia un modelo mental conceptual y luego elige una representación en un modelo lógico discreto”. Estos modelos varían de una representación a otra (sección 2.3). En los diferentes modelos de IG se identifican los siguientes tipos de heterogeneidad especialmente referida a SIG [Ter06]:

- *Heterogeneidad Sintáctica*: se refiere al modelo de datos utilizado por distintas soluciones. Diferentes SIG ofrecen soluciones que utilizan diferentes formatos y sus datos espaciales podrían estar representados con distintos modelos (vector o raster) o inclusive referenciar a distintos sistemas de coordenadas. En este sentido, el Lenguaje de Marcado Geográfico (*Geographic Markup Language - GML*) (OGC 2004)

es una especificación OGC que provee un formato común para representar IG y da una solución para los problemas de heterogeneidad sintáctica (ver sección 5.4)

Ejemplo 4.3.1 *Dos soluciones SIG S_1 y S_2 incluyen a la variable temática “cubierta forestal” con la representación de las áreas forestales de una región. Mientras que S_1 optó por el modelo vectorial y la descripción de cada área incluye al componente geométrico $GM_Surface$, S_2 eligió la representación espacial mosaico manteniendo para cada celda si pertenece o no a un área forestal.*

- *Heterogeneidad Estructural:* se relaciona con la forma en que se representan los modelos mentales, por ejemplo, en el caso de SIG usando distintos atributos. Los metadatos describen la estructura del esquema de representación y se requiere de herramientas que sean capaces de manejar esta heterogeneidad estructural.

Ejemplo 4.3.2 *Dos soluciones SIG S_1 y S_2 consideran el modelado del tema “cubierta forestal”. El modelo considerado por S_1 incluye información de cada área sobre si se trata de un área protegida o de un recurso económico. Por otro lado, S_2 no registra esas propiedades y sí mantiene información sobre la normativa que regula cada área y los planes de reforestación.*

- *Heterogeneidad Semántica:* se refiere al hecho que distintos agentes (personas u organizaciones) usan distintos esquemas mentales, esto es, categorizan el mundo real de diferentes maneras. En el caso de SIG, estas categorías corresponden a conceptos temáticos, y en consecuencia la semántica está principalmente relacionada con el componente temático de la IG que se está modelando.

Ejemplo 4.3.3 *Dos soluciones SIG S_1 y S_2 consideran el modelado del tema “cubierta forestal”. Sin embargo, S_1 considera como área forestal a las áreas cubiertas de montes, bosques o selvas, mientras que S_2 define como área forestal únicamente a las regiones de bosque nativo o implantado.*

Por *integración geográfica* se refiere a dotar a la información con “conocimiento” para que agentes externos de software puedan encontrar, integrar y compartir IG de distintas fuentes de origen. Las propuestas de solución para la gestión de heterogeneidad sintáctica se encaminan por el uso de modelos de representación estándar junto con traducciones

entre modelos. Por otra parte, los problemas de integración estructural y semántica han motivado importantes investigaciones en el área de Ciencias de la Información Geográfica. Las soluciones se vinculan con la definición de ontologías que enriquezcan cada solución SIG con la especificación formal de sus conceptos. También se requieren de servicios semánticos con capacidad para integrar datos que provienen de diferentes fuentes de información. La lógica que fundamenta a las ontologías provee la base para estos servicios.

El uso de sistemas especificados por ontologías trae nuevos problemas de integración: ¿cómo integrar sistemas de distintas ontologías? La clasificación provista por el Consorcio de Web Semántica [MSv⁺05] identifica los siguientes tres subniveles de heterogeneidad semántica ontológica:

1. El *nivel sintáctico*, que se refiere al hecho que distintas ontologías podrían estar expresadas en diferentes lenguajes (OWL o KIF).
2. El *nivel terminológico*, comprende las diferencias en la denominación de los conceptos. En este nivel se incluye a sinónimos, diferencias relacionadas con el idioma (Español, Inglés, etc), polisemias o sobrecarga de un término y derivados (sufijos o prefijos).
3. El *nivel conceptual*, incluye las diferencias relacionadas con el contenido de la ontología. Existen dos tipos de diferencias conceptuales: diferencias *metafísicas* que se refieren a “cómo está dividido el mundo en piezas” y diferencias *epistémicas*, que se refieren a “cómo se entienden las entidades”. En el cuadro 4.1 se clasifican y ejemplifican las diferencias semánticas entre ontologías.

La interoperabilidad semántica está relacionada con la noción de *Web Semántica*. Se buscan formas de representación de la información en la web que pueda ser interpretada por personas y máquinas. Esto permitiría hacer un uso más inteligente de la información, y compartir y reusar datos entre aplicaciones. En particular, en [Ege02] se define la necesidad de una *Web Semántica Geoespacial*, basada en un marco que abarque múltiples ontologías espaciales temáticas, así como una forma canonizada de especificar consultas geoespaciales. Relacionado con el concepto de interoperabilidad geoespacial está la noción de IDE. Un IDE se define como la infraestructura que provee el marco de trabajo para la optimización de tareas relacionadas con la creación, mantenimiento y distribución de IG

Integración de Información Geográfica basada en Ontologías			
<p>Heterogeneidad semántica en SIG: agentes distintos representan fenómenos del mundo de manera diferente.</p> <p>Las diferencias se vinculan con conceptos temáticos.</p>	<p>Diferencias metafísicas: cómo se divide el mundo en piezas (entidades, propiedades y relaciones).</p>	<p>Cubrimiento: ontologías que cubren diferentes porciones del mundo.</p>	<p>O_1 considera establecimientos educativos de una ciudad y O_2 de otra.</p>
		<p>Granularidad: ontologías con distinto grado de detalle para los mismos conceptos.</p>	<p>O_1 trata sobre información censal y toma valores por <i>provincia</i> mientras que O_2 considera por <i>distrito</i>.</p>
		<p>Perspectiva: ontologías diferentes para el mismo mundo real pero con distintos puntos de vista.</p>	<p>O_1 y O_2 modelan la misma región, una analiza la fauna desde el punto de vista del ecosistema y otra desde el punto de vista antropológico.</p>
	<p>Diferencias epistémicas: cómo se entienden las entidades, esto es, qué aserciones representan.</p>	<p>El conjunto de conocimientos condicionan las formas de entender e interpretar el mundo.</p>	<p>O_1 considera “vías de comunicación” como rutas y caminos y O_2 interpreta a los medios de comunicación e incluye “telecomunicaciones”.</p>

Cuadro 4.1: Heterogeneidad semántica en ontologías

a diferentes niveles (regional, nacional o global) e involucra tanto a instituciones públicas como privadas. Como una de las principales responsabilidades de los servicios IDE está recuperar metadatos y traducirlos a esquemas estándares.

Ejemplos de diferencias semánticas en IG

Como ya se observó, el hecho que los datos a procesar provengan de distintas comunidades es causa de heterogeneidad sintáctica, estructural y semántica [BEH⁺03]. Ocurre que distintos conceptos refieren al mismo tipo de información en distintas fuentes de datos o, por el contrario, diferentes orígenes de datos usen el mismo término para referirse a distinta clase de información. A continuación se plantean algunos ejemplos de diferencias semánticas en integración de IG.

Ejemplo 4.3.4 *Consideremos un organismo nacional que mantiene información sobre usos de la tierra en Argentina. Probablemente sus necesidades son diferentes de las de una municipalidad que lleva la misma información, pero reducida a su propio municipio, por ejemplo, Bahía Blanca.*

A nivel país, seguramente interesará mantener información que describa categorías de orden general: sector urbano, agricultura, industria, etc., mientras que a nivel local, el municipio usará conceptos más específicos relacionados con su realidad, objetivos y regulaciones locales. Por ejemplo, mientras que a nivel país puedan interesar las *áreas urbanas*, a nivel local, el municipio puede estar interesado en describir áreas urbanas que carecen del servicio de gas natural.

Seguramente, si existen soluciones implementadas por cada organismo, las mismas estructuraron la IG de distinta manera. Las entidades o conceptos se eligen de acuerdo con las necesidades específicas del propietario de la información. Así, aunque ambos proyectos comparten la misma realidad, la representan de manera distinta. Este ejemplo muestra un caso de heterogeneidad semántica metafísica, particularmente por utilizar conceptos con diferente granularidad. El nivel de detalle de los conceptos definidos está directamente afectado por la escala y el tamaño mínimo considerado para las unidades espaciales, que determina el tipo de representación y el modelo mental subyacente.

Ejemplo 4.3.5 *Un experto en ecología y un experto en geomorfología observan una misma porción del mundo real.*

En este ejemplo, aunque en ambos casos se utilice la misma escala de información, seguramente uno y otro se enfocarán en distintas categorías. Mientras que el primero puede distinguir secciones como “sector de coníferas”, el último tal vez use categorías como “zona de inundación”. Este es otro ejemplo de heterogeneidad semántica metafísica, aunque la misma es producida por perspectivas o enfoques diferentes para analizar una misma región.

Ejemplo 4.3.6 *Distintos productores y obviamente consumidores de IG mantienen modelos mentales diferentes para el mismo concepto “terreno edificable”.*

La IG generada por distintos productores es localmente correcta, pero ha sido producida para distintos propósitos, con distintos esquemas mentales y, consecuentemente, con distintos esquemas de representación. En este caso se dice que los distintos conjuntos de datos presentan heterogeneidad semántica epistémica. En este punto se resalta que de ninguna manera debe pensarse que alguna de las ontologías es incorrecta. El objetivo es encontrar una solución interoperativa a estos problemas, la que no se puede basar en restringir las libertades de los productores de IG tal como la entiendan y requieran.

Aunque en el entorno de las geociencias está consensuada la voluntad de comunicar y compartir, cada esquema de solución define su propia *ontología de aplicación*, que conceptualiza de distinta manera según las propias necesidades. Se requiere que las ontologías de aplicación se expresen de manera formal. Así, definir una ontología de mayor nivel que se obtenga a través de un proceso de “merging” o mezcla y represente el conocimiento completo especificando las relaciones semánticas entre distintas ontologías de aplicación. El problema de heterogeneidad semántica también causa conflictos durante el proceso de búsqueda de fuentes de información [LRP03]. Por ejemplo, un servicio cuya tarea se describe como *clasificación de zonas inundables en imágenes satelitales* podría no estar seleccionado con un buscador tradicional si se buscan las palabras claves *servicios de clasificación para datos sensados remotamente*.

4.4. Solución a los problemas anteriores

De lo expuesto en las secciones anteriores se concluye que se necesitan resolver problemas técnicos y semánticos en integración para IG. Los estándares tecnológicos desarrollados por OGC e ISO/TC 211 proveen las bases para la interoperabilidad sintáctica

y catálogos de datos geográficos y servicios. La solución a problemas semánticos, como ocurre en otros entornos de información, requiere de un marco que formalice las relaciones semánticas entre conceptos de distintas fuentes de datos. A continuación se presentan algunas soluciones parciales basadas en vocabularios y tesauros. En los capítulos siguientes se desarrollan en detalle las propuestas en estandarización (capítulo 5) y soluciones semánticas basadas en ontologías (capítulo 6).

4.4.1. Vocabularios y tesauros

Una solución parcial al problema de integración semántica es la definición de un *vocabulario compartido* por una comunidad. Un vocabulario está formado por un conjunto de términos relacionados por su significado. Las relaciones básicas que se definen entre términos de un vocabulario incluyen *término más amplio* (o más abarcativo), *término más específico* y *término sinónimo*. Existen algunos vocabularios normalizados para algunas subáreas de geociencias [Ter06]. Por ejemplo, el vocabulario usado en CORINE [BO00] sobre cobertura y uso del territorio de Europa.

Cuando el vocabulario está más formalizado y acuerda una jerarquía de conceptos se denomina *tesauro* (ejemplo 4.4.1). Informalmente un tesauro es una “estructura jerárquica” y “controlada” de términos. Los términos comprenden a una o más palabras que se organizan a través de un conjunto de relaciones recíprocas entre ellas. En el cuadro 4.2 se definen las relaciones estándar que se plantean entre términos de un tesauro para organizarlos como una jerarquía.

Ejemplo 4.4.1 *Ejemplo de un extracto de tesauro:*

- *Relieve*
 - *Elementos del relieve*
 - *Cerros Testigo*
 - *Colinas*
 - *Collados*
 - *Dunas*
 - *Gargantas*
 - *Litoral*

- ◇ *Bahías*
- ◇ *Acantilados*
- ◇ *Dunas*

Los nombres de estas relaciones (BT, NT, TT, RT, USE y UF) están regulados por los estándares ISO 2788:1986 [ISO07f] e ISO 5964:1985 [std85], normas internacionales para la definición de tesauros mono-lengua y multi-lengua respectivamente.

Ejemplos de vocabularios y tesauros en el ámbito geográfico

El vocabulario CORINE [BO00] sobre cobertura y uso del territorio de Europa distingue 44 tipos diferentes de cubiertas del territorio. Cada país, si lo desea, puede refinar estas categorías. CORINE Land Cover actúa como una articulación entre el conocimiento local (micro) y la Unión Europea (macro) sobre información ambiental y colabora con la visión micro/macro para la definición de políticas sobre el medio ambiente.

El tesauro GEMET [EEA01] definido sobre el dominio Gestión Ambiental, incluye además las relaciones “GROUP” y “THEME”, como opciones de indexación de términos por grupo y tema. En la figura 4.3 se observa la visualización en el tesauro GEMET del término “zona inundable”. Las relaciones subyacentes entre “zona inundable” con otros elementos del GEMET se muestran en el cuadro 4.3.

También están relacionados con vocabularios en el dominio geográfico los trabajos sobre Land Cover Classification System (LCCS) [DGJ98], que definen tipos de cubiertas del terreno. LCCS se utiliza en el marco de producción de mapas sobre destino de la tierra en varios países de África. Sin embargo, en el LCCS no es un vocabulario normalizado, sino que provee de un meta-vocabulario para crear nuevos vocabularios, especificando cómo cada clase del vocabulario se determina por distintitos clasificadores. Otra iniciativa relacionada con tesauros para términos geográficos en el contexto de los nomencladores de la web es el tesauro GETTY de nombres geográficos [Ins00]. Estas opciones no incluyen a todos los conceptos para mapas temáticos.

Un inconveniente en el dominio de IG es que aún no existe un vocabulario compartido global, ya que el dominio en sí mismo es muy amplio. Las distintas iniciativas definen jerarquías normalizadas de términos para algunos dominios más específicos.

Relaciones estándar			
Relación	Sigla	Descripción	Ejemplo
<i>TérminoAbarcativo</i>	BT	Representa la relación entre un término más general y otro más específico	CALLE “es BT de” AVENIDA
<i>TérminoEspecífico</i>	NT	Inversa de BT. Representa la relación entre un término más específico hacia un término más general	AVENIDA “es NT de” CALLE
<i>TérminoRaíz</i>	TT	Indica si un BT es raíz de la jerarquía	VIA “es TT” en la jerarquía de términos de un tesoro sobre vías de transporte
<i>TérminosRelacionados</i>	RT	Relación recíproca que representa un vínculo diferente del jerárquico	ZONA DE MADERAS “RT” REFORESTACIÓN
<i>Usado</i>	USE	Representa la relación entre un término no preferido y su preferido	AREA (término no preferido) “USE” ZONA (término preferido)
<i>UsadoPor</i>	UF	Inversa de USE. Sólo los términos preferidos participan de otras relaciones.	ZONA (término preferido) “UF” AREA (término no preferido), ZONA (término preferido) “UF” REGIÓN (término no preferido).
Relación optativa			
Relación	Sigla	Descripción	Ejemplo
<i>Define</i>	DEF	Representa la relación entre un término y su definición en formato texto	<i>Masa de agua acumulada en la depresión de un terreno</i> “DEF” LAGO

Cuadro 4.2: Definición de las Relaciones entre Términos de un Tesoro

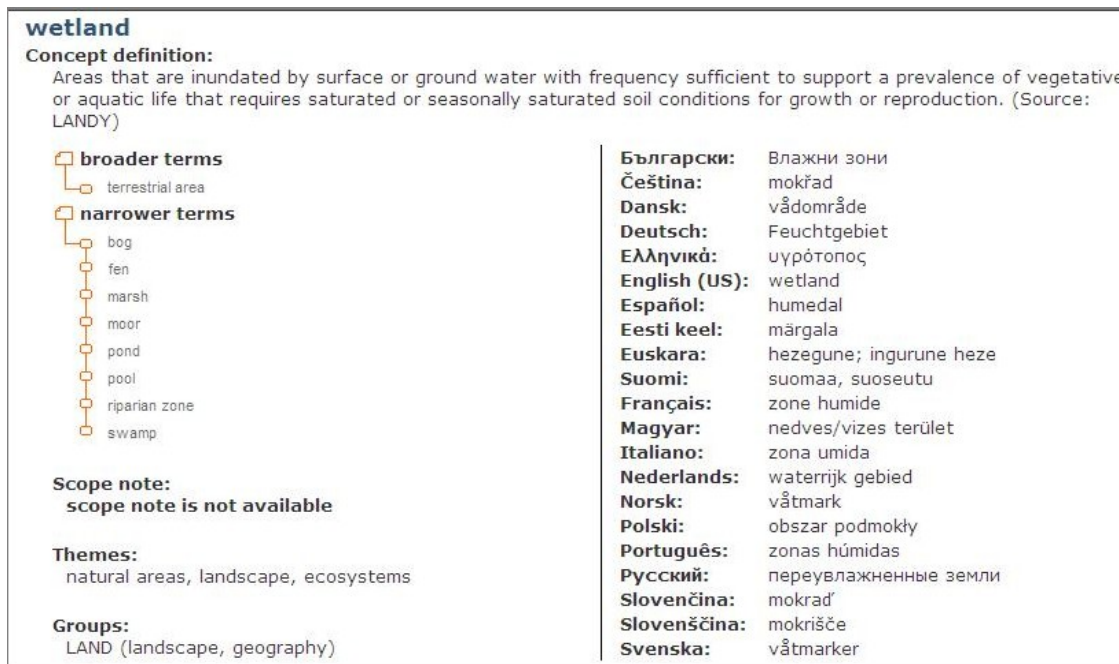


Figura 4.3: Definición del Thesaurus GEMET para el concepto “Wetland”

4.4.2. Limitaciones de los tesauros

El usar únicamente vocabularios o tesauros como base para una aproximación de interoperabilidad semántica presenta distintos problemas. Ningún tesoro cubre todas las subáreas de IG, sino que generalmente se especializa en algún subtema particular. Otra limitación es que los tesauros, como por ejemplo GEMET, tienen como objetivo definir conceptos temáticos, pero su estructura no está orientada a organizar modelos de datos. Esto implica que si, por ejemplo, buscamos el término “Foresta”, no existe un término abarcativo que describa a la foresta como un tipo específico de “Área Vegetal”. Esta información sería necesaria para representar las relaciones semánticas entre valores de distintos modelos de datos enfocados en problemas diferentes.

Un tercer problema es que los términos que se encuentran en un vocabulario o tesoro se definen teniendo presente un contexto particular y con el objeto de evitar ambigüedades, que es el uso principal de un buen tesoro. De esta forma, para que un tesoro sea claro y no ambiguo, debe restringirse a un contexto específico. Por ejemplo, en el caso de CORINE se enfoca sobre una escala 1:10000 donde la unidad espacial menor considerada tiene una extensión de al menos 25 hectáreas.

Zona Inundable	
Relación	Instancias de la relación
DEF	<i>Áreas que son inundadas desde la superficie o aguas subterráneas con la frecuencia suficiente para apoyar un predominio de vida vegetativa o acuática que tiene las condiciones de saturado para el crecimiento o la reproducción</i>
UF	ZONA ENCHARCADA
THEMA	AREA NATURAL, PAISAJE, ECOSISTEMA
GROUP	TIERRA (Paisaje, Geografía)
TT	<i>tierra</i>
BT	<i>area terrestre</i>
NT	<i>pantano, estanque, piscina, zona ribera</i>

Cuadro 4.3: Relaciones entre término ZONA INUNDABLE y otros elementos del tesauo

Un cuarto inconveniente está en que los términos del tesauo están definidos por descripciones textuales y no es posible derivar conclusiones lógicas. Esta limitación es particularmente significativa en el caso de temas modelados donde el mismo término, como por ejemplo “Terreno apto para construcción”, el cual tiene definiciones distintas según se consideren legislaciones locales o de planeamiento urbano.

Así, una aproximación basada únicamente en la definición de un tesauo restringe las posibilidades de descripción de un concepto real vinculado con IG. También es cierto que dominios que están en constante evolución requieren de la evolución de los tesauos. Esto trae como consecuencia problemas de mantenimiento y de vinculación de información diseñada bajo diferentes versiones de un mismo tesauo. Un resultado más adecuado es combinar el uso de vocabularios con el de ontologías que conviertan términos en clases ontológicas, con definiciones de términos más amplios y más específicos en relaciones de generalización-especialización, sinónimos en relaciones de equivalencia y otras asociaciones no restringidas entre clases. Así los valores se relacionan con clases que están claras, aunque el vocabulario en sí mismo no se referencie en la fuente de datos. El uso de ontologías para la integración de IG se va a desarrollar en el capítulo 6 de esta tesis.

4.5. Resumen del capítulo

El campo de aplicación de la IG es muy amplio incluyendo turismo, comunicación, defensa nacional, transporte, agricultura, etc. La mayoría de estas áreas están interesadas en compartir información y servicios en línea a través de la Web. Las calidades de interoperabilidad e integración de datos ganan importancia en el campo de los productos tecnológicos de información. Las aplicaciones que trabajan con IG al momento de intercambiar información enfrentan diferencias tanto a nivel de modelos del dominio como también de modelos de representación, especialmente para datos geográficos heredados desde sistemas de geoprosesamiento desarrollados careciendo de estandarización.

En este capítulo se organizaron los problemas de heterogeneidad en IG en dos grandes tipos: heterogeneidad sintáctica y heterogeneidad semántica. Desde la vista sintáctica, si bien existen diferentes formatos y modelos (vector o raster) y hasta distintos sistemas de coordenadas, desde la década del '90 se viene trabajando para definir normativas en el Comité Técnico Internacional ISO/TC 211 y la industria representada por Open Geospatial Consortium (OGC). En el capítulo 5 de esta tesis se presentan y analizan los principales resultados en estandarización alcanzados hasta el momento.

En la vista de heterogeneidad estructural y heterogeneidad semántica se considera cómo se representan los modelos mentales para el de caso desarrollos SIG usando organizaciones temáticas distintas. Distintas soluciones categorizan el mundo real de diferentes maneras. Una aproximación es la definición de tesauros compartidos; sin embargo, éstos no resuelven todas las relaciones entre valores temáticos de distintas fuentes de datos y contextos, ya que el tesoro en si mismo tiene una tendencia según su propio contexto. El problema de integración semántica de IG ha motivado importantes investigaciones en el área de Ciencias de la Información Geográfica. Las soluciones generalmente se vinculan con la definición de ontologías que provean la especificación formal de los modelos de representación de cada SIG. En el capítulo 6 de esta tesis se abordará el uso de ontologías aplicado al ámbito de IG.

Capítulo 5

Estándares Internacionales para Sistemas de Información Geográfica

A lo largo de las décadas distintas organizaciones han recolectado y gestionado IG siguiendo modelos y políticas individuales. Hoy en día hay una fuerte exigencia por integrar distintos servicios web a los entornos operacionales. El último impulso surge con la industria de servicios basados en ubicación (LBS). Como se vio en el Capítulo 2 existen varias aproximaciones para modelar el espacio geográfico y los fenómenos que lo componen. Las distintas implementaciones SIG han ido adoptando diferentes modelos de representación, lo que hace difícil la integración tanto de datos como de servicios. Con este marco en los '90 nacen dos organizaciones que se ocupan de la estandarización de IG: ISO/TC 211 [ISO94] y el OpenGIS Consortium (OGC) [OGC94] (más adelante denominado Open Geospatial Consortium).

OGC e ISO/TC 211 conforman una propuesta de trabajo coordinado que alienta el desarrollo cooperativo y que apunta a minimizar la superposición de tareas. OGC envía sus especificaciones a la ISO para que terminen siendo aprobadas vía ISO/TC 211. El consorcio industrial OGC tiene su propio programa de conformidad y testing para las especificaciones que desarrolla y un programa de interoperabilidad para el desarrollo rápido de prototipos de sus especificaciones. En general, OGC es responsable de las especificaciones de interfaces, mientras que ISO/TC 211 desarrolla estándares para modelos de datos geográficos, transferencias y servicios. Algunos resultados de este proceso incluyen: la especificación de la *Interface OGC Web Map Service* [wms04, ISO05e] que ha sido implementada por más de 100 proveedores SIG, bajo la conformidad cooperativa OGC e

ISO 19128. Otra especificación OGC incluida en ISO/TC 211 es el Lenguaje de Marcas Geográfico (GML)¹. En lo que sigue de este capítulo se recopilan y analizan estos y otros resultados del proceso de estandarización.

5.1. Conceptos generales

Originalmente, el contar con IG en diferentes formatos propietarios restringió la posibilidad de compartir información entre aplicaciones SIG que no compartieran la misma plataforma de base. La gran diversidad y esfuerzos duplicados junto con la necesidad de propiciar interoperabilidad e integración entre sistemas motiva a diversas organizaciones a trabajar en normativas estándares para información y servicios geográficos digitales. A continuación se definen los conceptos norma, estándar, recomendación y especificación, utilizados en el ámbito del vocabulario técnico de normalización:

Norma es todo documento que armoniza aspectos técnicos de un producto, servicio o componente, definido como tal por algún organismo oficial de normalización, como son ISO, CEN o AENOR. También se llaman *normas de jure* o *normas de derecho*.

Estándar es un documento o práctica que, sin ser norma, está consagrado y aceptado por el uso. También se conocen como *normas de facto* o *normas de hecho*. En el ámbito de SIG las especificaciones de Open Geospatial Consortium y los formatos DGN serían estándares. Un estándar es el estadio anterior a una norma regulada.

Recomendación es una directriz que promueve un organismo que busca definir prácticas y usos para una comunidad determinada. El éxito depende de la influencia que es capaz de ejercer el organismo que la propone. Por ejemplo: EUROSTAT produce recomendaciones para armonizar las prácticas estadísticas en Europa; el Consejo Superior Geográfico define recomendaciones acerca de la cartografía en España, entre otros. La diferencia con *estándar* está en que este último está “aceptado por el uso” mientras que una *recomendación* busca promover el uso.

Especificación, es una descripción técnica, detallada y exhaustiva de un producto o servicio, con la información necesaria para su producción.

En las siguientes secciones se recopilan los principales resultados alcanzados en el proceso de regulación en representación y servicios geográficos por los organismos internacio-

¹En Inglés *Geography Markup Language* (GML).

nales ISO/TC 211 y el consocio OGC. Ambas agrupaciones trabajan colaborativamente, la primera con facultad de definir normas y la segunda a través del consenso de la industria, agencias gubernamentales y universidades, definiendo especificaciones y estándares aprobados por el uso.

5.2. Organismos internacionales de estandarización

El desarrollo y uso de aplicaciones SIG data de los '70. En sus orígenes, cada SIG diseñó sus modelos de datos, formatos de almacenamiento y operaciones de análisis, buscando satisfacer sus requerimientos particulares. Los primeros esfuerzos para resolver las dificultades de interoperabilidad entre sistemas los realizó un grupo heterogéneo integrado por representantes del gobierno, de las ciencias de investigación, y socios de la industria que fundaron en 1994 el OpenGIS Consortium con el objetivo de definir y divulgar estándares para representar y manipular IG y favorecer la interoperabilidad entre herramientas SIG. OGC crea dos tipos de productos: *especificaciones abstractas* y *especificaciones sobre implementación*. El propósito de las especificaciones abstractas es crear y documentar el modelo conceptual que sirva de base a las especificaciones de implementación. Las especificaciones de implementación [OGC06, OGC07] definen e implementan interfaces estándares.

Contemporáneamente, ISO a través del Comité Técnico ISO 211 (ISO/TC 211) también comienza a ocuparse del desarrollo de normas y estándares de carácter internacional aplicables al entorno IG. El propósito de ambas organizaciones es crear especificaciones para que las tecnologías SIG resulten interoperables. Las especificaciones son desarrolladas, para luego ser finalmente adoptadas siguiendo un proceso de consenso entre los miembros. El objetivo es amplio, por lo que ambas organizaciones han distribuido el trabajo de especificación en unidades más pequeñas abordadas en paralelo. Estas unidades reciben el nombre de *tópicos abstractos*² por OGC [OGC05, Pau05, OGC99, Cli09] y de *estándar internacional* por ISO [ISO03c, ISO05c, ISO05e]. Actualmente, ambas organizaciones trabajan colaborativamente.

²En Inglés *abstract specification topics*.

5.2.1. Open Geospatial Consortium (OGC)

El proyecto OpenGIS precedió al lanzamiento formal del actual Open Geospatial Consortium (OGC) en septiembre de 1994. OGC es un consorcio internacional que comprende a más de 360 compañías, agencias gubernamentales y universidades, participando de un proceso consensuado para el desarrollo de especificaciones de interfaces de código abierto aplicables a sistemas de geoprocésamiento. Los resultados de OGC se denominan *Especificaciones OpenGIS* y proveen soluciones interoperativas para hacer “geo-disponible” la información de la Web y los servicios wireless de ubicación. La misión del OGC [OGC05] es conducir hacia desarrollos que promuevan el uso de arquitecturas que permitan la integración de aplicaciones con servicios y datos geoespaciales.

OGC desarrolla, distribuye y promueve el uso de estándares abiertos para procesamiento espacial desde sus programas “Especificación”, “Interoperabilidad”, “Evaluación” y “Difusión y Adopción”. El Programa de Especificación trabaja en el proceso de consenso formal para llegar a especificaciones OpenGIS aprobadas o *adoptadas*. El Programa de Interoperabilidad promueve iniciativas de ingeniería práctica para acelerar los desarrollos y aceptación de las especificaciones OpenGIS. El Consejo de Evaluación crea y mantiene documentación de los puntos de referencia básicos, procedimentales y técnicos. La Junta de Evaluación examina los documentos antes de ser publicados. El Programa de Difusión y Adopción distribuye y publicita especificaciones e interfaces espaciales de código abierto para su uso globalizado. El objetivo es lograr tecnologías de geoprocésamiento interoperativas, según el esquema “plug and play”. De hecho, se asocia la marca registrada OpenGIS a aquellos productos que cumplen con sus especificaciones.

Los miembros de OGC además participan en ISO/TC 211 a través del Consejo Consultivo Conjunto ISO/TC211-OGC. La misión es coordinar los esfuerzos de ambos organismos y asegurar una única norma de referencia. A pesar de tener orígenes distintos, OGC reemplazó tópicos relacionados con la representación de datos por los estándares definidos por ISO/TC 211.

Especificaciones OpenGIS

Las *Especificación OpenGIS* son documentos técnicos que definen las interfaces para productos y servicios geoespaciales. Las especificaciones OpenGIS son el “producto” o resultado generado por OGC y se desarrollan respetando el desafío de interoperabilidad. Se

asegura que las aplicaciones desarrolladas bajo estas especificaciones garanticen sistemas o componentes interoperables.

Las especificaciones OpenGIS se clasifican según su propósito en dos grupos: los modelos abstractos y las especificaciones de implementación. Las primeras son conceptuales mientras que las especificaciones de implementación son técnicas y proveen detalles de implementación. El Comité Técnico desarrolló un modelo de arquitectura denominada *Especificación Abstracta OpenGIS* [OGC05] según su visión de la tecnología geoespacial y las necesidades de interoperabilidad. Los componentes de este modelo proveen el marco conceptual para los distintos desarrollos. Los protocolos e interfaces de código abierto se construyen respetando este modelo abstracto para garantizar interoperabilidad entre distintas marcas y tipos de sistemas de procesamiento espacial. El cuadro 5.1 muestra los dominios de modelos abstractos ordenados por tópicos tal como los define OGC. Las *especificaciones OpenGIS* son especificaciones técnicas que detallan las interfaces de los componentes. Una especificación de interface se considera implementada, cuando habiendo sido desarrollada independientemente por al menos dos ingenieros de software, las componentes resultantes pueden interoperar.

Otros documentos incluyen el *Modelo de Referencia OpenGIS (ORM)* [OGC08] que da el marco a las tareas continuas del OGC, el documento *Prácticas Recomendadas* con las posiciones oficiales de los miembros OGC entre otros. Todos los productos OGC son de dominio público, resultan del trabajo colaborativo y consensuado de sus miembros, y están disponibles en el sitio oficial OGC³.

5.2.2. Organización Internacional para la Estandarización ISO/TC 211

Paralelamente en 1994 la Organización Internacional de Estándares (ISO) crea el comité Técnico para normalización en Geomática e Información Geográfica ISO/TC 211. Al comité se le asignó la responsabilidad de definir normativas de referencia para IG digital y transferencia de datos y servicios, requeridos especialmente por la industria IDE o SIG distribuidos. Así surge la familia ISO 19100, un conjunto de normas relacionadas con *fenómenos geográficos*. Estas normativas tratan sobre los métodos, herramientas y servicios para la gestión de datos, adquisición, procesamiento, análisis, acceso, presentación y

³ <http://www.opengeospatial.org/>

Tópico 0	Revisión.
Tópico 1	Geometría de Fenómenos.
Tópico 2	Referencia Espacial por Coordenadas.
Tópico 3	Estructuras Geométricas de Ubicación.
Tópico 4	Funciones Almacenadas e Interpolación.
Tópico 5	Fenómenos.
Tópico 6	El Tipo Cobertura.
Tópico 7	Imágenes de la Tierra.
Tópico 8	Relaciones entre Fenómenos.
Tópico 10	Colecciones de Fenómenos.
Tópico 11	Metadatos.
Tópico 12	La Arquitectura de Servicios OpenGIS.
Tópico 13	Catálogo de Servicios.
Tópico 14	Semánticas y Comunidades de Información.
Tópico 15	Servicios de Explotación de Imágenes.
Tópico 16	Servicios de Transformación de Coordenadas de Imágenes.
Tópico 17	Servicios de Ubicación de Base de Móviles.
Tópico 18	<i>Geospatial Digital Rights Management Reference Model (GeoDRM RM).</i>
Tópico Domain 1	Dominio de Telecomunicaciones.

Cuadro 5.1: Modelos Abstractos OGC

transferencia de IG en formato digital entre diferentes usuarios, sistemas y localizaciones. ISO/TC 211 comenzó con un programa de trabajo que comprendió el desarrollo paralelo de varios estándares relacionados con IG para avanzar rápidamente en interoperabilidad.

La familia ISO 19100 se basó en el paradigma orientado a objetos y mantiene relación con otras normas y estándares internacionales como Lenguaje Unificado de Modelado UML, ISO 9000 sobre gestión de la calidad, ISO 8601 de formatos de fechas y horas, entre otras. Comprende alrededor de cincuenta normas organizadas según la siguiente clasificación [Pas08]:

- **Normas de propósito general:** ISO 19101: Modelo de Referencia [ISO02c], ISO

19103: Lenguaje de Modelado Conceptual, ISO 19104: Terminología, ISO 19105: Conformidad y Pruebas, ISO 19106: Perfiles, ISO 19107: Modelo Espacial [ISO03c], ISO 19108: Modelo Temporal [ISO02d], ISO 19111: Referencia Espacial por Coordenadas [ISO07e], ISO 19112: Referencia Espacial por Identificadores Geográficos [ISO03b], ISO 19115: Metadatos [ISO], ISO 19115-2: Metadatos para Imágenes y Mallas, ISO 19139: Metadatos. Esquema de Implementación XML [ISO07d], ISO 19137: Perfiles más usados del Modelo Espacial [ISO02a].

- **Normas sobre calidad:** ISO 19113: Principios de Calidad [ISO02b], ISO 19114: Procedimientos para Evaluación de Calidad [ISO03a], ISO 19138: Medidas de Calidad [ISO06], ISO 2859 e ISO 3951: Procedimientos de Muestreo para la inspección de atributos y variables, ISO 9000: Gestión de Calidad.
- **Normas de servicios geográficos:** ISO 19119: Servicios, ISO 19128: Interface de Servidor Web de Mapas [ISO05e], ISO 19133: Servicios de Rastreo y Navegación Basados en Localización [ISO05a], ISO 19134: Servicios de Enrutamiento y Navegación Basados en Localización de Modo Múltiple [ISO07c].
- **Normas sobre aplicaciones, formato y representación:** ISO 19109: Reglas para Esquemas de Aplicación [ISO05c], ISO 19110: Metodología para la Creación de Catálogos de Elementos [ISO05d], ISO 19117: Representación [ISO05b], ISO 19131: Especificación de Productos de Datos [ISO07a], ISO 19136: Geographic Markup Lenguaje (GML)[ISO07b].
- **Normas relacionadas con representaciones raster y malla:** ISO 19101-2, ISO 19121, ISO 19123, ISO 19124, ISO 19129, ISO 19130.

En la próxima sección se analizan en detalle los estándares y especificaciones principales que tienen relación con el diseño del modelo de datos de una aplicación SIG.

5.3. Estándares para información geográfica

En cualquier sistema o aplicación que incluya “gestión de datos” generalmente se construyen dos modelos de datos: *el modelo conceptual* y *el modelo lógico*. El modelo conceptual sirve a los efectos de clasificar, identificar y representar fenómenos del mundo real

que se está modelando. Se utiliza como herramienta de comunicación entre el equipo de desarrollo y los clientes o usuarios del sistema. El modelo conceptual utiliza abstracciones de alto nivel que no son directamente representables en la máquina, por ejemplo, con IG el modelo conceptual se abstrae de cómo representar conjuntos infinitos de puntos. El modelo lógico o modelo discreto se ocupa de la representación adecuada en la computadora. La figura 5.1 muestra la relación entre los elementos del mundo real y sus abstracciones en un sistema de información.

La construcción del modelo de datos de un problema sigue un proceso iterativo que pasa por los estadios: identificar los fenómenos del mundo a modelar, construir el esquema conceptual, y traducirlo al esquema lógico. El esquema conceptual de aplicación surge de un modelo conceptual representado usando las abstracciones de algún lenguaje de representación conceptual. En la sección 2.2 del capítulo 2 de esta tesis se presentaron las abstracciones requeridas para la construcción de esquemas conceptuales en problemas que incluyan IG y ejemplos de lenguajes de modelado existentes. Una vez que el esquema conceptual se discute con los usuarios y se refina hasta llegar al diseño conceptual adecuado para el problema, el mismo se traduce a un esquema lógico o discreto. La traducción consiste en encontrar para cada elemento del esquema conceptual su correspondiente abstracción en el modelo discreto según las componentes de un lenguaje de modelado lógico. Sobre abstracciones y lenguajes de nivel lógico para representación de IG se discutió en la sección 2.3 del mismo capítulo.

En esta sección se presentan los resultados más relevantes para esta tesis relacionados con estándares para la representación de IG a nivel de modelos conceptuales y de modelos lógicos.

5.3.1. ISO/TC211 19101 Modelo de Referencia

El estándar internacional *ISO 19101: Modelo de Referencia* [ISO02c] provee el marco que contiene el alcance de la serie de estándares ISO 19XXX sobre modelos geoespaciales. En este estándar se introduce el "Modelo de Referencia de Dominio", que provee la arquitectura de información para IDE's. En ISO 19101 están las bases para la estandarización que luego son abordadas por otra serie de estándares. El núcleo del modelo está en la base de información con datos geoespaciales o "*dataset*". Un dataset contiene las instancias de "Fenómenos Geográficos" que son descritos por sus "metadatos". Estos metadatos

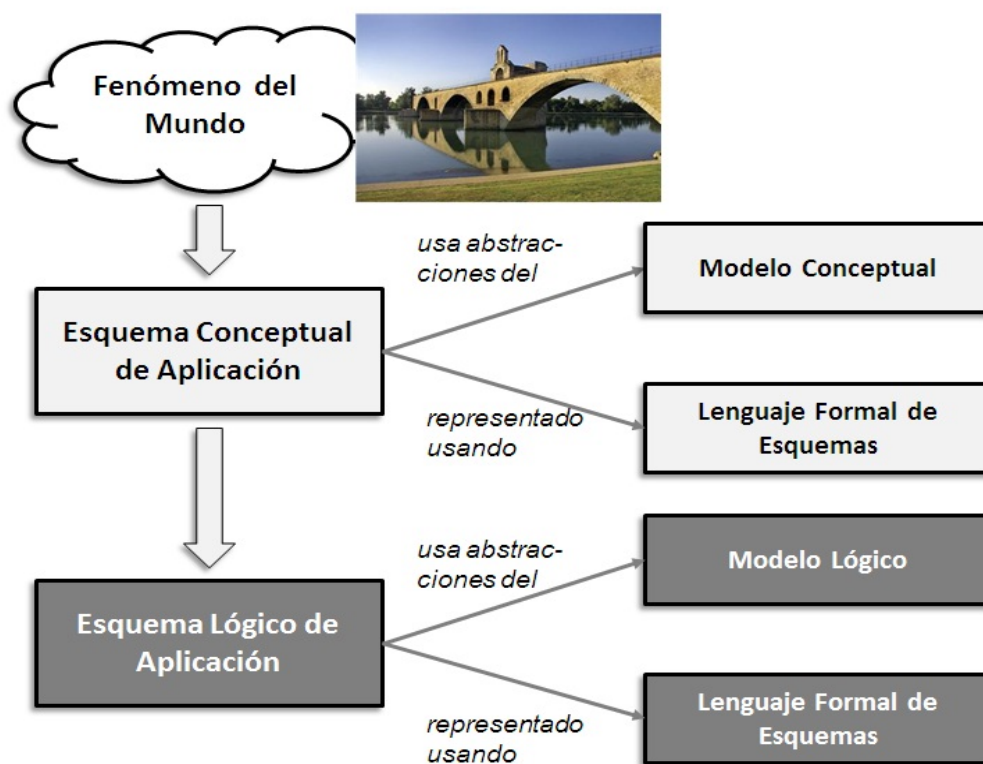


Figura 5.1: El proceso de modelado

especifican la estructura lógica y semántica a través de su “esquema de aplicación”. Los esquemas de aplicación especifican tipos de fenómenos con sus restricciones, relaciones y metadatos. Un esquema de aplicación puede definir los “tipos de fenómenos” propios o referir a “Catálogos de Fenómenos” externos.

Fenómeno Geográfico

El término **fenómeno** se utiliza para nombrar una abstracción del mundo real, por ejemplo, un río, una parcela, el mapa de un continente, etc. Los fenómenos se integran en el “Modelo de Fenómeno General” (GFM)⁴. La Figura 5.2(a) muestra las etapas del proceso de modelado adaptadas a la construcción de modelos de datos para sistemas geográficos. La Figura 5.2(b) muestra cómo se realiza el proceso de transformación desde los objetos del mundo real al conjunto de datos geográficos que los modela usando GFM. ISO 19109

⁴En Inglés *General Feature Model*(GFM).

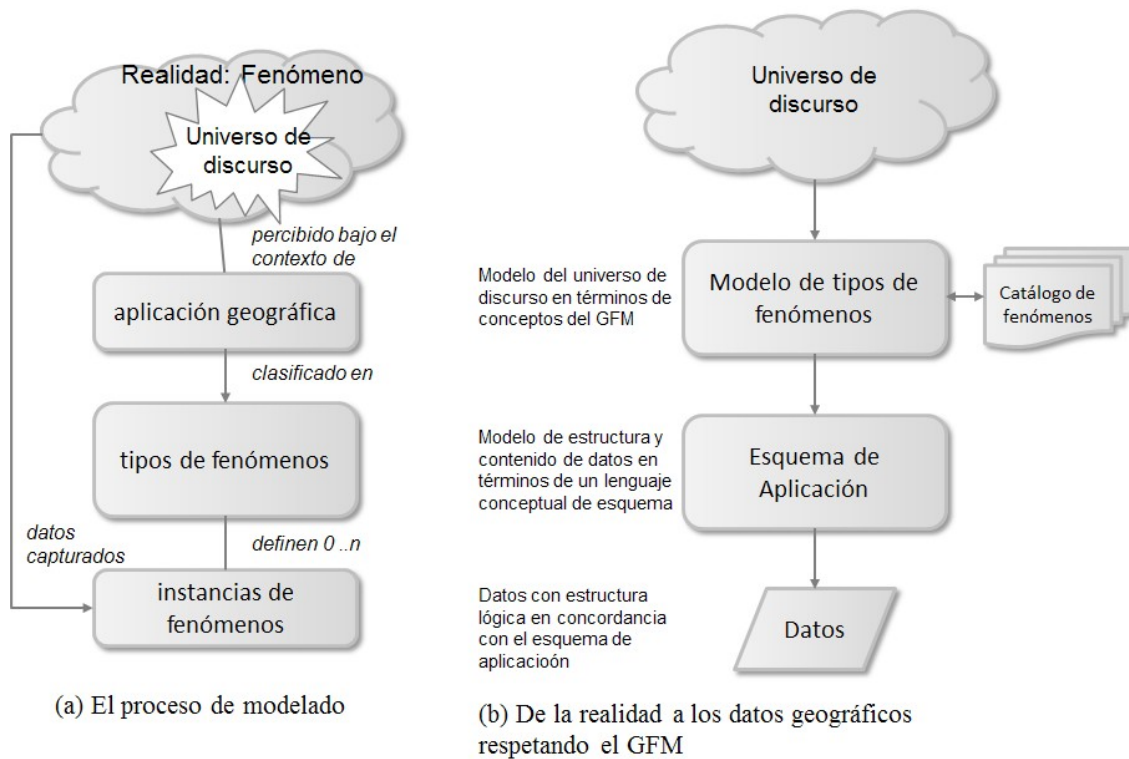


Figura 5.2: Proceso de Transformación: de la realidad a los datos

proporciona un modelo abstracto o metamodelo, el cual clasifica una vista del mundo real en el GFM, un metamodelo para la representación de IG.

El GFM define un fenómeno abstracto con atributos y operaciones. Los atributos mantienen la información estática de un fenómeno como pueden ser: la calidad del fenómeno o sus propiedades geométricas. Las operaciones contienen información sobre los cambios del fenómeno a causa de influencias externas como puede ser el mostrar una carretera para unos intervalos de escalas determinados. Existen otros conceptos adicionales que también puede tener asociado un fenómeno: asociaciones con otros fenómenos, relaciones de generalización y especialización, y restricciones. El siguiente ejemplo explica la interpretación de las figuras 5.2(a) y (b) en un problema concreto:

Ejemplo 5.3.1 *Asumamos que el **universo de discurso** o problema de interés es “Ríos de Argentina”. Desde el punto de vista del proceso de modelado se identifica el **tipo de fenómeno** FT_RÍO. Como instancias de FT_RÍO figurarán Paraná, Grande, Bermejo, Negro, entre otros. Desde el punto de vista del modelo de datos, para los fenómenos de*

FT_RÍO se identifican los atributos **nombre**, bajo el dominio cadena de caracteres; **longitud en km** y **caudal** ambos definidos en el dominio de los Reales y el atributo espacial **curso** para el que se seleccionó la representación geométrica curva. Esta información del **Modelo de Tipos de Fenómenos** forma parte del **Catálogo de Fenómenos**. El siguiente paso traduce el modelo conceptual a un **Esquema de Aplicación** siguiendo un modelo espacial, por ejemplo, al modelo vector. Los distintos ríos con sus valores de atributos conforman los **Datos** del sistema.

5.3.2. ISO/TC211 19109 Modelo conceptual estándar para Esquemas de Aplicación

El estándar internacional **ISO 19109: Reglas para Esquemas de Aplicación** [ISO05c] fija las reglas para crear y documentar Esquemas Conceptuales de Aplicación, incluyendo los principios para la definición de fenómenos geográficos. Con la definición del GFM se proporciona la descripción formal para el esquema de datos, esto es, la descripción general de los elementos tipo de fenómeno y sus atributos. Las definiciones de ISO 19109 han sido adoptadas como Modelo Conceptual de Aplicación estándar por OGC.

GML, especificado por ISO 19109, define un metamodelo para tipos de objetos espaciales, sus tipos de propiedades (atributos, asociaciones, roles y operaciones) y las restricciones. También sirve como metamodelo para el catálogo de fenómenos, dando la estructura para representar la semántica de la IG.

La Figura 5.3 muestra una vista parcial de GFM, expresada utilizando notación UML. Se observa que *GF_FenomenoGeografico* o fenómeno, es la unidad para IG. Está definida como una “metaclase” que se instancia con las clases que representan fenómenos geográficos individuales. Las instancias de una clase que representa un fenómeno geográfico individual se las denomina instancias de fenómeno.

Ejemplo 5.3.2 *Continuando con el ejemplo 5.3.1, supongamos la siguiente simplificación para del Esquema Conceptual de Aplicación que se muestra en la figura 5.4. En el modelo se puede observar que se identificaron a los tipos de fenómenos RIO y TRAMO, bajo la asociación de composición por la que una instancia de RIO resulta de la composición de varias instancias de TRAMO. En el modelo además se puede observar que el tipo de fenómeno RIO está descrito por los atributos nombre, longitud, caudal y curso (atributo*

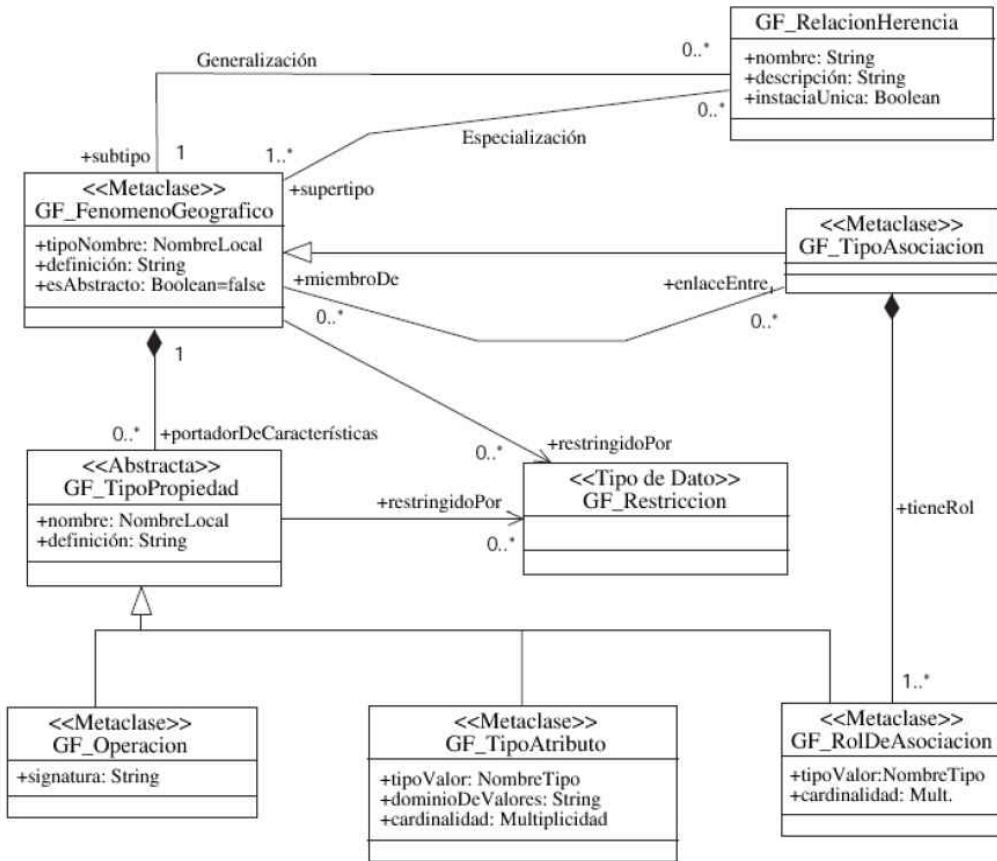


Figura 5.3: Fragmento del Modelo de Fenómeno General

espacial) y el tipo de fenómeno TRAMO cuenta con el atributo descriptivo número y los atributos espaciales punto de inicio, punto de finalización y segmento. El modelo de objetos UML que resulta de instanciar el Esquema Conceptual de Aplicación de la figura 5.4 al estándar GFM (figura 5.3) se muestra en la figura 5.5. El patrón de colores es al único efecto de visualizar las instancias que corresponden a clases de GFM distintas.

A continuación se describe en detalle los componentes de GFM de la figura 5.3.

- *TipoPropiedad* (GF_TipoPropiedad): es la metaclase para cualquier tipo de propiedad de un fenómeno geográfico. Son tipos de propiedad: atributos, comportamiento, y/o roles de asociación. Cada tipo de propiedad posee el nombre, la definición (atributo, comportamiento o rol) y una asociación con un fenómeno geográfico para indicar con cual de ellas está ligada. Como tipo de propiedad se consideran:

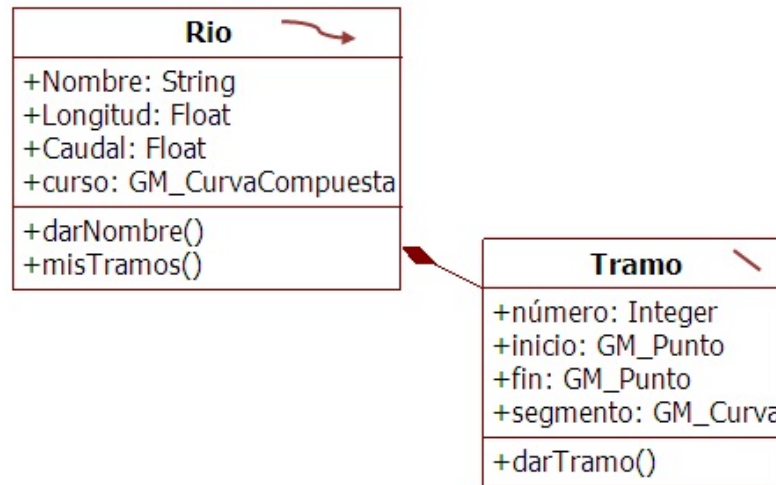


Figura 5.4: Fragmento del Esquema Conceptual de Aplicación “Ríos de Argentina”

- *TipoAtributo* (GF_TipoAtributo): es la metaclassa para la definición de tipos de atributos. Posee los descriptores tipo de dato, nombre, dominio de valores y cardinalidad. Los tipos de atributos están organizados en *tipos de atributos espaciales*, que asumen valores del dominio espacial geométrico o topológico, *tipos de atributos temporales*, cuyos valores pertenecen al dominio temporal, *tipo de atributos de ubicación*, cuyos valores son identificadores a un índice geográfico, atributos de *tipo metadada* y *atributos temáticos* o descriptivos (Figura 5.6). La jerarquía de atributos espaciales referencia a la norma ISO 19107. En este estándar también se define una jerarquía de tipos de asociaciones posibles, las que incluyen agregaciones, asociaciones espaciales o asociaciones temporales (Figura 5.7).

En la figura 5.5-GFM para el modelado “Ríos de Argentina”, sobre el fenómeno RIO se definen las instancias de GF_TipoAtributo descriptivos *nombre*, *longitud* y *caudal* y un GF_TipoAtributo espacial *curso*, en este caso de tipo GM_CompositeCurve. Por simplicidad en la figura 5.5 no se incluyó su instanciación con la jerarquía de tipos espaciales. Para el fenómeno TRAMO se define un único GF_TipoAtributo descriptivo *número* y GF_TipoAtributo espaciales *inicio* y *fin* de tipo GM_point y *segmento* de tipo GM_curva.

- *Roles de asociaciones* (GF_RolDeAsociacion): es la metaclassa para la definición

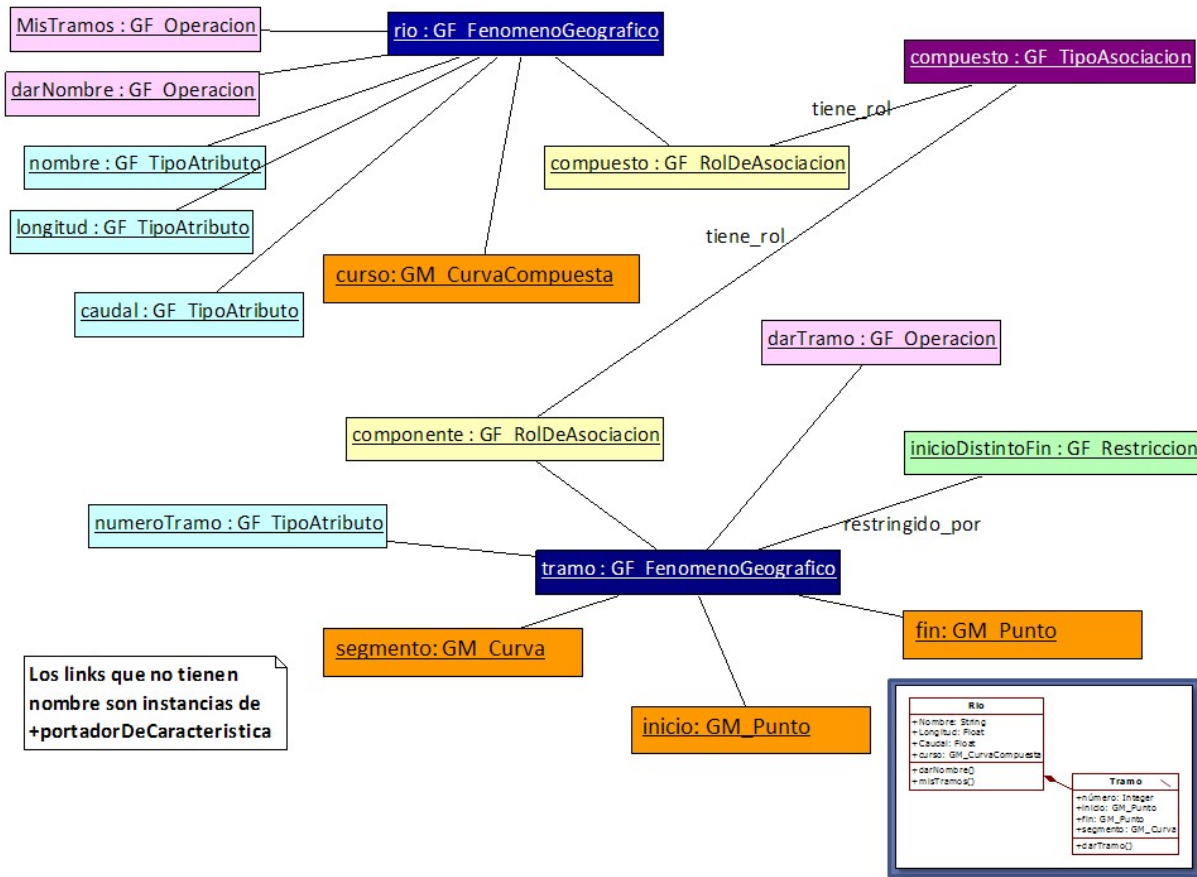


Figura 5.5: Modelo de Objetos de GFM para el modelado de “Ríos de Argentina”

de clases de roles. Indica el rol que juega un fenómeno geográfico en términos de la asociación y especifica el contexto para estos tipos y sus instancias. Posee el nombre del rol y la cardinalidad denotando el número de instancias que un fenómeno geográfico puede tener con respecto a ese rol.

El modelo GFM de referencia (figura 5.5) incluye para el fenómeno RIO por la propiedad GF_RolDeAsociacion el rol *compuesto* y el fenómeno TRAMO tiene asignado en GF_RolDeAsociacion el rol *componente*

- *Operaciones* (GF_Operacion): es la metaclass para describir el comportamiento de los fenómenos geográficos en términos de sus operaciones. Las instancias de esta clase pueden ser de tres tipos diferentes: de observación, de cambio o de construcción. Las operaciones de observación retornan los valores actuales de los atributos, las de cambio poseen acciones que cambian el valor de esos

atributos y los constructores crean nuevas instancias.

En la figura 5.5-GFM para el modelado de Ríos de Argentina, se observan las instancias de `GF_Operacion` *misTramos* y *darNombre* asociadas con el fenómeno RIO y *darTramo* asociada a TRAMO. En este punto queremos aclarar que las operaciones que se incluyeron en el modelo son a título ilustrativo del ejemplo y no representan el conjunto completo de operaciones requeridas. Por ejemplo, no se incluyó ninguna operación de construcción.

- *Tipos de asociaciones* (`GF_TipoAsociacion`): es la metaclass para describir asociaciones entre fenómenos geográficos. Puede contener atributos ya que esta clase es una especialización de la clase `GF_TipoPropiedad`. A su vez, posee una asociación con esta misma clase llamada “vinculadaCon” para denotar los fenómenos geográficos que se relacionan. También incluye una asociación de agregación llamada “Rol” entre esta clase y la clase `GF_RolDeAsociacion`. La misma indica los roles que posee la asociación (al menos uno).

En el ejemplo de la figura 5.5 existe una única instancia de `GF_TipoAsociacion` *compuesto* con la asociación de composición que existe entre RIO y TRAMO.

- *Relaciones de herencia* (`GF_RelacionHerencia`): es la clase que representa una relación genérica entre un fenómeno geográfico más general (supertipo) y uno o más fenómenos geográficos más específicos (subtipos). Cualquier miembro de un fenómeno geográfico especializado es miembro también de uno más general. Como atributos posee nombre, descripción e instancias, un valor booleano que indica si una instancia del supertipo puede o no ser una instancia de más de un subtipo. Los roles “generalización” y “especialización” denotan justamente las instancias más específicas y más generales entre uno o más fenómenos geográficos.

En este caso el ejemplo de la figura 5.5 no plantea instancias de `GF_RelacionHerencia`. Se podría incluir una instancia de `GF_RelacionHerencia` con una superclase `GF_FenomenoGeografico` *FUENTESDEAGUA*, de la cual especialice RIO.

- *Restricciones* (`GF_Restriccion`): es la clase para representar restricciones tanto de los fenómenos geográficos como de las propiedades. Posee una descripción en lenguaje natural indicando el significado de la misma y una asociación con rol “restringidoPor” `GF_TipoPropiedad` o `GF_TipoCaracteristica`.

En el ejemplo de la figura 5.5 se incluyó una instancia de `GF_Restriccion` *inicioDis-*

tintoFin vinculado con el fenómeno TRAMO para indicar que los puntos de inicio y fin de tramo deben ser distintos.

Entre las ventajas que ofrece utilizar el modelo GFM se pueden decir que modela cierta información semántica además de la representación, las instancias de fenómeno llevan asociadas el tipo de fenómeno y es posible asociar más de una geometría a un tipo de fenómeno. La ISO 19109 se complementa con la norma ISO 19110 [ISO05d]. Un esquema de aplicación se crea para proceder a la definición de los fenómenos. Todos los detalles de cada uno de estos fenómenos, una vez definidos, se muestran mediante un catálogo de fenómenos. La norma Internacional ISO 19110 describe la metodología para construir estos catálogos.

5.3.3. ISO/TC211 19107 Modelo conceptual estándar para Fenómenos Geográficos

La norma *ISO 19107: El Modelo Espacial* proporciona los elementos conceptuales para describir y manipular las características espaciales asociadas a fenómenos geográficos. Dichas características están representadas por uno o más atributos espaciales, los que pueden ser *cuantitativos* o *cualitativos* y cuyo valor se vincula con un objeto geométrico (GM_Object) o un objeto topológico (TP_Object).

- La *geometría* proporciona la descripción geográfica *cuantitativa* de las características espaciales, por medio de coordenadas y funciones matemáticas. En ella se incluyen dimensiones, posición, tamaño, forma y orientación. Las funciones matemáticas que se usan para describir la geometría de un objeto dependen del tipo de sistema de coordenadas de referencia usado para definir la posición espacial. La geometría es el único aspecto de la IG que cambia cuando la información se transforma de un sistema geodésico de referencia o de coordenadas a otro. Por ejemplo, “la parcela número 80150 tiene la esquina inferior izquierda en las coordenadas UTM (450228; 4499053), es rectangular y la orientación del lado mayor es noroeste, su superficie es de 5 ha y sus lados miden 500 m x 100 m”. Ante una deformación del espacio, estas características cambian.

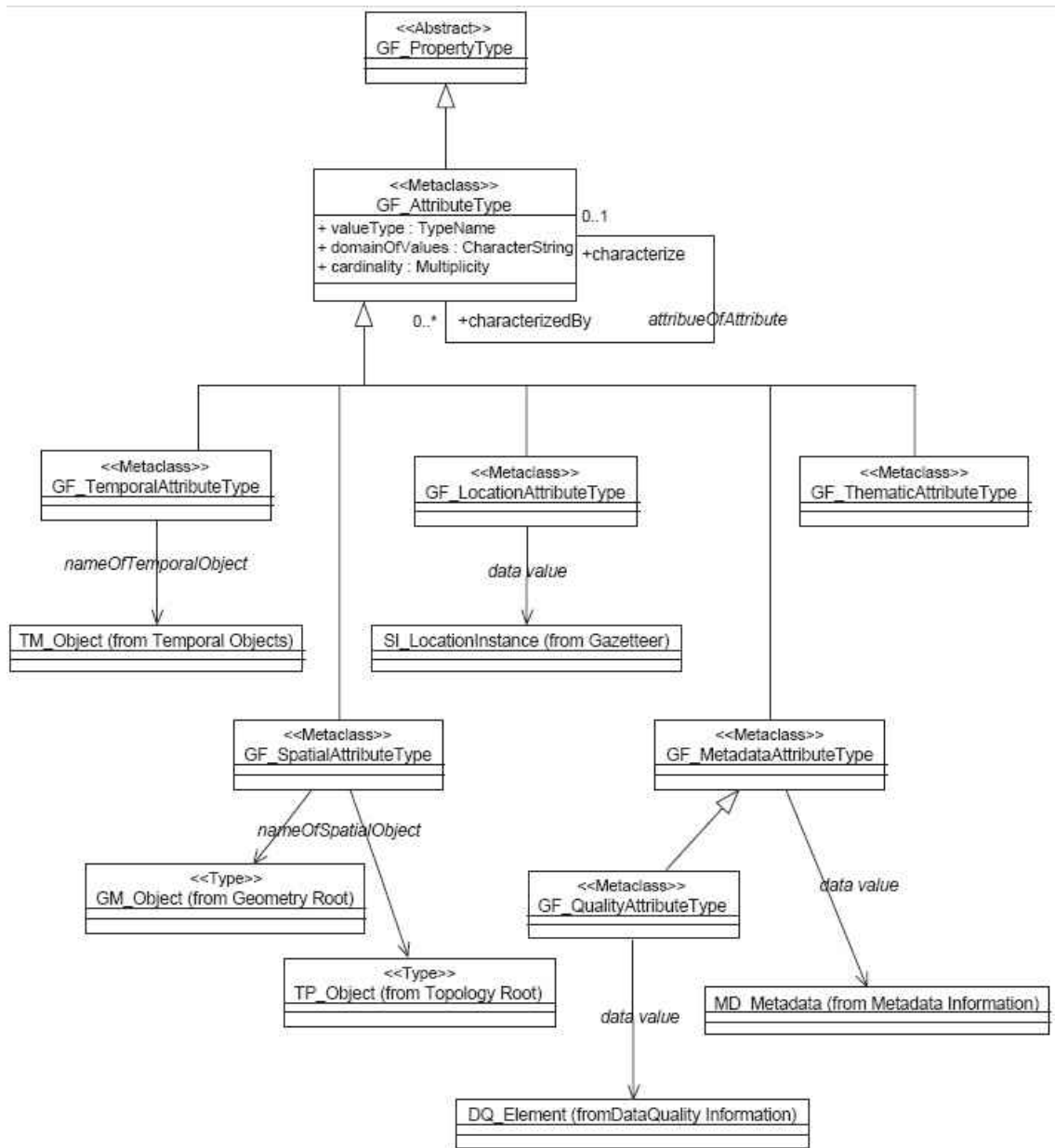


Figura 5.6: Jerarquía de Tipo de Atributo

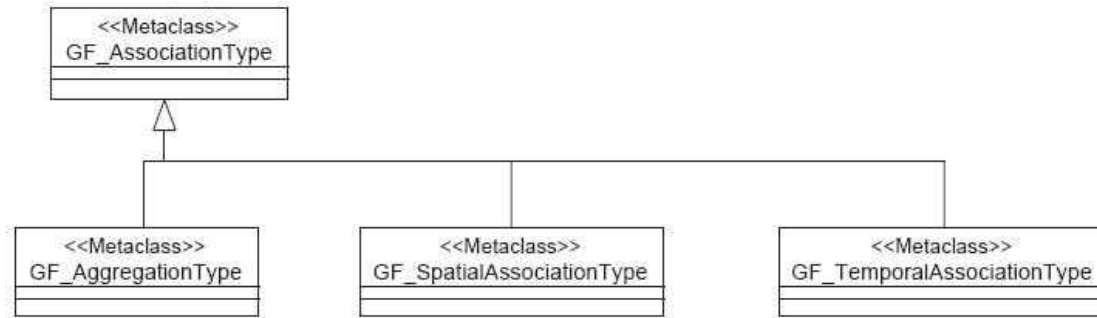


Figura 5.7: Jerarquía de Tipo de Asociación

- La *topología* proporciona una descripción geográfica *cuantitativa*, y se encarga de las características de las figuras geométricas que permanecen invariantes frente a deformaciones elásticas y continuas del espacio, como por ejemplo las transformaciones de un sistema de coordenadas a otro. Dentro del contexto de la IG, la topología se suele usar para describir la conectividad, propiedad que es invariante bajo cualquier transformación continua y que se deriva de la propia geometría. Por ejemplo, “la parcela número 80150 contiene un pozo, se encuentra incluida dentro del polígono catastral 80 y es adyacente a las parcelas 80151, 80152 y 80149”. Aún ante una deformación elástica y continua del espacio, estas características espaciales permanecen invariantes.

Las definiciones de la norma ISO 19107 están organizadas siguiendo la distribución en paquetes de UML. Se distinguen paquetes geométricos y topológicos. Las clases del modelo geométrico se desprenden de la raíz *GM_Object* y heredan una asociación opcional a un sistema de referencia de coordenadas. La figura 5.8 muestra las dependencias entre los paquetes geométricos y el listado de clases de cada uno. El nombre de las clases geométricas comienza con el prefijo *GM*.

Un objeto geométrico pertenece a uno de los siguientes tipos:

- Primitiva** (*GM_Primitive*): representa a los elementos gráficos básicos que componen al conjunto de datos geográficos. Las primitivas geométricas, según su dimensión son:
 - Punto* (*GM_Point*): primitiva geométrica cero dimensional que representa una posición.

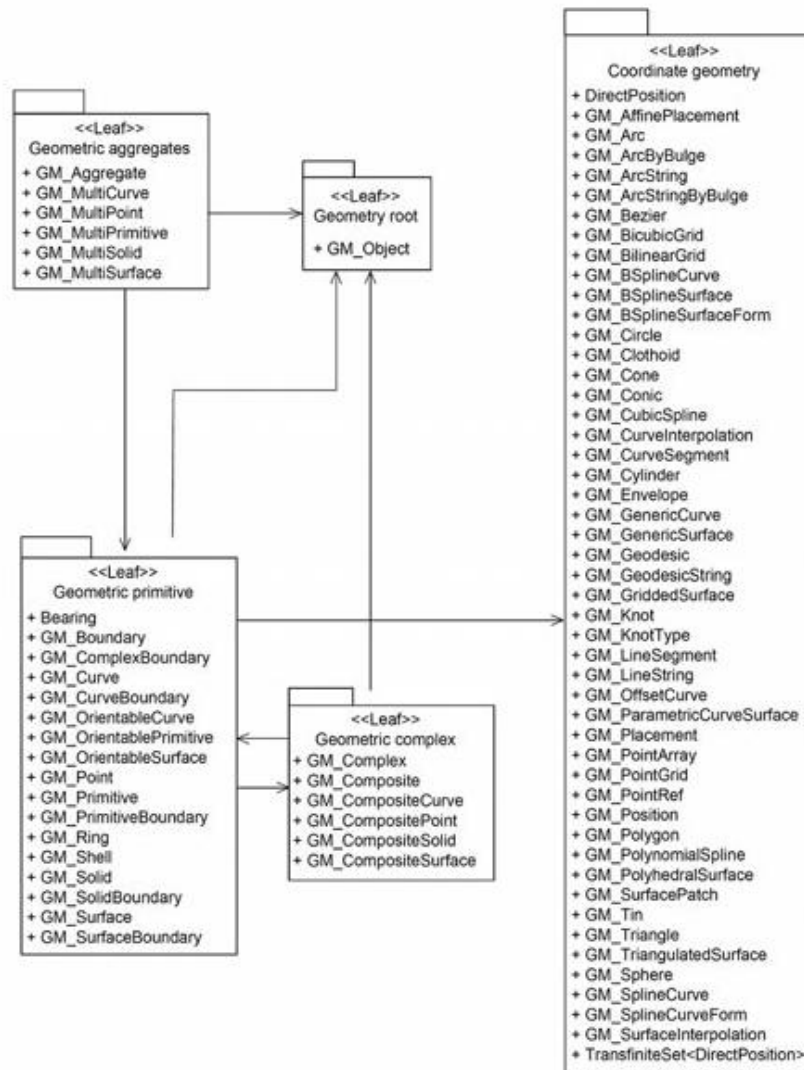


Figura 5.8: ISO 19107 - Paquetes y sus clases geométricas

- *Curva* (GM.Curve): primitiva geométrica unidimensional formada por una secuencia de posiciones; se compone de uno o más segmentos curvos (GM.CurveSegment). Diferentes métodos de interpolación entre los puntos de control permiten generar distintos tipos de segmentos curvos: clotoides, cónicas, geodésicas, arcos, etc.
- *Superficie* (GM.Surface): primitiva geométrica bidimensional que se define como superficies poligonales conectadas entre sí por las curvas fronteras de manera que se genere una superficie continua sin agujeros (superficie poliédrica),

o como una malla rectangular de puntos (GM_GriddedSurface) en el espacio unidos por filas y columnas utilizando funciones matemáticas (superficie paramétrica).

- *Sólido* (GM_Solid): primitiva tridimensional.

Las primitivas existen por sí mismas y son abiertas, esto es, no contienen a sus límites como posiciones directas, aunque pueden tenerlos como referencia.

- **Complejo** (GM_Complex): es una colección de GM_Primitives disjuntas. Si una GM_Primitive (exceptuando GM_Point) pertenece a un GM_Complex, entonces existe un conjunto de primitivas de una dimensión menor en el objeto complejo que pertenecen a la frontera de dicha primitiva. Los complejos se usan en esquemas de aplicación donde es importante compartir la geometría, como es el caso en topología digital. Un complejo se dice *compuesto* si se componen de primitivas de igual dimensión. Así se distinguen:

- *Punto compuesto* (GM_CompositePoint): objeto complejo compuesto de un único punto GM_Point. Se incluye en el modelo por completitud.
- *Curva compuesta* (GM_CompositeCurve): curva formada por un conjunto de curvas donde cada una (excepto la primera) empieza en el punto final de la anterior de la secuencia.
- *Superficie compuesta* (GM_CompositeSurface): superficie formada por un conjunto de superficies conectadas entre sí por sus fronteras que son curvas.
- *Sólido compuesto* (GM_CompositeSolid): conjunto de sólidos unidos por sus superficies límites formando un único sólido.

El uso de objetos complejos permite introducir restricciones a partir de compartir primitivas y bajo la asociación “*contiene*” y los roles de subcomplejo y supercomplejo. Por ejemplo, la restricción “para toda PROVINCIA P debe existir una REGIÓN R tal que $\text{Supercomplex}(P)=R$ ” expresa que toda provincia está contenida en una región.

- **Agregado** (GM_Aggregate): agrupa elementos geométricos sin limitaciones. El ejemplo típico es un conjunto de puntos de elevación los cuales, sin agregados, sólo podrían ser descritos como puntos individuales; sin embargo, la agregación da la

posibilidad de hacer referencia a los mismos como una unidad. Las agregaciones se denominan “Multi_primitivas” cuando son composición de primitivas de igual dimensión, y de esta forma existen: GM_Multipoint, GM_Multicurve, GM_Multisurface y GM_Multisolid.

Las clases GM_Object y GM_Primitive son abstractas, ningún objeto o estructura de datos de una aplicación las instancia directamente, sino por medio de sus subclasses no abstractas como GM_Point, GM_Curve o GM_Surface. Este no es el caso de GM_Complex o GM_Aggregate que pueden instanciarse.

La Figura 5.9 muestra las clases principales del paquete geometría. Cualquier objeto que hereda la semántica de GM_Object actúa como un conjunto de posiciones directas. Su comportamiento estará determinado por las posiciones directas que contiene. Los objetos bajo GM_Primitive son abiertos, eso es, no contienen a sus límites. Los objetos bajo GM_Complex son cerrados, esto es, contienen a sus límites. Esto conduce a cierta ambigüedad. Una representación de una línea como una primitiva debe referenciar sus puntos finales, pero no contendrá estos puntos como un conjunto de posiciones directas. Una representación de una línea como un complejo también referenciará sus puntos finales, y contendrá estos puntos como un conjunto de posiciones directas. Esto significa que representaciones digitales semejantes tendrán semántica diferente dependiendo de ser referidos como primitiva o complejo.

El estándar usa la clase GM_Position para almacenar una posición, la que puede estar instanciada con un objeto de la clase DirectPosition o indirectamente haciendo referencia a un objeto (GM_PointRef). GM_Point tiene asignada una posición directa, mientras que otros objetos como curvas, superficies y sólidos están formados por una secuencia de referencias.

Los objetos de una clase compartirán las mismas operaciones. Las que se enumeran a continuación son algunas de las operaciones definidas que se aplican a clases geométricas:

- **mbRegion**: devuelve la región que contiene al objeto geométrico en el Sistema de Coordenadas de Referencia.
- **boundary**: aplicada a un objeto, devuelve su límite que tendrá una dimensión menos que el propio objeto.

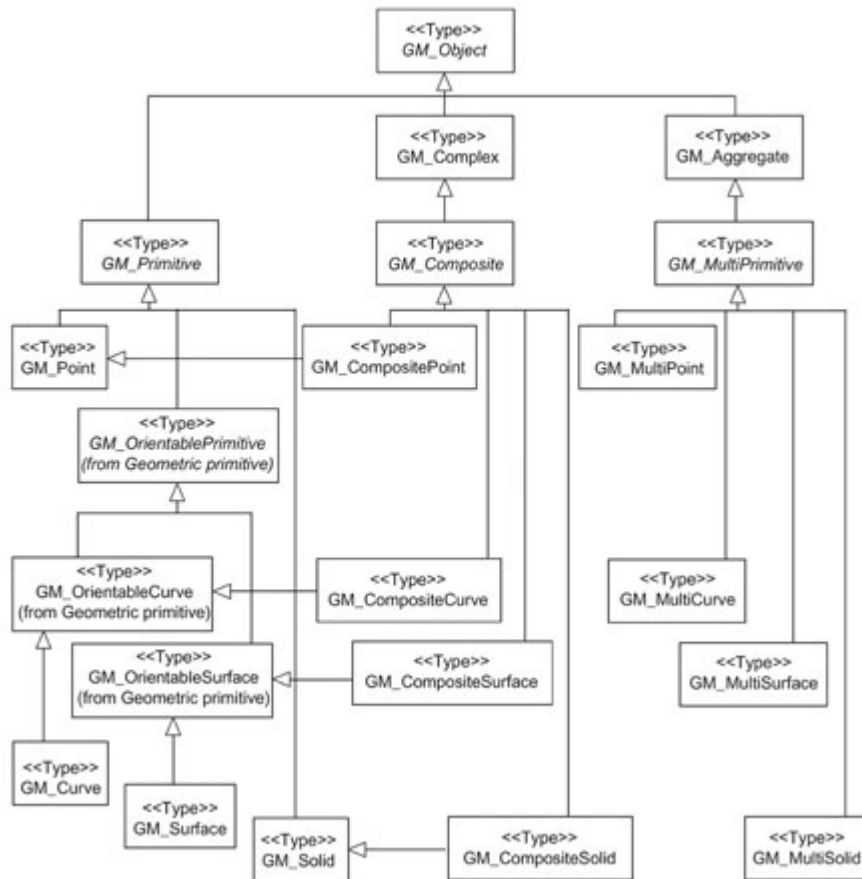


Figura 5.9: Jerarquía de clases de geométricas

- **distance:** devuelve la distancia entre dos objetos geométricos. La distancia se define como la longitud mínima entre las posibles calculadas por cada par de puntos de los dos objetos.
- **closure:** que combina un objeto con su límite. Por ejemplo, aplicado a GM_LineString la operación devolverá la línea más sus puntos inicial y final.

Las relaciones espaciales topológicas son cualitativas y se deducen de la geometría. Algunos ejemplos de las propiedades cualitativas son conectividad, inclusión, vecindad o coincidencia. La topología describe las propiedades geométricas que son invariantes a deformaciones continuas. Así un cuadrado es topológicamente equivalente a un rectángulo o un trapecioide. Un aspecto importante de la topología es que permite trabajar con

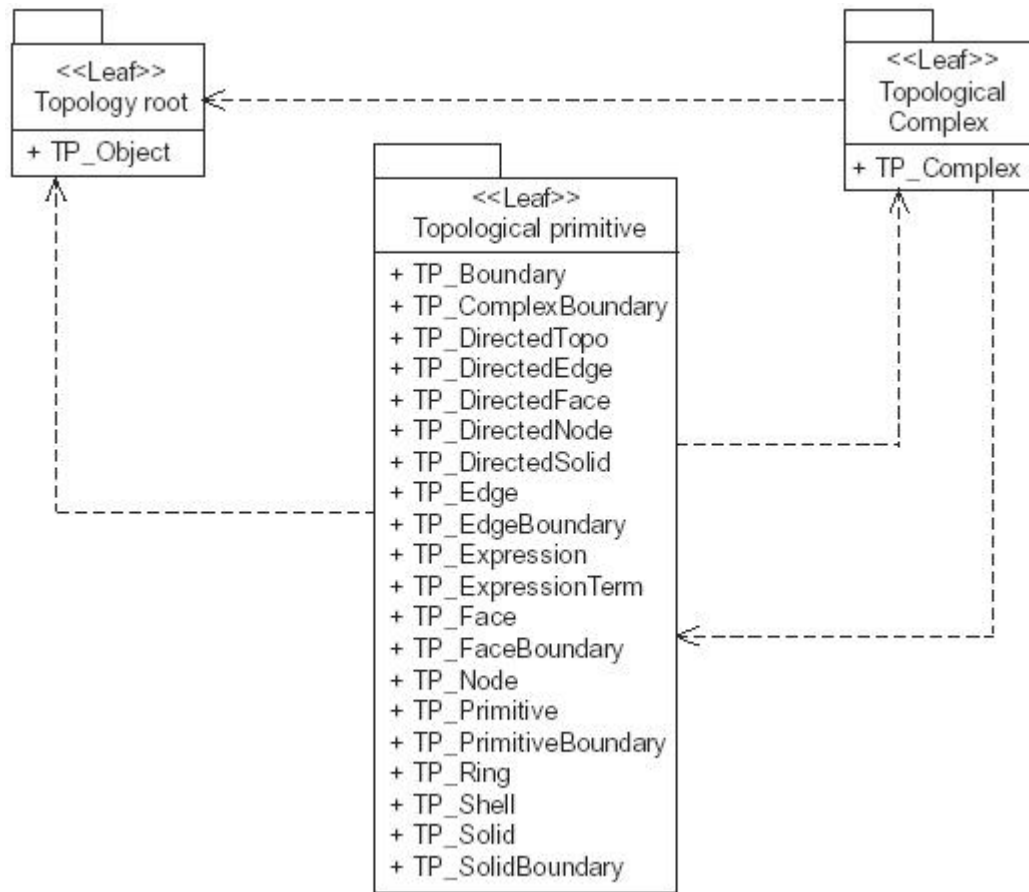


Figura 5.10: Paquetes Topológicos

algoritmos combinatorios para relacionar fenómenos geográficos y esto tiene un impacto positivo sobre la performance. La manera de lograrlo es asociando explícitamente fenómenos y objetos geométricos junto con sus relaciones geométricas implícitas. Es por esto que se definieron paquetes topológicos que son paralelos a los paquetes geométricos. La figura 5.10 muestra estos paquetes y sus dependencias.

El estándar incluye en el modelo de clases una jerarquía que tiene como raíz el objeto topológico (TP_Object) (figura 5.11) . Los objetos topológicos a su vez se clasifican en uno de dos tipos: primitivo (TP_Primitive) o complejo (TP_Complex). Las primitivas topológicas están definidas según su dimensión como:

- **Nodo** (TP_Node): primitiva topológica de dimensión cero. Los nodos representan a los puntos topológicos donde se cortan dos o más curvas. Identifican los extremos inicial y final de los arcos.

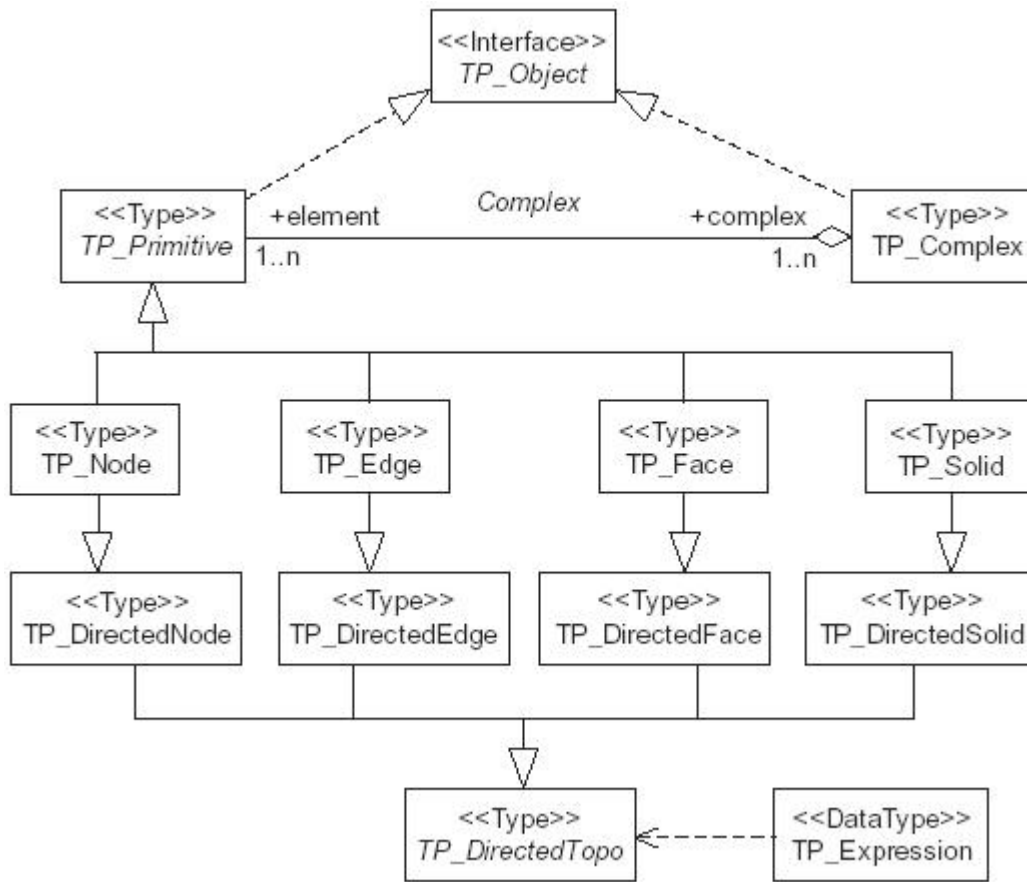


Figura 5.11: Diagrama de Clases Topológicas

- **Arco** (TP_Edge): primitiva topológica unidimensional. Representan a las curvas topológicas que están delimitadas por dos nodos y que, a su vez, delimitan caras.
- **Cara** (TP_Face): primitiva topológica 2D. Una cara es una región cerrada delimitada por arcos.
- **Sólido** (TP_Solid): primitiva topológica 3D. Son cuerpos definidos y delimitados por nodos, arcos y caras.

En la figura 5.12 se puede ver un ejemplo gráfico de las primitivas topológicas. La estructura del diagrama de clases muestra a la clase raíz TP_Objeto, de la que se desprenden TP_Primitivo y TP_Complejo. Claramente, esta organización es análoga a la que se mostró en el diagrama de clases de geometría con GM_Primitivo y el GM_Complejo. Como

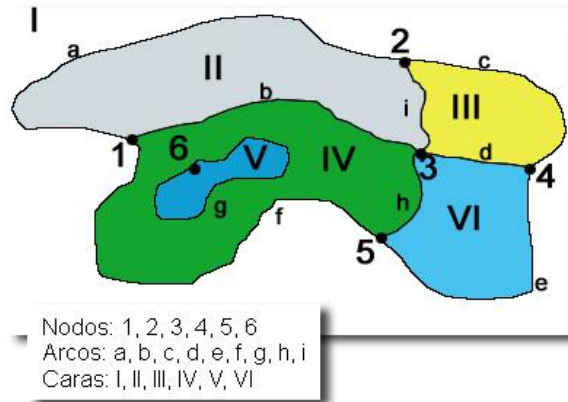


Figura 5.12: Primitivas topológicas

se muestra en la figura 5.13, hay un paralelismo entre las maneras en que los primitivos y los complejos están relacionados en los dos sistemas de clases.

En la representación de IG se distinguen dos dimensiones: espacial y temporal. El modelo espacial es más complejo y está considerado en toda su extensión en ISO 19107 [ISO03c]. Posteriormente la norma ISO 19137 definió un conjunto reducido de perfiles del esquema espacial sólo con el conjunto de elementos geométricos más utilizados.

Por su parte, el conjunto de clases estándares definidas para la dimensión temporal es más simple. En la definición de este conjunto de normas participó activamente el OGC. Incluso en algunos casos el comité ISO/TC 211 adoptó estándares ya definidos por OGC. Para manipular las características espaciales de los fenómenos geográficos se definió además un conjunto de operadores espaciales acordes con dichos modelos conceptuales.

5.3.4. Otras normas de la familia ISO 19000

Las normas ISO 19111 [ISO07e] e ISO 19112 [ISO03b] consideran el sistema de referencia de los fenómenos, ya sea por posicionamiento directo o coordenadas (ISO 19111), o por medio de identificadores geográficos (ISO 19112). ISO 19111 cubre las operaciones de conversión y transformación de coordenadas y establece los atributos para dar información precisa sobre las características y propiedades del sistema de referencia.

Relacionadas con interoperabilidad existen cuatro normas que se orientan de manera general a los servicios. Las normas ISO 19133 [ISO05a] e ISO 19134 [ISO07c] se refieren

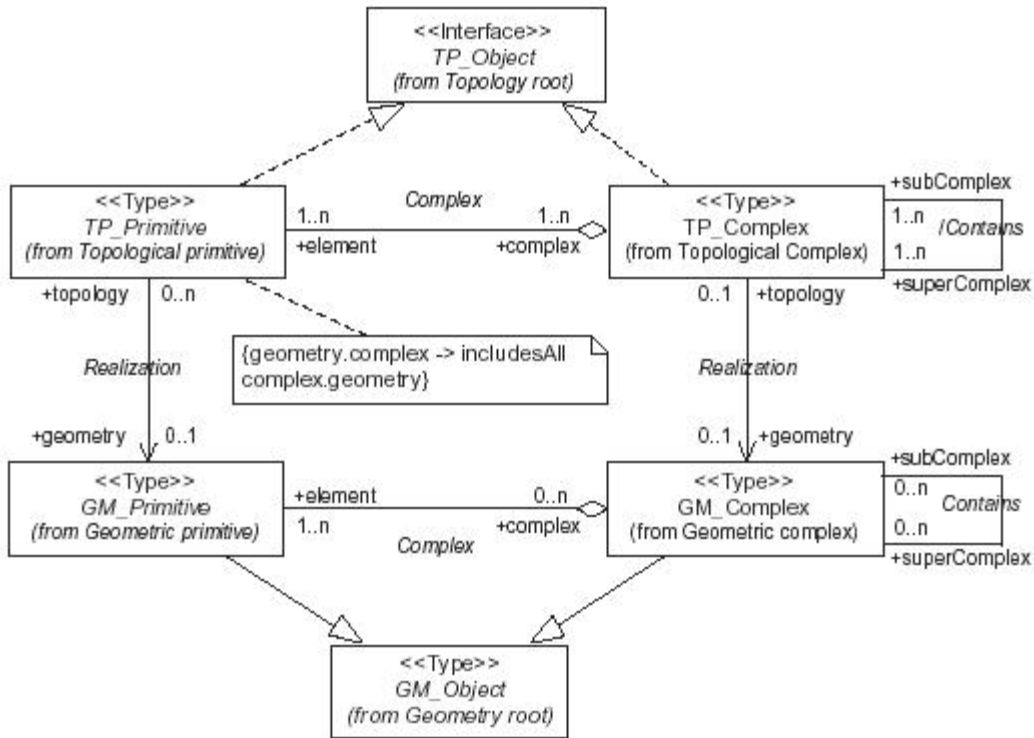


Figura 5.13: Relación entre geometría y topología

a servicios basados en la posición del objeto: seguimiento, enrutamiento y navegación en redes lineales. Las normas ISO 19119 e ISO 19128 [ISO05e] referencian a los servidores de mapas y servicios cartográficos de edición, transformación de coordenadas, rectificación, cartometría, etc.

La norma ISO 19110 [ISO05d] establece la organización para la confección de catálogos de fenómenos. Esta estructura facilita la posibilidad de comparar, conocer y explotar diferentes catálogos. Los elementos del catálogo dan una idea más concreta de los tipos de fenómenos y las propiedades que los definen.

La norma ISO 19117 [ISO05b] se ocupa de la obtención de salidas gráficas con valor cartográfico. Su objetivo es definir el esquema que se aplica a los fenómenos para obtener su representación. La especificación de representación y las reglas de representación no son parte de la base de datos y se almacenan independientemente. La especificación de representación se almacena externamente utilizando una URL como referencia universal. Por su parte, las reglas de representación se almacenan en el catálogo de representación.

Esto prevé un sistema que despega al dato de su representación y con ello se tiene una gran versatilidad.

Las normas ISO 19115 [ISO] e ISO 19131 [ISO07a] se ocupan de los metadatos que hacen posible el uso IG en la red. ISO 19115 es un modelo que ofrece versatilidad y flexibilidad por medio de la definición de perfiles, extensiones y soporte multilingüe. Constituye la referencia para trabajar en el campo de los metadatos para datos y servicios geográficos. ISO 19131 define cómo describir las especificaciones de datos geográficos. Su aplicación está condicionada por la adopción de las restantes normas de la serie y el nivel de madurez técnica de los usuarios. La mayor dificultad reside en conocer los objetivos de utilización de la IG que, naturalmente, condicionan la especificación del producto, y para los cuáles una norma con un nivel de abstracción necesariamente elevado no puede presentar indicaciones concretas.

Las normas ISO 19113 [ISO02b], 19114 [ISO03a] y 19138 [ISO06] se centran en aspectos complementarios y relativos a la calidad: identificar factores relevantes de calidad, evaluar calidad, usar medidas y métodos estándar para informar sobre la calidad. Estas normas no marcan niveles de calidad. Los niveles de calidad se deben establecer de mutuo acuerdo entre productores y usuarios en función del propósito de cada producto.

5.4. GML: El Lenguaje de Mercado Geográfico

El Lenguaje de Mercado Geográfico (GML) es un lenguaje de marcado que sirve para transportar IG a través de la red de forma que pueda ser compartida [LBTR04]. GML cuenta con capacidades para describir objetos geográficos complejos. Fue desarrollado en el marco OGC y está basado en OpenGIS Abstract Specification y la serie ISO 19100 de ISO/TC 211. Desde sus inicios ha evolucionado hasta lograr actualmente soportar un modelo de datos para IG robusto. GML ha sido aprobado por un gran número de compañías internacionales y organizaciones que definen en gran parte los lineamientos que siguen las tecnologías SIG, entre las que se encuentran Oracle, Esri, Intergraph y NTT Data.

Históricamente traducir datos geográficos de un formato a otro ha sido difícil. Como consecuencia, los usuarios con importantes bases de datos geográficas han estado restringidos a usar las herramientas de un único proveedor. El GML busca disminuir estas

dificultades y aumentar la posibilidad de compartir información geográfica. GML está basado en el Lenguaje de Marcado Extensible (XML)⁵. Las especificaciones XML proveen una forma estándar para definir lenguajes de marcado para “documentos de texto”. Los documentos XML se almacenan como texto plano lo que trae como ventajas que no se requiere de herramientas de edición sofisticadas, son fácilmente transferibles entre plataformas y no están sujetos a ningún proveedor; por esto es simple compartir información. Dado que GML está definido con XML hereda estos beneficios.

5.4.1. Características de GML

En GML las entidades geográficas se denominan *fenómenos*. Los fenómenos pueden ser concretos, como por ejemplo ríos, construcciones o calles; o pueden ser abstractos, por ejemplo, límites políticos o de distritos de salud y se describen por sus propiedades, sean estas geométricas y/o descriptivas. Los fenómenos no son representados en forma directa en GML, sino que son definidos en esquemas de aplicación según lo determine el diseñador. Esto permite que se puedan definir elementos con variadas características.

GML está basado en XML por lo que resulta muy adecuado para transferir información a través de Internet. Entre sus componentes se encuentran geometrías complejas, sistemas de referencia espaciales y temporales, tipos de proyección, topología y unidades de medida. Siguiendo a XML, GML separa contenido de representación. Así, un mismo grupo de elementos puede ser representado de múltiples formas de acuerdo a las necesidades de los clientes o usuarios de los datos, utilizando otros formatos intermedios para su visualización gráfica como son SVG o VML, mediante la aplicación de transformaciones y hojas de estilo CSS o XSL. Es importante señalar la diferencia entre datos geográficos y las interpretaciones gráficas de éstos como las que se ven en un mapa o en cualquier otra forma de representación. Los datos geográficos atañen a la representación del mundo en términos espaciales que es independiente de cualquier forma particular de visualización de los mismos. Así, al referir a *datos geográficos* se considera a las propiedades y la geometría de los objetos. La *forma cómo son representados* en un mapa, por ejemplo, qué colores o qué tipo de líneas usar no se considera parte del dato.

GML codifica entidades geométricas y tiene capacidad para trabajar con conjuntos de entidades. Una entidad geográfica asociada a una locación se puede conceptualizar

⁵En Inglés *eXtensible Markup Language* (XML).

como un conjunto de entidades correlacionadas espacialmente. El estado de la entidad queda definido por un conjunto de propiedades, cada propiedad está definida por la trílogía <nombre, tipo, valor>. Utiliza el concepto de *agregación*, lo que permite que un conjunto de entidades sea considerado como otra entidad. Por ejemplo, las entidades RECORRIDO, LÍNEA DE COLECTIVO y PARADA se podrían reunir en una nueva entidad TRANSPORTE URBANO DE PASAJEROS. A la nueva entidad se le podrían definir atributos propios, como pueden ser horarios picos, cantidad promedio de pasajeros, etc.

El siguiente ejemplo muestra una codificación básica GML.

Ejemplo 5.4.1 *Instancia de fenómeno*

```
<Feature fid="12" featureType="universidad" Description="universidad nacional">
  <Description>UNS Universidad Nacional del Sur</Description>
  <Property Name="NumCarr" type="Integer" value="57"/>
  <Property Name="NumAlumnos" type="Integer" value="22000"/>
  <Polygon name="extent" srsName="epsg:27354">
    <LineString name="extent" srsName="epsg:27354">
      <CDATA>
        491888.999999459, 5458045.99963358
        491904.999999458, 5458044.99963358
        491908.999999462, 5458064.99963358
        491924.999999461, 5458064.99963358
        491925.999999462, 5458079.99963359
        491977.999999466, 5458120.9996336
        491953.999999466, 5458017.99963357
      </CDATA>
    </LineString>
  </Polygon>
</Feature>
```

Un aspecto poderoso de GML es que da libertad a los usuarios para definir sus esquemas de aplicación personalizados. Del ejemplo anterior se puede observar que define al fenómeno real concreto “Universidad Nacional del Sur (UNS)”. En el primer renglón de la definición se lo instancia como fenómeno y se lo relaciona con tipo de fenómeno UNIVERSIDAD NACIONAL. El fenómeno UNS está descrito por las propiedades descriptivas número de carreras (*NumCarr*) y número de alumnos (*NumAlumnos*). Además tiene una propiedad espacial *polígono* y su referencia con la superficie terrestre.

5.4.2. Descripción de GML schema

GML como la mayoría de las gramáticas basadas en XML, tiene dos partes: el documento de definición de esquema y el documento de instancias que contiene los datos. Da libertad a usuarios y desarrolladores para diseñar datos con geografías genéricas (puntos, líneas y polígonos) según los requerimientos del problema. De todas formas cuenta algunos esquemas predefinidos para ciertas comunidades de aplicación específicas⁶. La posibilidad de definir esquemas de aplicación permite que los usuarios puedan referirse a objetos como *caminos*, *carreteras* ó *puentes* en lugar de puntos, líneas o polígonos. Si una comunidad de uso acuerda en utilizar el mismo esquema, aplicaciones distintas podrán intercambiar datos fácilmente y estar seguros que los datos no pierden identidad. Todos los clientes y proveedores con interfaces que implementen el OpenGIS Web Feature Service Interface Standard⁷ podrán leer o escribir GML data.

Respetando las reglas de ISO 19109, los tipos de fenómenos de una aplicación se especifican en el documento de *esquema de aplicación*. El esquema de aplicación GML especificado en XML Schema se importa a GML Schema. Para definir en GML los tipos de datos de una aplicación se pueden extender o restringir los tipos definidos en GML Schema o usarlos directamente en el esquema de aplicación si no requieren cambios. Todos los esquemas de aplicación se deben modelar siguiendo el Modelo de Fenómeno General de ISO 19109.

La figura 5.14 muestra el modelo de un esquema de aplicación que incluye a los fenómenos PARCELA, CONSTRUCCIÓN y PERSONA. Se observa que PARCELA se relaciona por asociación de propiedad con PERSONA y por agregación con CONSTRUCCIÓN. El diagrama está modelado usando UML e incluye los estereotipos de las clases definidas en ISO 19109. Una posible codificación para el tipo de fenómeno PARCELA podría ser:

Ejemplo 5.4.2 Descripción en GML del tipo de fenómeno PARCELA

```
<!-- Parcela - AbstractFeature -->
<element name="Parcel" substitutionGroup="gml:AbstractFeature">
  <complexType>
    <complexContent>
```

⁶en.wikipedia.org/wiki/GML_Application_Schemas

⁷<http://www.opengeospatial.org/standards/wfs>


```

    <extension base="gml:AbstractFeatureType">
      <sequence>
        <element name="area" type="gml:AreaType"/>
        <element name="extent" type="gml:SurfacePropertyType"/>
        <element name="owner" type="ex:PersonPropertyType" maxOccurs="unbounded">
          <annotation>
            <appinfo><gml:reverseProperty>ex:owns</gml:reverseProperty></appinfo>
          </annotation>
        </element> <element name="hasBuilding" type="ex:BuildingPropertyType" minOccurs=
      </sequence>
    </extension>
  </complexContent>
</complexType>
</element>
<complexType name="ParcelPropertyType">
  <sequence minOccurs="0">
    <element ref="ex:Parcel"/>
  </sequence>
  <attributeGroup ref="gml:AssociationAttributeGroup"/>
  <attributeGroup ref="gml:OwnershipAttributeGroup" />
</complexType>

<!-- ===== -->

```

El código asociado a la definición del esquema de aplicación PARCELA muestra cómo los elementos del modelo se definen respetando las clases del Modelo de Fenómeno General del estándar ISO 19109. El desarrollo completo de este ejemplo se puede ver en el apéndice A.

GML especifica las codificaciones XML de ciertas clases conceptuales definidas en la serie ISO 19100. Los modelos conceptuales relevantes incluyen los elementos de definidos en: ISO/TS 19103 *Conceptual Schema Language (units of measure, basic types)*; ISO 19107 *Spatial Schema (spatial geometry and topology)*; ISO 19108 *Temporal Schema (temporal geometry and topology, temporal reference systems)*; ISO 19109 *Rules for Application Schemas (features)*; ISO 19111 *Spatial referencing by coordinates (coordinate reference systems)*; ISO 19123 *Schema for coverage geometry and functions (coverages, grids)*.

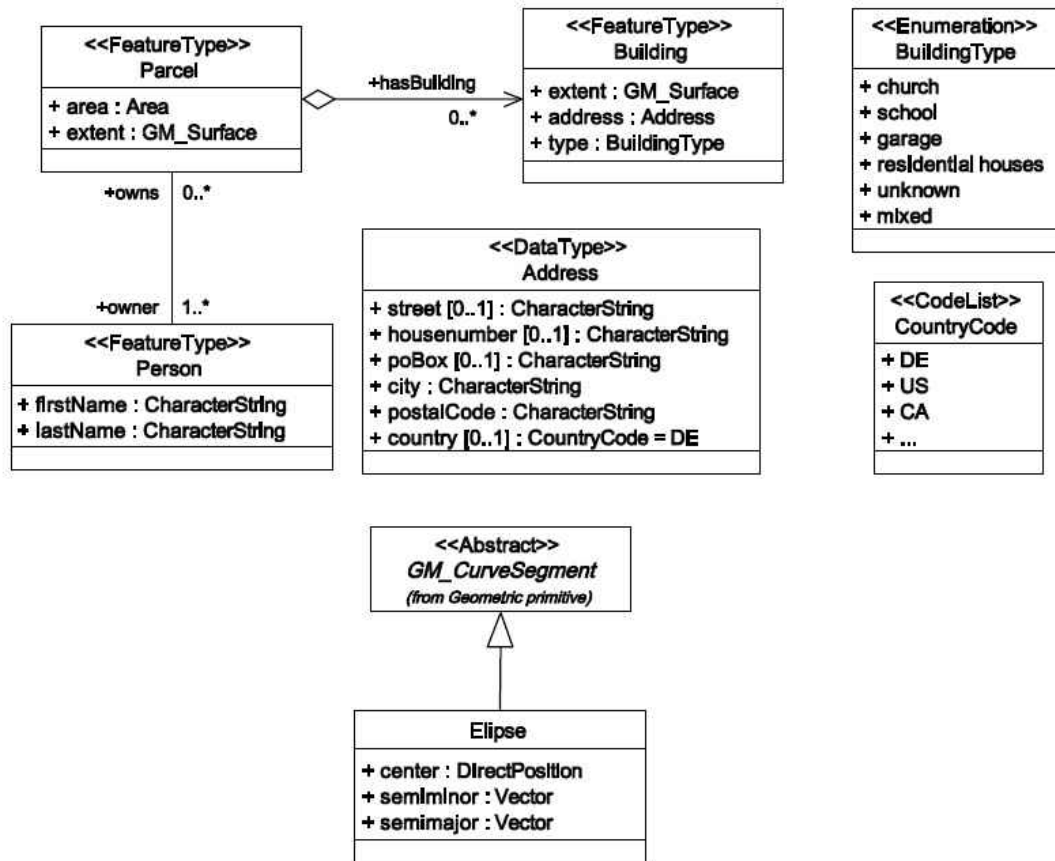


Figura 5.14: Ejemplo de esquema de aplicación Mapa Urbano

5.4.3. Ventajas de GML

A continuación se describen algunas de las ventajas de uso del lenguaje de marcado geográfico GML, y como éstas pueden impactar de forma positiva en el tratamiento de la IG [dSR04]:

- El estar basado en XML lo hace adecuado para transferir la IG a través de Internet. Por otro lado, GML define una codificación XML que permite que los datos geográficos puedan ser transferidos entre sistemas dispares de forma sencilla.
- Los usuarios pueden ver el mapa resultante utilizando un navegador estándar, sin necesidad de un software GIS propietario en el cliente. Esta característica lo hace mucho más versátil y potente que los métodos convencionales basados en imágenes.

- Los mapas son de mayor calidad. GML codifica información sobre entidades geográficas y objetos, y éstos pueden visualizarse con tanta resolución como se requiera.
- Permite manejar estilos de mapa personalizados. En GML se almacena únicamente el contenido del mapa, dónde están los fenómenos, su geometría, tipos y atributos, pero no cómo se deben mostrar. Se pueden aplicar diferentes estilos a los datos geográficos de manera que se muestren como el usuario desee. Por ejemplo, un usuario puede elegir el símbolo de la cruz para situar centros de salud, mientras que otro pondría representarlos con un texto. Los datos GML son los mismos en ambos casos; cada visualización simplemente utiliza un estilo diferente a través de las hojas de estilo.
- Relacionado con el ítem anterior, está el control sobre los contenidos. En GML resulta sencillo aplicar funciones de filtrado que permitan a los usuarios descargar sólo los tipos de fenómenos que deseen visualizar en sus mapas. Por ejemplo, si no se requiere información sobre las líneas de ferrocarril, no se necesitan descargar del servidor, aunque la información esté disponible. El contenido del mapa puede ser controlado incluso después de ser descargado sobre el navegador Web del cliente.
- Una vez que el mapa ha sido descargado y visualizado en el navegador (transformado de GML a SVG) es simple aplicar funciones de edición: añadir texto, agregar marcas, etc.
- Un navegador Web no es la única forma de pensar donde usar la IG, sino que se puede ver como un formato de intercambio de propósito general. Los datos pueden ser compartidos con cualquier otro dispositivo que esté en condiciones de procesar código XML (SIG, teléfonos móviles, etc).
- GML reutiliza la experiencia existente en modelos de información geográfica.

Por supuesto también puede traer algunas desventajas que deberán ser analizadas a la hora de considerar un proyecto usando GML [dSR04]. Genera archivos de texto grandes que desde el punto de vista del almacenamiento no están optimizados y presenta dificultades para manipular archivos raster de gran tamaño.

5.5. Resumen del capítulo

Los requerimientos de estandarización para IG digital en cuanto a transferencia de datos y servicios provienen de la industria IDE y la necesidad de cooperar con sistemas distribuidos. El potencial de los entornos web podrá ser verdaderamente explotado en la medida en que el número de fuentes de IG accesibles sea cada vez mayor. Los avances en estandarización han significado que los entornos de servicios web para IG hayan aumentado considerablemente.

Los resultados más importantes logrados en estandarización e interoperabilidad están conducidos por los organismos internacionales OGC e ISO. Desde sus ámbitos de trabajo ambas organizaciones apuntan a definir una forma normalizada para servicios, catálogos y metadatos, y representación de datos geográficos. Actualmente ambas organizaciones trabajan cooperativamente. Aún el grado de aplicación de las normas es bastante desigual. En general, las normas que son base para la interoperabilidad de paquetes informáticos tienen un alto grado de aplicación en los productores de software. La explosión de las IDE está potenciando enormemente por todo lo relacionado con metadatos y servicios. OGC como representante de la industria, se ocupa de las especificaciones de implementación para servicios. Su objetivo es lograr tecnologías de geoprocésamiento interoperables siguiendo el esquema “plug and play”. Entre los resultados más destacados se pueden nombrar las especificaciones de implementación para diversos tipos de servicios: servicios de mapas web (WMS)⁸[wms04], servicios de fenómenos web (WFS)⁹ [Ope02] y el lenguaje de marcas geográfico GML. Estos resultados han sido llevados adelante respetando los estándares geográficos de la serie ISO 19000.

Los catálogos y metadatos favorecen que distintos clientes puedan encontrar y acceder a recursos, datos y servicios, disponibles en los servidores de los proveedores, aunque éstos no necesariamente sean conocidos previamente por el cliente. Hacer la información geo-disponible justamente se relaciona con dar la posibilidad a los clientes de encontrar proveedores de servicios de geoprocésamiento y datos buscando a través de la descripción de los mismos en metadatos interpretables. En este capítulo se mostraron estándares para datos geográficos y metadatos de esquemas definidos por ISO/TC-211 y adoptados por las implementaciones OGC.

⁸En Inglés *Web Map Service* (WMS).

⁹En Inglés *Web Feature Service* (WFS).

La normalización conlleva notables beneficios; sin embargo, dada la libertad que tienen los usuarios para definir sus esquemas de aplicación y servicios personalizados hace que en muchos casos exista la información que se está buscando, pero que no se llegue a ella porque los patrones de búsquedas siguen siendo por coincidencia y no por significado. Por ejemplo, si un cliente está buscando “zonas de ganado vacuno”, podría no vincularse con servicios de un proveedor que ofrece “cría de Aberdeen Angus”. Esto es así porque para un buscador automático no hay coincidencia en los patrones de búsquedas, aunque semánticamente cliente y proveedor están relacionados. Esto forma parte de los nuevos desafíos semánticos para la información. En el siguiente capítulo se presentan las propuestas de incrementar el significado de la información mediante el uso de ontologías.

Capítulo 6

Ontologías para Información Geográfica

La Web Semántica se relaciona con la noción de interoperabilidad e integración de información. Para ello se requiere de compartir y transferir “conocimiento”. El modelo de comunicación distingue dos entidades: el *emisor* y el *receptor*. Un medio para posibilitar esta comunicación podría estar dado por el uso de *ontologías mediadoras*. Una ontología es una especificación formal de los términos de un dominio junto con las relaciones que existen entre ellos [Gru93]. La ontología define un lenguaje común para los usuarios de una comunidad que desean compartir información. Más interesante aún es que las definiciones de los conceptos y sus relaciones serían potencialmente interpretables por agentes automáticos.

Los diversos repositorios distribuidos con IG y el costo que significa generar y procesar datos geográficos que ya existen en otros medios trae como consecuencia la necesidad de poder evaluar, combinar y reusar recursos de fuentes múltiples. Buscar, recuperar e integrar información es un proceso de comunicación que involucra a un cliente y a uno o varios proveedores. Se espera que en el proceso de comunicación el cliente y los proveedores no interactúen directamente sino a través de un sistema de información que funcione como distribuidor y administre las negociaciones entre ellos. Para que esto sea posible se necesitan resolver los problemas relacionados con la heterogeneidad semántica causada por conceptualizaciones diferentes, terminologías, contexto o información incompleta.

Como se vio en el capítulo anterior la normalización trae beneficios importantes. Al unificar los modelos de representación bajo las ISO 19XXX se ha establecido un marco común para todas las herramientas SIG y se ha favorecido la interoperabilidad actual que existe entre aplicaciones. La norma ISO 19109 permite especificar cómo están definidos los fenómenos geográficos del modelo conceptual de una aplicación concreta, conformando su diccionario de datos accesible. Por su parte, GML es un lenguaje de marcado con capacidad de describir objetos geográficos complejos que sirve para transportar IG a través de la red. Estos resultados han mejorado las posibilidades de integración entre aplicaciones. Sin embargo, aún faltan recursos para hacer realmente geo-disponible a la IG. En la medida que los algoritmos de búsqueda se basen únicamente en “coincidencias” de términos en metadatos y/o datos, el resultado será pobre, ya que no se considera a la conceptualización subyacente. En este capítulo se estudian a las ontologías como medio para enriquecer semánticamente a los modelos de datos espaciales.

6.1. Ontologías y representación de conocimiento

El concepto de *ontología* originalmente se asoció a la investigación en el campo de la Filosofía. Sin embargo, actualmente, varias líneas de investigación en Ciencias de la Computación que aplican a diversas áreas también se ocupan del estudio de ontologías: Ingeniería del Conocimiento, Representación del Conocimiento, Diseño e Integración de Bases de Datos y Recuperación y Extracción de Información, entre otras. La proliferación en la aplicación de ontologías tiene dos aspectos a observar: el *metodológico* y el *arquitectónico* [Gua98]. Desde lo metodológico la característica relevante es la adopción de una apuesta fuertemente interdisciplinaria. La definición de una ontología es el resultado de trabajos colaborativos entre especialistas del dominio, profesionales de ingeniería de software y expertos en filosofía y lingüística, cada uno jugando un rol fundamental para analizar la estructura de una realidad con un alto nivel de generalidad y formulando un vocabulario claro y riguroso.

Desde la vista arquitectónica, se destaca el rol que puede ocupar una ontología en un sistema de información (SI); en una perspectiva futura se habla de *SI conducidos por ontologías*. El enfoque en el uso de ontologías en SI dependerá del momento en que se le dé la oportunidad de intervención a las distintas ontologías previas. A grandes rasgos se pueden evaluar dos instancias diferentes: en la etapa de desarrollo del sistema o en tiempo

de producción. El primer caso propicia SI semánticos, integrables y que eventualmente hasta pueden resultar en nuevas ontologías. La intervención de ontologías en tiempo de producción permite la comunicación entre agentes de software en términos ontológicos y usando contenido semántico.

6.1.1. ¿Por qué definir ontologías?

Las ontologías unifican la interpretación de *conceptos* o *términos* de forma tal que dicha interpretación sea única. Esta afirmación es válida entre personas así como también si en la comunicación participan agentes automáticos. Las comunicaciones personales pueden resolver la heterogeneidad semántica a causa de conceptualizaciones diferentes, terminologías, contexto o información incompleta. Para las personas, el significado de las cosas lo tiene “el término” mismo, como expresión del lenguaje natural junto con las relaciones implícitas que tiene con otros términos. La relación superconcepto/subconcepto entre elementos cotidianos es comprendida claramente por las personas. Por ejemplo, dados los términos “transporte” y “automóvil” entre ellos existe la relación superconcepto/subconcepto que identifica al automóvil como una especialización de un transporte. Sin embargo, esta relación no es trivial para los algoritmos de búsqueda disponibles actualmente que se basan en encontrar coincidencias de términos en los esquemas y los datos.

La tarea de una ontología es la especificación formal de los conceptos y sus relaciones. La figura 6.1 presenta un ejemplo simple de especificación ontológica de conceptos y relaciones. La representación gráfica utilizada es la propuesta por UML [RJB04]. Los conceptos están representados por rectángulos. Entre los conceptos se distinguen las *clases* (en color), que definen las propiedades comunes a los elementos del mismo tipo y las *instancias* (en blanco) que representan a un concepto concreto, ejemplar o “instancia”. En el figura se plantea una relación de jerarquía entre las clases FUENTE DE AGUA DULCE y RÍO por medio de la asociación distinguida *es_un*, en el gráfico representada por una línea con un triángulo blanco sobre el concepto más abarcativo y que define a RÍO como un tipo de FUENTE DE AGUA DULCE. El uso de la asociación distinguida *es_un* refleja la relación entre un concepto especializado y su padre. Definir conceptos y relaciones *es_un* no recursivas entre los mismos explicita una jerarquía o taxonomía. Otras asociaciones definidas en el ejemplo son *desemboca* que especifica que un RÍO *desemboca* en el OCEÁNO y *termina* para representar que RÍO *termina* en una LOCALIDAD. Las instancias de clases

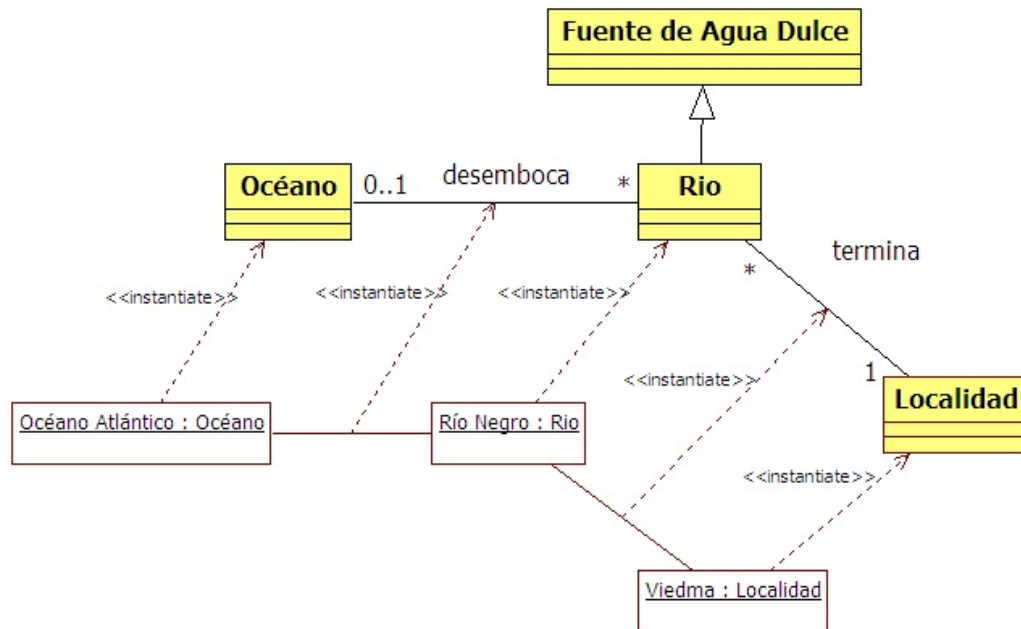


Figura 6.1: Ejemplo de Modelado de Conceptos y Relaciones

y asociaciones están representados por la dependencia estereotipada «*instantiate*». Con esta simple definición es posible obtener algunas conclusiones básicas, como por ejemplo:

- El Río Negro es un RÍO.
- El Océano Atlántico es un OCEANO.
- Viedma es una LOCALIDAD.
- El Río Negro *desemboca* en el Océano Atlántico.
- El Río Negro *termina en* Viedma.

Y otras conclusiones implícitas más elaboradas:

- El Río Negro *es una* FUENTE DE AGUA DULCE.
- Viedma *está próxima* al Océano Atlántico.
- Viedma *cuenta con* una FUENTE DE AGUA DULCE.



Figura 6.2: Composición de una ontología

Lo atractivo de contar con este modelo de representación explícito es que a las mismas conclusiones podría llegar un agente automático, tal como si comprendiera o razonara.

6.1.2. Uso de ontologías en Ciencias de la Computación

El Diccionario de la Real Academia Española define como *ontología* a “*la parte de la metafísica que trata del ser en general y de sus propiedades trascendentales*”. La definición anterior se aplica al área de Ciencias de la Filosofía y su análisis está fuera de los objetivos de esta tesis. Para la comunidad de Ciencias de la Computación, la aplicación del concepto ontología originalmente se vinculó con Inteligencia Artificial. Según la definición de Inteligencia Artificial una ontología es “*un artefacto de ingeniería, constituido por un vocabulario específico usado para describir una determinada realidad, más un conjunto de hipótesis explícitas sobre el significado de las palabras del vocabulario*”. En la figura 6.2 se muestra gráficamente el alcance de esta definición. Las hipótesis deben responder a la estructura de una Lógica de Primer Orden, los términos del vocabulario se utilizan como predicados unarios o binarios, conformando respectivamente los conceptos y las relaciones. En su expresión más simple, una ontología describe una jerarquía de conceptos vinculados por las relaciones de generalización/especialización. En casos más sofisticados, añaden axiomas para expresar otras relaciones entre los conceptos y restringir o precisar la interpretación pretendida.

Las ontologías en Ciencias de la Computación se utilizan como medio para especificar formalmente el significado que se le asigna a la parte del mundo real que se está modelando en un programa o sistema. La especificación formal y explícita permite que el conocimiento que tiene dicha aplicación sobre el mundo real se pueda compartir y reutilizar.

6.2. Formalización

6.2.1. Definición

Una ontología es una especificación formal de los conceptos de un dominio de aplicación. Define un *vocabulario* y *relaciones* de uso común entre varios usuarios que desean compartir información relativa a un dominio [dB93], de forma que las definiciones de conceptos y sus relaciones puedan ser interpretadas por agentes automáticos. El estudio de ontologías aún constituye un tema de investigación. En la bibliografía técnica existen diferentes definiciones para el término ontología, y todas ellas pretenden capturar el mismo espíritu. Hemos seleccionado la siguiente por considerarla la más concisa y clara [Gru93]:

“Una ontología es una especificación explícita y formal de una conceptualización compartida.”

donde el término *conceptualización* se refiere a que utiliza la estrategia de definir modelos abstractos para representar elementos del mundo real trasladados a la computadora con algún propósito. Además, una ontología es una *especificación* porque define de forma concreta tales conceptos. Es *explícita* porque todos los conceptos, relaciones y restricciones deben estar definidos explícitamente. Es *formal* porque las definiciones deben estar expresadas sin ambigüedades de forma tal que sean interpretables por componentes automáticos. Por último, la calidad de *compartida* refleja la noción de que captura conocimiento consensuado por una comunidad de expertos del dominio.

6.2.2. Componentes de una ontología

Conceptos

El *concepto* es la unidad básica a la que la ontología le asigna un significado. Un concepto en una ontología representa a un conjunto de elementos o instancias del mundo real que comparten una misma definición, esto es, significado, propiedades y relaciones. Por ejemplo, una ontología para sistemas hídricos incluirá la definición del concepto RÍO. La especificación deberá ser representativa para todas las instancias o ejemplares de RÍO. Establecida la porción del mundo a modelar por una ontología, ésta deberá incluir la definición del cada uno de los conceptos que maneja para estar completa.

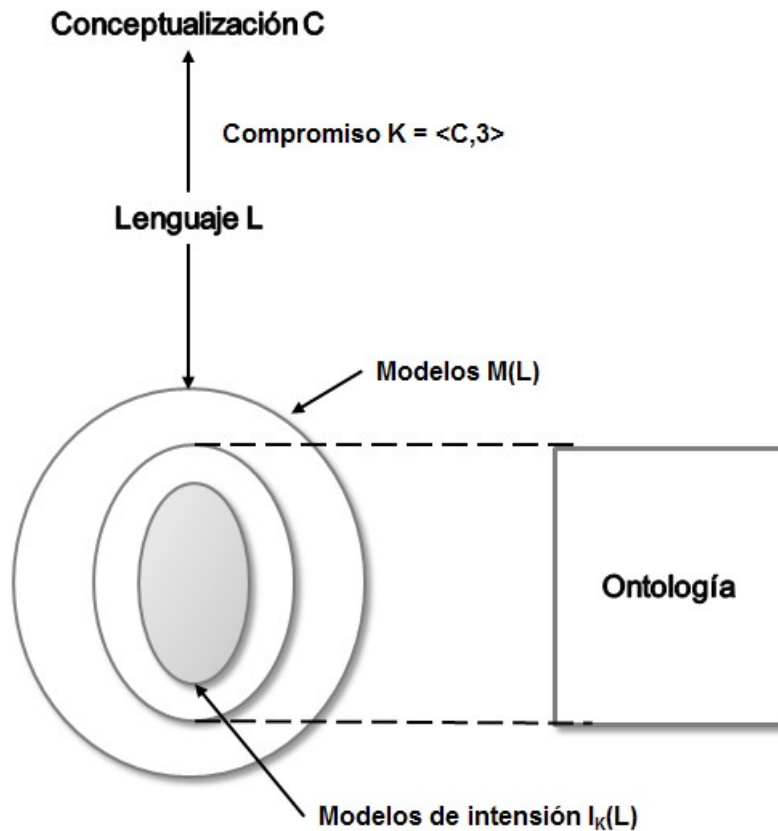


Figura 6.3: Conceptualización y ontología

El significado que se da en este entorno al vocablo “concepto” no se debe confundir con el significado filosófico más amplio que se les da a los términos “Concepto” o “Conceptualización”. La figura 6.3 muestra gráficamente la relación. Mientras que una conceptualización filosófica es la definición que se le asigna a un elemento del mundo real a partir de su participación en varias situaciones o mundos posibles, las ontologías en Ciencias de la Computación definen conceptos en un sentido más estricto y, en parte, restringidas por el poder expresivo del lenguaje ontológico utilizado [Gua98].

Instancias

La especificación de un concepto define el patrón o molde para un grupo de *instancias* o ejemplares. Luego, para dicho concepto pueden existir varias instancias, cada una representando a un objeto del mundo real que se distingue de otros objetos. Por ejemplo, el río “Paraná” es una instancia del concepto RÍO. Las instancias son distinguibles unas

de otras por los valores particulares de sus propiedades y/o sus asociaciones con otros términos.

Propiedades

Las *propiedades* son las características o atributos usados para describir a los conceptos. Para cada concepto se definen un conjunto de una o más propiedades. Las instancias de los conceptos asocian sus valores a las propiedades. Por ejemplo, nombre y longitud podrían ser propiedades que describen al concepto RÍO. La instancia río Paraná tiene los valores de propiedad “Paraná” y “2.570 km.” respectivamente.

Relaciones

Las *relaciones* mantienen las asociaciones que existen entre conceptos, propiedades e instancias. Cuando se define una relación entre conceptos ésta es heredada por sus instancias. Un tipo particular de relaciones frecuentemente utilizado en la definición de ontologías es *generalización/especialización*. Por ejemplo, dados los conceptos FUENTE DE AGUA DULCE y RÍO, entre ellos se define la relación *es_un*¹ que especifica que todo RIO *es una* FUENTE DE AGUA DULCE. Así, la instancia “(Paraná, 2.570 km)”, por ser un ejemplar de RÍO, también es una FUENTE DE AGUA DULCE. El tipo de relaciones capaces de ser definidos por una ontología depende del poder expresivo del lenguaje ontológico que utilice.

6.2.3. Definición formal

Habiendo presentado la definición conceptual de ontología y los elementos que la construyen, estamos en condiciones de proveer una definición formal. Una ontología es 4-*tupla* $O = \langle C, P, I, A \rangle$ donde C representa un conjunto de clases o conceptos, P un conjunto de propiedades, I un conjunto de instancias y A un conjunto de axiomas. Las clases son los patrones para las entidades u objetos del mundo real; las propiedades están asociadas a las clases como atributos o representan relaciones entre las mismas (tal como relaciones de generalización/especialización, agregación, etc); las instancias son los individuos de las

¹Nombre genérico dado a la relación generalización-especialización.

clases; y los axiomas modelan restricciones adicionales que involucran tanto clases como propiedades.

6.3. Tipos de ontologías

Existen diversas ontologías, construidas para diferentes dominios de aplicación. Entre ellas difieren en el nivel de detalle que expresan. La visión de la ontología como un vocabulario más la especificación del significado de los términos de dicho vocabulario sirve como base para organizarlas según el grado de formalidad utilizado en la especificación de significados. Las ontologías más informales usan definiciones en lenguaje natural, las ontologías semiformales proporcionan axiomatizaciones débiles tales como taxonomías y las ontologías formales definen la semántica del vocabulario por una axiomatización completa y efectiva. Los siguientes son algunos ejemplos de ontologías.

Ejemplo 6.3.1 *Ejemplos de ontologías*

- **WordNet**[WOR98] es una ontología que cubre términos del idioma Inglés con descripción en lenguaje natural. Incluye relaciones para sinónimos o términos equivalentes y relaciones de hiper/hipónimo con otros términos. WordNet es una ontología de amplio alcance de aplicación y con nivel de detalle bajo.
- **Cyc**[RL02] es otro ejemplo de ontología de alcance amplio. Formaliza conocimientos básicos o de sentido común. Está desarrollada con alto nivel y alto detalle. Establece varias relaciones formales estrictas entre términos.

De los ejemplos anteriores se pueden tomar dos dimensiones para caracterizar una ontología: *generalidad* y *expresividad*. Generalidad se refiere al campo de aplicación o uso de la ontología. Cuanto más general, mayores serán los dominios donde se pueda emplear. La expresividad se refiere al nivel de formalidad con el que está definida y la capacidad de razonar que provee.

6.3.1. Clasificación según generalidad

Partiendo de que una ontología es una especificación sobre una conceptualización compartida, expertos del dominio, usuarios y diseñadores deben acordar el conocimiento

especificado en la ontología para que la misma sea *usable*, al menos por los miembros de la comunidad. Naturalmente este acuerdo es difícil de lograr. Una forma de abordar una solución es dividir el conocimiento compartido en distintas ontologías, según su alcance o nivel de generalidad. De esta forma, aquellas ontologías más generales se puede usar en distintos problemas, más allá de no estar necesariamente de acuerdo con las definiciones de ontologías de mayor nivel de detalle. Distintos autores proponen formas levemente diferentes en cuanto a la manera de nombrar y clasificar ontologías según el nivel de generalidad [IIMM99, Gua98, Usc98]. Una práctica común es distinguir los siguientes niveles de generalidad:

- *Ontologías genéricas (Top-Level)*: capturan conocimiento de propósito general, independiente del dominio de aplicación, por ejemplo: espacio, tiempo, evento, acción, etc. En teoría se espera que estas ontologías se adopten por una amplia comunidad de usuarios.
- *Ontologías de dominio y ontologías de tarea*: definen el conocimiento usado por un dominio (medicina, geografía) o por una actividad específica (comercio) describiendo su vocabulario mediante la especialización de los términos introducidos en la ontología de alto nivel.
- *Ontologías de aplicación*: capturan el conocimiento necesario de un sistema o aplicación específica. Describen conceptos que dependen tanto del dominio como de la actividad, los cuales frecuentemente son especializaciones de ambas ontologías anteriores. A menudo se corresponde con roles desarrollados por las entidades mientras realizan una actividad.

6.3.2. Clasificación según expresividad

Otra forma de organizar las ontologías es por su grado de expresividad. En este sentido se abre un amplio espectro de posibilidades. En [McG03] se identifican las siguientes posibilidades, en orden ascendente según su poder expresivo:

- *Vocabulario controlado*: es una lista de conceptos que carece de relaciones entre ellos. Las definiciones están dadas en lenguaje natural.

Ejemplo 6.3.2 *Se sugiere el uso de un vocabulario controlado para fijar los términos de un campo o atributo. Por ejemplo, acordar para un atributo **idioma** usar el valor “español” y evitar otros sinónimos como “española”.*

- *Tesaurus*: define algunas relaciones básicas entre conceptos, como sinónimos, término abarcativo, término menor, término raíz. En la sección 4.4.1 del capítulo 4 se describen en detalle las posibles relaciones entre términos de un tesaurus.
- *Taxonomía informal*: incluye una jerarquización explícita de conceptos (generalización y especialización), pero no usa herencia estricta; esto es, una instancia de una subclase no necesariamente es instancia de la superclase.
- *Taxonomía formal*: se diferencia de la anterior en que incluye herencia estricta.
- *Marcos*: definen especies de clases que asocian propiedades a los conceptos. Las propiedades son heredadas por las subclases y las instancias.
- *Con restricciones sobre valores*: agregan la posibilidad de restringir los valores que pueden asumir ciertas propiedades como, por ejemplo, restringir el dominio.
- *Con restricciones lógicas generales*: pueden restringir los valores de las propiedades definiendo para ellos fórmulas lógicas o matemáticas basadas en valores de otras propiedades.
- *Con restricciones de Lógica de Primer Orden*: los lenguajes de ontología más expresivos como Ontolingua o CycL permiten definir restricciones en Lógica de Primer Orden entre términos y relaciones más refinadas como clases disjuntas, cubrimientos, relaciones inversas, relaciones todo-parte, etc.

Otra forma de organizar las ontologías según el nivel de expresividad consiste en dividir las en *ontologías livianas* y *ontologías pesadas*. Las ontologías livianas incluyen conceptos, propiedades que describen los conceptos, relaciones entre conceptos y taxonomías de conceptos. Las ontologías pesadas además agregan axiomas y restricciones. El cuadro 6.1 muestra algunos ejemplos de ontologías conocidas organizadas según la clasificación por poder expresivo presentada.

Nombre	Descripción
<i>Dublin Core</i>	Vocabulario controlado destinado a especificar metadatos de documentos. Contiene términos como “autor”, “título”, “editor”, etc.
<i>WordNet</i> [WOR98]	Tesauro de palabras del idioma Inglés que incluye algunas relaciones básicas entre términos (sinónimos e hipónimos).
<i>AGROVOC</i> [LTM82]	Tesauro estructurado multilingüe del dominio de agricultura, silvicultura, pesca, alimentos y tópicos relacionados como medio ambiente. Consta de palabras y expresiones en diversos idiomas, organizados a través de sus relaciones “término genérico”, “término específico” y “término relacionado”.
<i>Yahoo!Dictionary</i>	Define una taxonomía informal utilizada para clasificación de sitios web. Incluye categorías generales como “economía y negocios”, que a su vez contiene subcategorías como “shopping”, “servicios”, etc.
<i>UNSPSC</i> [UNS]	Taxonomía formal de productos de Estados Unidos. Clasifica los productos desde el punto de vista del proveedor.
<i>Cyc</i> [CYC02]	Es una ontología desarrollada EN KIF (Knowledge Interchange Format). Este lenguaje tiene la expresividad total de un lenguaje de lógica de primer orden.

Cuadro 6.1: Ejemplos de Ontologías

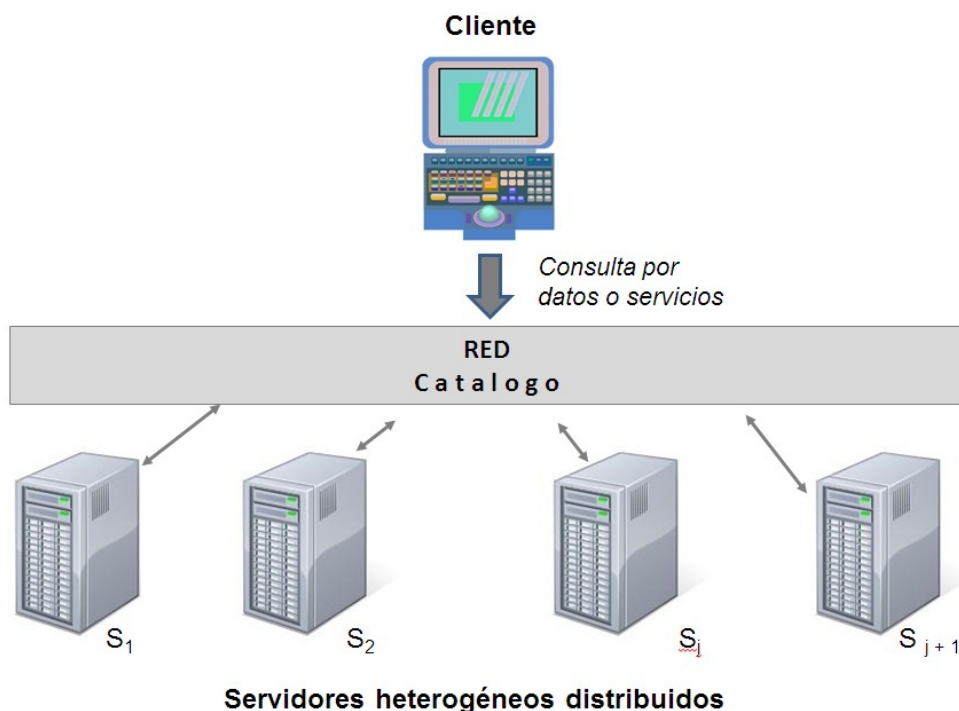


Figura 6.4: El entorno Cliente-Proveedores en el proceso de búsqueda de IG

6.4. Interoperabilidad e Integración para Información Geográfica

6.4.1. Recuperación de IG basada en catálogos

Existe un interés creciente en investigar interoperabilidad e integración semántica dentro del campo de las Ciencias de Información Geográfica. Los repositorios distribuidos y el procesamiento de IG nacen como consecuencia de la necesidad de los clientes de evaluar y combinar recursos de fuentes de múltiples proveedores. La tarea crucial es entonces encontrar y acceder a la IG adecuada distribuida entre los servidores de sus proveedores. En la actualidad, los repositorios de información cuentan con información descripta por catálogos, pero esto en muchos casos resulta insuficiente. La figura 6.4 presenta una representación gráfica de este entorno.

Así, en el proceso de búsqueda de IG se distinguen dos actividades, descubrir a los potenciales proveedores de fuentes de datos y recuperar la información que se necesita.

- **Descubrir IG.** Bajo el contexto de sistemas distribuidos heterogéneos, clientes y proveedores de datos (o servicios) están arbitrariamente distribuidos en grandes redes y cada uno desconoce la existencia del resto. La falta o insuficiencia de documentación hace difícil o imposible a los clientes descubrir las posibles fuentes de datos o determinar si una fuente de datos es útil para sus tareas. Las búsquedas son por catálogo. Los proveedores ofrecen datos y servicios geoespaciales descritos con los metadatos del catálogo. El catálogo reúne metadatos de proveedores conocidos y suele proveer funciones de librería para mostrar y consultar el catálogo.
- **Recuperar IG.** Los formatos heterogéneos hacen difícil comprender y plantear consultas sobre el conjunto de datos de otros sistemas. En varias oportunidades requiere de conversiones. Las soluciones en estandarización favorecen la normalización de interfaces para servicios y formatos de intercambio. Por ejemplo, la especificación Web Feature Service (WFS) de OGC [Ope02] propone interfaces para describir las operaciones de manipulación de fenómenos geográficos y capacidades para recuperar fenómenos basados en restricciones espaciales y no espaciales (`GetFeature`) y para generar la descripción de tipos de fenómenos (`DescribeFeatureType`).

Los problemas de heterogeneidad semántica causados por la ambigüedad del lenguaje natural persisten en la búsqueda por palabras claves en los catálogos, cuando se formulan las consultas para acceder a datos o servicios, causando serios conflictos durante la búsqueda de fuentes de datos y/o servicios apropiados [LRP03]. Por ejemplo, un servicio cuya tarea está descrita por los metadatos “clasificar las áreas de inundación en imágenes satelitales” podría no ser encontrado si se busca por las palabras claves “servicios de clasificación para datos sensados remotamente”. Las dificultades de interoperabilidad semántica ocurren a tres niveles de interpretación [BEH⁺03]:

- Interpretación de metadatos.
- Interpretación de datos.
- Integración de información.

En la interpretación de metadatos ocurre que los términos utilizados para un mismo concepto varían. Esto a veces tiene que ver con el uso de distintos idiomas, pero también dentro de un mismo idioma (sinónimos).

Ejemplo 6.4.1 *Supongamos tres proveedores S_1 , S_2 y S_3 que mantienen información sobre las mediciones de nivel de agua de un río. S_1 registra sus datos con el término “Medida”, mientras que S_2 usa el término “Nivel-de-Agua”. Para las personas, no es difícil ver que ambos términos se refieren a nivel de agua, especialmente en el contexto de mediciones de ríos. En contraste, para un agente de búsqueda automático no existe coherencia entre los términos, y no puede asumir ningún tipo de similitud por lo que los términos son tratados como si fueran totalmente diferentes.*

También ocurre, que en ciertos casos se omiten metadatos. Por ejemplo, el proveedor S_3 provee sus datos en forma de registros sin descripción:

```
... 04.06.2003; 00:45; 138.00;
    04.06.2003; 01:00; 138.00;
    04.06.2003; 01:15; 138.00;
    04.06.2003; 01:30; 138.00;
    04.06.2003; 01:45; 137.00; ...
```

Sin metadatos para un buscador automático es imposible trabajar con hechos ocultos.

En cuanto a interpretación de los datos geográficos ocurren los mismos problemas semánticos. Por ejemplo, sea una fuente de información con datos sobre la cuenca de un río. Estos datos incluyen descripción del tipo de terreno. Existen varios vocabularios para clasificar el uso o destino del suelo. Algunos están estandarizados y su uso está difundido, por ejemplo, ATKIS [atk03] o CORINE [BO00], mientras que otros son creados por los mismos proveedores de datos. La clase de problemas que surge cuando se interpretan datos es similar a los problemas que existen con los metadatos. En la interpretación de metadatos los términos usados para describir los tipos de propiedades (en el modelo relacional los encabezados de las columnas) son ambiguos; en el caso de interpretación de datos los términos usados como valores (en el modelo relacional los valores de atributos) son ambiguos.

El problema de integración de información se deriva de los dos anteriores: cuando datos de fuentes diferentes necesitan ser combinados e integrados en una única colección de datos, se requiere superar ambigüedades de la interpretación de metadatos y datos. Así, en el resultado, cada campo descriptivo de metadato debe pertenecer a un vocabulario definido explícitamente y debe estar claro desde qué vocabulario se toman los términos para comprender el significado de los mismos.

6.4.2. Recuperación de IG basada en ontología

Como vimos, encontrar y recuperar eficientemente IG heterogénea distribuida es un factor clave. Las especificaciones de OGC favorecen la interoperabilidad sintáctica y la clasificación de IG por catálogo. Los problemas causados por heterogeneidad semántica constituyen aún un desafío en recuperación de IG para entornos abiertos distribuidos. Las aproximaciones usando ontologías para hacer conocida la información semántica de fuentes de datos se pueden organizar de distintas maneras. En [WVV⁺01] se introducen distintas arquitecturas usando ontologías para integrar información (ver figura 6.5). En lo que sigue se explican las distintas arquitecturas con referencia al dominio de IG.

La aproximación de la figura 6.5-a) u *ontología única* define una ontología global que provee el vocabulario compartido para especificar la semántica de todas las fuentes de información. Tal solución se puede aplicar a problemas donde las descripciones semánticas disponibles han sido creadas en el marco de una visión de dominio común, también compartido por los consumidores o clientes.

En contraste, en la aproximación de la figura 6.5-b) o *múltiples ontologías*, cada fuente de información está descrita por su propia ontología local u ontología de aplicación. Es posible que una ontología local resulte de la combinación de varias otras ontologías. Sin embargo, no se puede asumir que varias ontologías locales compartan el mismo vocabulario. La falta de un vocabulario común hace difícil comparar diferentes ontologías de aplicación.

La aproximación *ontología híbrida* (figura 6.5 c)) también usa un vocabulario compartido global, que contiene los términos básicos del dominio (primitivas). Estos pueden ser combinados para describir las semánticas más complejas de cada fuente de dato o consulta, resultando en ontologías de aplicación separadas. En contraste con la propuesta múltiples ontologías, los conceptos de cada una de estas ontologías de aplicación son comparables ya que están fundados sobre las primitivas del vocabulario compartido.

Para los casos ontología única y ontología híbrida se asume que la semántica de las primitivas es conocida y comprendida por los clientes y proveedores del dominio. En algunos casos hasta es útil representar al vocabulario compartido como una ontología de dominio. La propuesta ontología híbrida provee mayor flexibilidad para construir ontologías de aplicación [Buc09], mientras que al mismo tiempo las ontologías de aplicación resultan comparables, lo que es crucial al momento de realizar búsquedas semánticas.

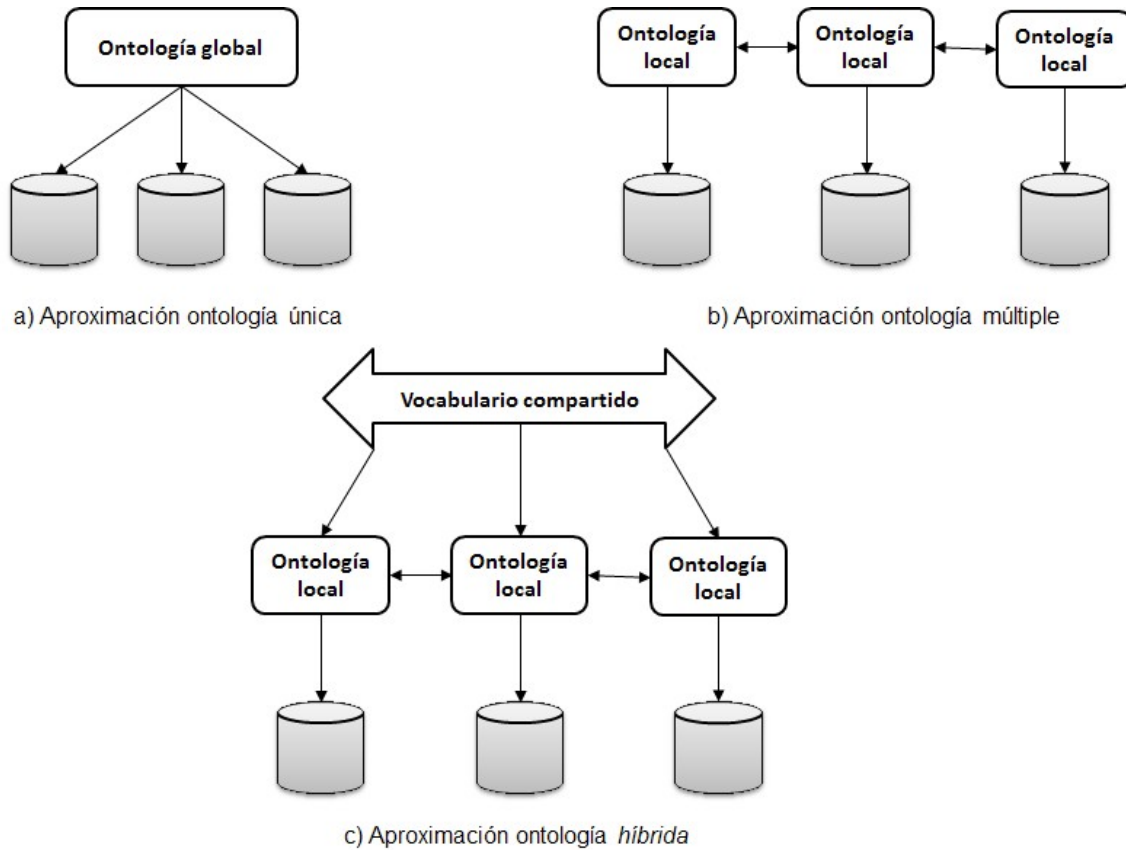


Figura 6.5: Formas de usar ontologías para incluir conocimiento

6.5. Diseño de ontologías para información geográfica

El diseño y definición de ontologías para IG es una actividad creativa compleja en la que deben participar ingenieros de software, responsables de construir los modelos semánticos, y especialistas del dominio encargados de definir, precisar y validar la definición conceptual del dominio. La dificultad no está en la traducción del modelo conceptual a un lenguaje ontológico, sino en la construcción y validación del modelo mismo. El diseño del modelo ontológico destaca el uso de los principios de ingeniería de software de generalidad, abstracción y separación de intereses [GJM02]. En esta sección se presentan consideraciones generales para construir y usar modelos ontológicos.

6.5.1. Creación de ontologías

La construcción de *ontologías de dominio* necesita considerar la definición de un esquema conceptual general para uno o varios dominios de uso. Durante el diseño de ontologías de dominios, las relaciones o roles entre conceptos tienen un significado relevante y deberían ser usados en la descripción semántica siempre que fuera posible [SHW⁺04]. Esto es, se deben incluir vínculos taxonómicos y no taxonómicos entre conceptos. De esta forma, se pueden proveer significados conceptuales más flexibles y precisos que los que tienen con definiciones jerárquicas estáticas.

Ejemplo 6.5.1 *Uso de roles en las definiciones semánticas.*

Continuando con el ejemplo 6.4.1 sobre representación de información para un problema hídrico, entre los conceptos CANTIDAD y FENÓMENO se podría anotar semánticamente a CANTIDAD como lo “observable” de FENÓMENO (figura 6.6).

En la figura 6.6 se muestra una representación gráfica usando notación UML para dos ontologías de dominio relacionadas: *mediciones* e *hidrología* [Kli08]. Se observa que existen asociaciones jerárquicas entre conceptos, representadas por relaciones *es_un*, por ejemplo, entre CANTIDAD y su especialización CANTIDAD HÍDRICA. También existen asociaciones no jerárquicas o “por roles”, expresadas como asociaciones dirigidas con nombre, por ejemplo, MEDIDA *tiene por resultado a* CANTIDAD. La escala de colores está asociada a los conceptos dónde están definidos.

Por otra parte, el enfoque para diseñar *ontologías de aplicación* está en describir semánticamente los conceptos que representan a los tipos de fenómenos geográficos de la aplicación. Esto es, representar la semántica de **Feature Type**. Los conceptos de la ontología de aplicación deberían derivarse como subconceptos de los conceptos de alguna ontología de dominio restringiendo su definición. Esta estrategia asegura no solo información explícita sino también información implícita derivada de la ontología de dominio. La ontología de aplicación no tiene que describir todas las relaciones semánticas posibles en el área, sino describir en forma precisa y formal el contenido de su aplicación.

La propuesta de integración con ontologías híbridas trabaja definiendo ontologías de aplicación y de dominio. Para explicarlo mejor vamos a utilizar el ejemplo de la figura 6.7 tomado de [Kli08]. Sobre la izquierda, siguiendo OGC Reference Model se observa la definición del *esquema para fenómenos geográficos* para ‘‘FT.tierrasBajasInundables’’ en

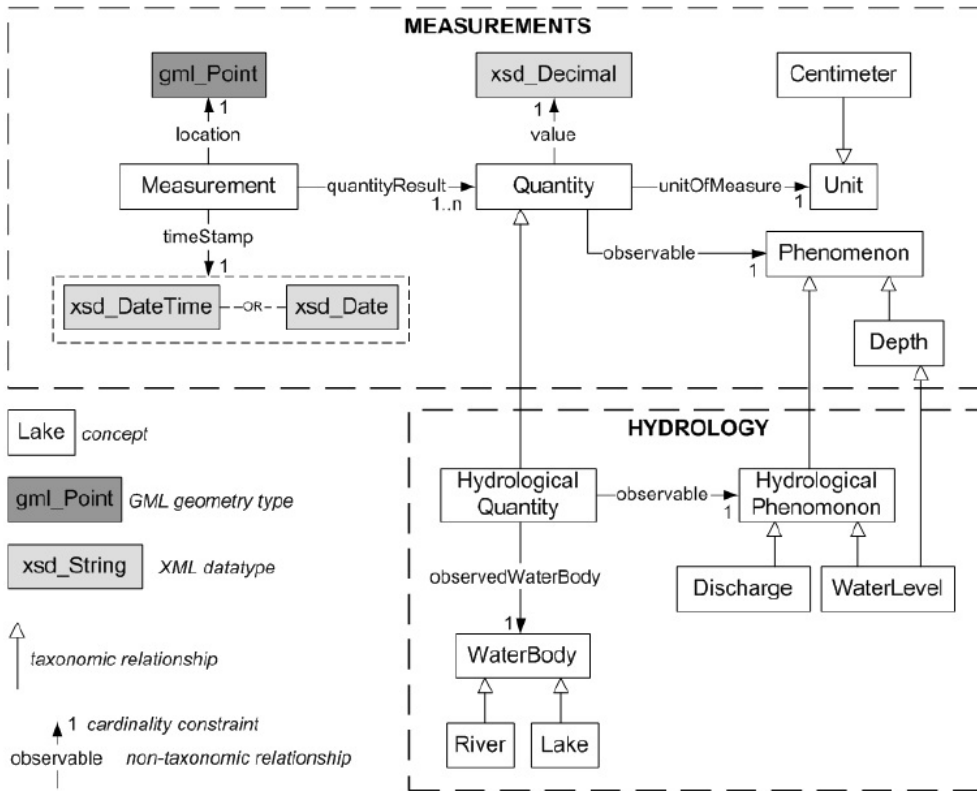


Figura 6.6: Un ejemplo de ontologías de dominio

GML. Esta definición juega el rol de ontología de aplicación. La aplicación contará con múltiples instancias de ‘‘FT_tierraBajaInnundable’’, en el gráfico simbolizadas con la imagen de un fenómeno real. A nivel de ontología de dominio el autor propone una abstracción superior a la que denomina *categoría*. La ontología de dominio para TIERRAİNUNDADA provee una caracterización formal para las entidades que pertenecen a dicha categoría. Finalmente, establece la asociación *anotación semántica* entre la ontología de dominio TIERRAİNUNDADA y la ontología de aplicación del esquema ‘‘FT_tierraBajaInnundable’’ que representa al tipo de fenómeno TIERRABAJAİNUNDABLE definido por una aplicación. La asociación *anotación semántica* no se define como una relación de taxonomía porque el esquema de aplicación y la categoría de la ontología de dominio describen entidades diferentes. También se establecen enlaces *categoriza* entre la ontología de dominio y las instancias de TIERRABAJAİNUNDABLE.

Las anotaciones semánticas establecen vínculos entre *tipo de fenómeno* y *categoría*, y entre *categoría* y *fenómeno*. La flexibilidad extra de la propuesta en [Kli08] está en que

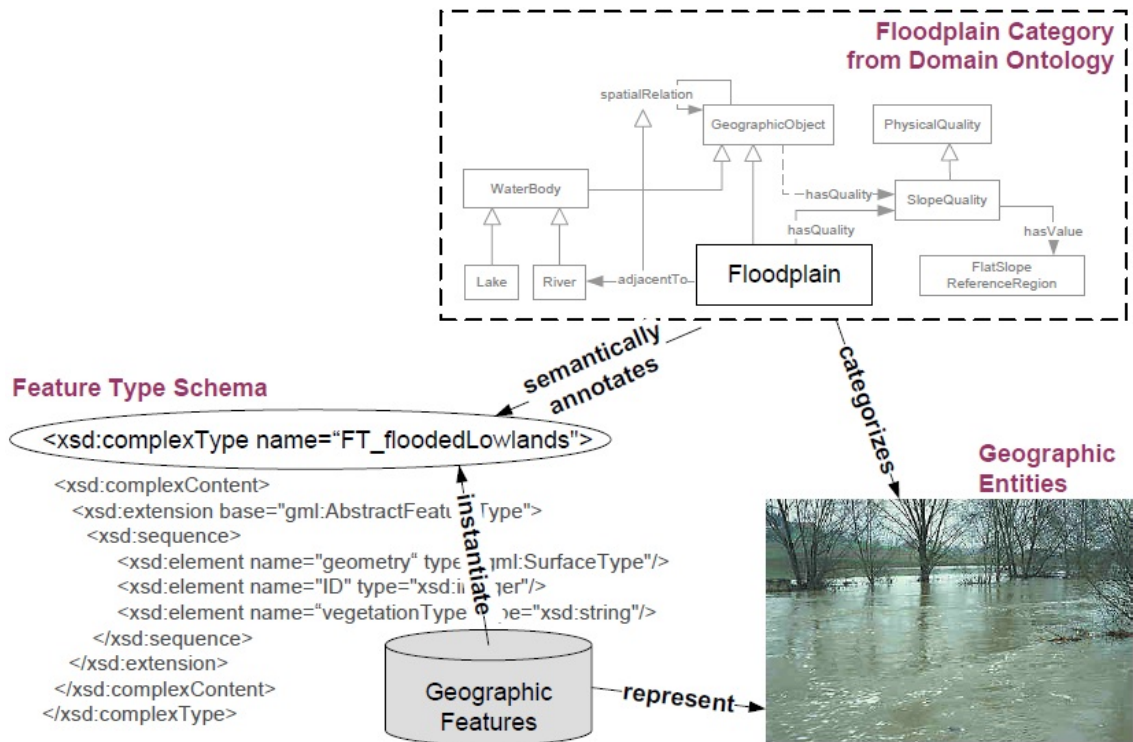


Figura 6.7: Asociaciones semánticas entre Fenómenos, Tipos de Fenómenos y Ontologías de Dominio

se permite a un mismo fenómeno asociarse por *categoría* con más de una ontología de dominio. En la figura 6.8 se muestra cómo un mismo objeto es categorizado por distintas ontologías de dominio: TIERRAINUNDADA, ZONARECREATIVA, HÁBITAT y TIERRASBAJAS. Cada categoría de dominio conceptualiza su visión de dominio. De esta forma, un fenómeno geográfico exhibe las propiedades especificadas por su dueño así como las calidades y relaciones indirectamente definidas en alguna ontología de dominio. Por ejemplo, si una zona inundable está cercana a un río, en el caso de que una zona inundable también juegue el rol de área recreativa estará también próxima a un río.

6.5.2. Cohesión entre una aplicación y una ontología de dominio

A partir de lo visto en la sección anterior, la propuesta de integración con ontología híbrida se muestra la figura 6.9. Cada aplicación cuenta con su ontología de aplicación, la que eventualmente podría estar compartida por más de una aplicación en la medida que la definición conceptual de tipos de fenómenos sea compartida. A mayor nivel de genera-

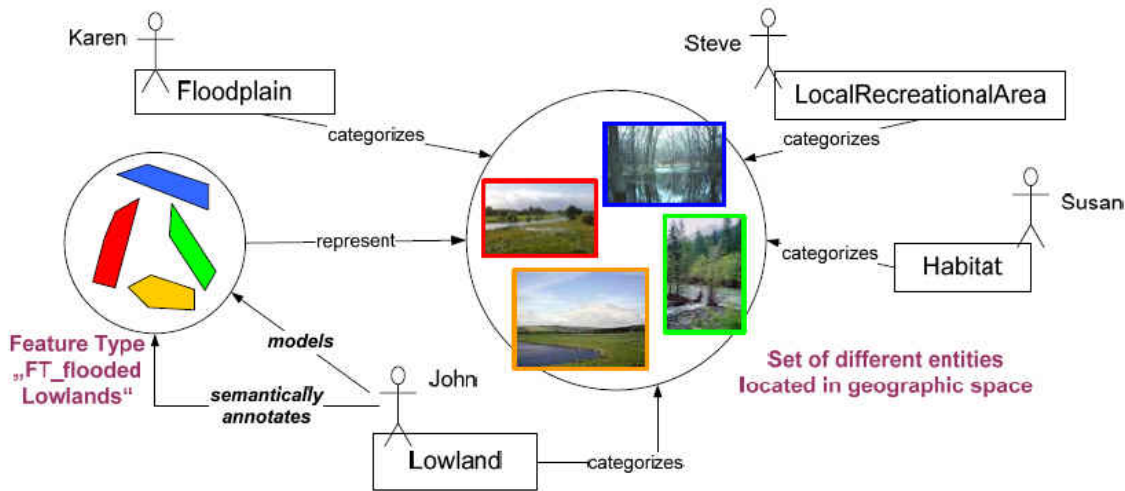


Figura 6.8: Vistas múltiples de un fenómeno geográfico

lidad se ubican las ontologías de dominio que, como se vio, modelan la conceptualización semántica consensuada por el dominio de uso. Esta propuesta supone la existencia de ontologías de dominio de forma que cada aplicación que requiera compartir IG se encargue de definir la conexión entre su ontología de aplicación y la/las ontología de dominio.

A continuación se propone una técnica para determinar el grado de interrelación o cohesión entre una ontología de dominio y un problema particular. En particular, nos interesa la relación de cohesión que existe entre los conceptos definidos por una ontología de dominio y los conceptos de la ontología de aplicación de un SIG.

Si bien la definición de nivel de cohesión no está formalizada, para ganar en rigurosidad se presenta la siguiente definición: sea C_{SIG} el conjunto de conceptos definidos en la ontología de aplicación de un SIG y C_{OD} el conjunto de conceptos definidos por la ontología de dominio, podemos identificar el nivel de cohesión entre el sistema y la ontología según las siguientes relaciones de pertenencia entre los conjuntos:

- $C_{SIG} \subseteq C_{OD} \wedge |C_{SIG}| \approx |C_{OD}|$, se presenta como una situación de máxima cobertura de la ontología de dominio y alta precisión. Representa la relación óptima entre ontología y SIG. Se trata del caso para el cual la ontología de dominio cubre los conceptos requeridos por la aplicación. La ontología contiene las definiciones semánticas requeridas y su semántica se aproxima a las necesidades del problema.

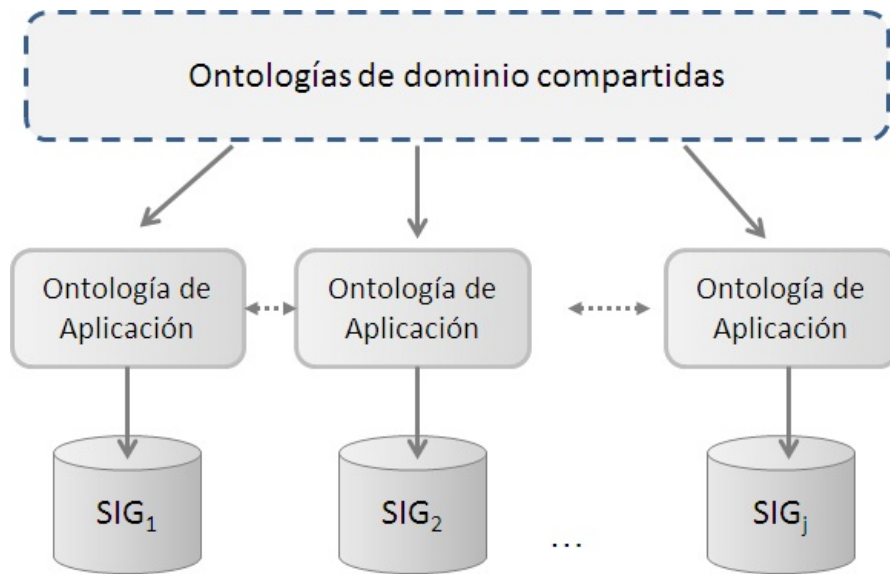


Figura 6.9: Integración entre aplicaciones usando ontología híbrida

- $C_{SIG} \subset C_{OD} \wedge \neg(|C_{SIG}| \approx |C_{OD}|)$, se presenta como una situación de máxima cobertura, pero de baja precisión. La relación entre ontología de dominio y SIG se puede definir como “buena”. En este caso también la ontología cubre los conceptos requeridos por la aplicación. Sin embargo, el contenido semántico de la ontología no está próximo al contenido semántico requerido por el SIG y potencialmente es menos preciso.
- $C_{OD} \subset C_{SIG}$. La ontología de dominio no cubre el universo de conceptos requeridos por la aplicación. Para ciertos conceptos definidos por el SIG no existen conceptos relacionados en la ontología de dominio, esto es, existen conceptos en el SIG que no mapean a conceptos de la ontología. La relación ontología–SIG no es buena. Cuanto mayor sea el número de conceptos que no están en C_{OD} peor es la relación. Una solución puede ser usar más de una ontología de dominio para cubrir todos los conceptos.
- $\neg(C_{SIG} \subset C_{OD} \vee C_{OD} \subset C_{SIG}) \wedge (C_{SIG} \cap C_{OD}) \neq \emptyset$, de ambos lados existen conceptos que no mapean en el otro universo, con lo que se puede afirmar que ambos conjuntos no modelan el mismo dominio del problema. Como en el caso anterior se pueden considerar usar más de una ontología de dominio. Si se usa más de una ontología de dominio es importante que entre ellas no compartan la definición de conceptos.

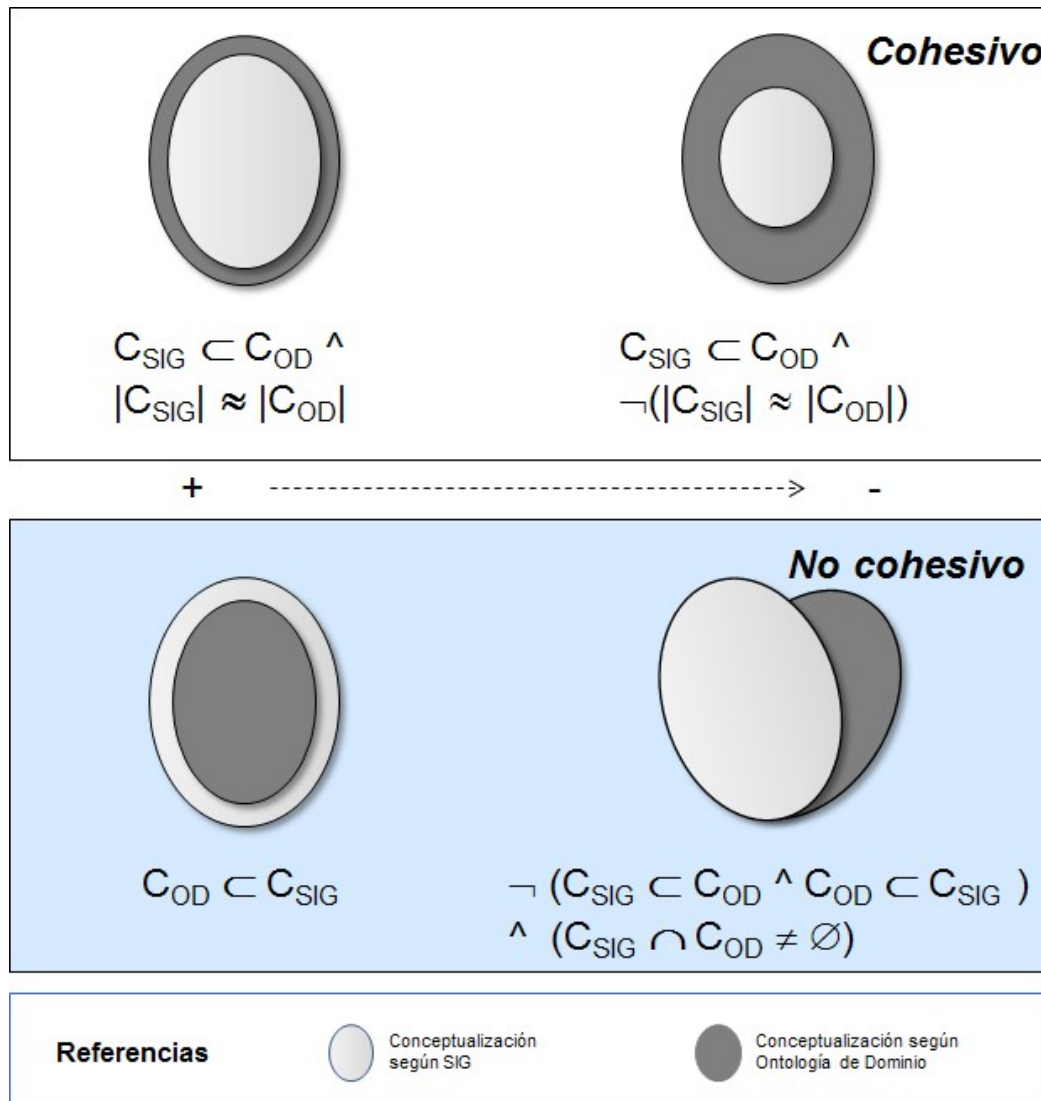


Figura 6.10: Cohesión: Ontología de Aplicación y Ontología de Dominio

En la figura 6.10 se muestra una representación gráfica de estas relaciones.

6.6. Sistemas de Información Geográfica conducidos por ontologías

Hoy día existen numerosas aplicaciones SIG que colaboran con el trabajo cotidiano de empresas, organismos gubernamentales y educativos y de casi cualquier otro ámbito de

aplicación. Generalmente estas aplicaciones cuentan en sus bases de información con datos geográficos acumulados por varios años. La información acumulada resulta una fuente de conocimiento que en muchos casos también sería importante poder compartir y relacionar.

En un sentido figurado, todo SIG cuenta con su propia ontología de aplicación desde el momento en que le asocia significado a los símbolos que usa para representar su visión particular de un problema, la que además, está aceptada por la comunidad que utiliza la aplicación. El problema está en que existen muchas de estas ontologías para usos superpuestos. Esto es, cada una de estas ontologías de aplicación es de generalidad acotada a su entorno de uso. Por otra parte, en cuanto a expresividad también son pobres y muchas veces el significado de los conceptos no está explícito. Todo esto lleva a que a pesar de la cantidad de datos que existen y de que muchas organizaciones están interesadas en compartirlos, se hace difícil cualquier intento por integrarlos.

En forma paralela, la definición de ontologías con distintos niveles de generalidad y expresividad sigue avanzando, buscando satisfacer las necesidades de la web semántica y el análisis de la información a través de agentes automáticos. A continuación se muestran y analizan las distintas alternativas de participación de ontologías en el ciclo de vida de un sistema.

6.6.1. El ciclo de vida del sistema

Si se considera el estado del arte en SIG, existen incontables sistemas funcionando que dan soporte a millones de usuarios y que fueron diseñados bajo sus propias definiciones de conceptos no estándares. Estos sistemas perduran en el tiempo produciendo muchos de los datos que interesa relacionar, analizar y disponer de su contenido semántico. Por otra parte, la dinámica en el ámbito de producción de sistemas indica que constantemente surgen nuevos proyectos de desarrollo. Los nuevos sistemas cubren las necesidades no satisfechas hasta el momento o reemplazan a viejos sistemas adecuándolos a las nuevas alternativas tecnológicas. Estas nuevas ofertas también son generadoras de datos con las mismas necesidades de información semántica explícita.

De lo anterior se puede inferir que una ontología consensuada, con conceptos, relaciones y definiciones semánticas surge mientras diversas aplicaciones SIG que comparten su dominio transitan por diferentes estadios de sus ciclos de vida. Así, el rol de una ontología en SIG se puede analizar bajo las siguientes situaciones:

- Desarrollar una nueva aplicación SIG siguiendo una o más ontologías de dominios existentes.
- Vincular un SIG en producción con una o más ontologías que surgieron posteriormente al sistema.

Para el análisis se tienen en cuenta la clasificación de ontologías según el nivel de generalidad presentada en la sección 6.3.1.

6.6.2. Inclusión de ontologías en la etapa de desarrollo

Ontologías de dominio

Dado el caso de un nuevo proyecto SIG, existe al menos una ontología de dominio/tarea cuyo contenido semántico se adecua al requerido por la nueva aplicación. En este entorno, incluir en las etapas análisis, diseño y desarrollo actividades de conciliación con la/s ontología/s de dominio previas será altamente favorable para el resultado final. El contenido semántico de la/s ontología/s involucradas se trasladará al SIG, donde fácilmente se reducen costos y tiempos de análisis conceptual y se asegura la creación de un SIG que a posteriori se integrará naturalmente con otros sistemas de su comunidad con los que comparta definiciones semánticas. Desde lo técnico, favorece equipos de desarrollo con habilidades para pensar en términos de generalidad, reutilización y mecanismos para compartir conocimiento del área de aplicación y vocabulario estandarizado entre diversas plataformas. Se dice que el SIG resultante es conducido por la ontología, y el sistema se encuadra en un componente clasificado como ontología de aplicación.

El punto desfavorable es que aún no existen ontologías consensuadas de dominio y/o tarea para la mayor parte de los problemas de aplicación con IG. El escenario planteado es difícil de alcanzar. También es cierto, que la de inclusión de ontologías en la fase de desarrollo del sistema aún no está incorporada como una práctica estándar en los modelos de proceso de desarrollo.

Ontologías de nivel general

Esta situación se presenta en el caso de nuevo proyecto de desarrollo para el que existe al menos una ontología de alto nivel cuyos conceptos se ajustan a los tipos de

datos requeridos por el SIG. En la práctica, este entorno comparado con el anterior, es más factible de lograr. Existen ontologías de nivel general que se adecúan a distintos entornos de problemas: un ejemplo son las definiciones ISO 19107, 19108, 19110 y en general catálogos como GEMET [EEA01]. Igual que en la situación anterior, diseñar desde el principio el SIG exigiendo que se ajuste a las definiciones provistas por la ontología de nivel general será favorable para el resultado final del proyecto y su futura integración con otros sistemas. También es cierto que es más simple de aceptar y comprender el alcance de las ontologías de nivel general. El conocimiento semántico de estas ontologías es reducido y de buena calidad. La ontología cumple el rol de asegurar un nivel de abstracción lógico para administrar objetos con atributos que pertenecen a un dominio no trivial, como si fuera un TDA. Al momento de buscar integrar aplicaciones que pertenecen al mismo dominio, esta situación simplificará el trabajo, pero no garantizará que compartan el significado conceptual de los fenómenos.

6.6.3. Vinculación de un SIG en producción con nuevas ontologías

La inclusión de ontologías a un sistema que está en producción es la propuesta para permitir la comunicación de componentes de software existentes en términos ontológicos. En este punto es cuando efectivamente se requiere de la integración del SIG y la ontología. Para ello es necesario mapear los conceptos definidos en el SIG primero al diseño de su propia ontología de aplicación, para luego relacionarlos con los conceptos de la ontología de dominio. Si el SIG desarrollado mantiene concordancia semántica con los términos de la ontología, la relación será más directa estableciendo los mapeos necesarios; de lo contrario, se requerirá de la definición de mapeos parciales entre conceptos, que dependiendo del caso podrán o no ser exactos y causar cierto grado de pérdida de información.

6.7. Resumen del capítulo

La Web Semántica Geográfica se relaciona con las calidades de interoperabilidad e integración de servicios y datos geográficos. Hoy en día existen numerosas fuentes distribuidas con IG heterogénea. Es una necesidad poder evaluar, combinar y reusar los recursos

geográficos en forma transparente. Para que esto sea posible se deben enriquecer los modelos de datos y servicios con su significado semántico y estas definiciones semánticas deberían ser interpretadas por agentes automáticos.

Una propuesta de solución está dada por el uso de ontologías. En Ciencias de la Computación las ontologías sirven para especificar formalmente el significado que se le asigna a los conceptos que modela un programa o sistema. La especificación formal y explícita permite que el conocimiento que tiene dicha aplicación se pueda compartir y reutilizar. Además, si la especificación es formal, es posible navegar el modelo semántico permitiendo que agentes automáticos no solo capturen las definiciones explícitas sino también que pueden hacer inferencias u obtener conocimiento implícito.

El problema está en que el conocimiento semántico de la ontología debe ser consensuado y aprobado por toda la comunidad de uso. Esto, en general, es difícil de alcanzar, especialmente si ya existen aplicaciones con sus propias definiciones conceptuales que no pueden dejarse de utilizar. La solución a este inconveniente está en organizar los conceptos en distintas ontologías clasificadas por generalidad. Así se definen ontologías genéricas, ontologías de dominio y ontologías de aplicación (sección 6.3.1). Existen diversos modelos de integración basados en el uso de ontologías (sección 6.4.2). El modelo de integración más flexible es el de ontología híbrida que establece al menos dos niveles ontológicos conectados: el nivel de aplicación y el nivel de dominio. Aún no existen definiciones ontológicas para muchos de los dominios de la IG. Para el diseño de modelos ontológicos se requiere de un trabajo interdisciplinario que destaque el uso de los principios de generalidad, abstracción y separación de intereses (sección 6.5). El diseño de ontologías de aplicación requiere la adopción de estándares en la definición de los esquemas de los tipos de fenómenos y su vinculación con los conceptos definidos por la ontología de dominio.

Finalmente en este capítulo se elaboró un criterio para determinar el grado de cohesión entre las definiciones semánticas que provee una ontología de dominio y los conceptos definidos por un SIG. Para lograr un nivel de integración óptimo con mínima pérdida de información se requiere encontrar y establecer conexiones semánticas con ontologías de dominio con las que exista buena cohesión (no siempre disponibles). Una forma de lograr esto a futuro será mediante el desarrollo de nuevas aplicaciones que consideren desde su diseño la vinculación con ontologías de dominio existentes. Esto se conoce como desarrollo de sistemas conducidos por ontologías.

Capítulo 7

Conclusiones y Trabajos Futuros

En el capítulo final de esta tesis, se hace un análisis sobre las conclusiones y principales aportes de la investigación realizada para este proyecto. En primer lugar, se presenta un resumen con los puntos principales de estudio y cómo fueron abordados. En la segunda sección se retoman los objetivos presentados en el capítulo 1 y se analizan los resultados obtenidos. En tercer lugar, se presentan las contribuciones teóricas producidas en esta investigación. En la última sección del capítulo se presentan las líneas de investigación futura que quedan abiertas a la investigación.

7.1. Resumen de la investigación

Los sistemas con gestión de IG han ganado relevancia en las últimas décadas. Muchas aplicaciones usadas en ámbitos industriales, gubernamentales y educativos tienen a la IG como componente central. Para muchos otros sistemas las tecnologías de gestión de IG resultan un valor agregado que no se había considerado previamente [Lua04] por ejemplo, si se decide incorporarles servicios basados en ubicación.

La IG es un tipo especial de información que no se representa, manipula y visualiza con los métodos tradicionales usados por los SI. Al comienzo de esta investigación se realizó un trabajo de recopilación de las soluciones propuestas para representar IG, desde lo más concreto, como son los modelos de representación física, hasta lo más conceptual. Esto permitió comprender la necesidad de los SIG de contar con las dos conceptualizaciones del espacio: basada en objetos y basada en campos (sección 2.2), cada una de ellas apropiada

para distintas familias de problemas y con diferentes representaciones y capacidades para consultas y análisis espacial.

Luego, el segundo orden de prioridades en la investigación fue estudiar y caracterizar a las tecnologías SIG y sus aplicaciones. La investigación se abordó desde distintos enfoques: desde la visión de la arquitectura, identificando organización y componentes; desde la visión funcional, clasificando los servicios esperados de un SIG y desde los modelos de datos espaciales y sus capacidades de análisis espacial. Se consiguió ampliar el conocimiento sobre las necesidades y usos de la tecnología SIG, abstraer las particularidades y comprender las exigencias de los modelos de representación en cuanto a generalidad, abstracción y reutilización (capítulo 3).

Las primeras aplicaciones SIG surgieron en la década del '70 y fueron pensadas como soluciones particulares autónomas e independientes. En particular, cada SIG definió sus formatos de representación y almacenamiento para los datos geográficos espaciales. Esto naturalmente representó una dificultad extra al pretender avanzar hacia la interoperabilidad e integración de sistemas heterogéneos. Actualmente, el Consorcio OpenGIS y el Comité Técnico ISO 211 trabajan en la especificación de estándares para la representación, manipulación y visualización de IG que permiten diseñar aplicaciones SIG que usen estándares. Las exigencias en estandarización e interoperabilidad son muchas y se trabaja en paralelo desde diversos frentes. Como parte del trabajo de esta tesis se estudió el estado del arte, los resultados alcanzados, usos y avances en cuanto a la integración de información espacial (capítulo 5.2).

El análisis anterior nos permitió determinar que las soluciones basadas en estándares para la gestión de IG no son suficientes para alcanzar los siguientes objetivos en cuanto interoperabilidad e integración de sistemas con gestión de IG:

- Aprovechar el potencial de los entornos de servicios web.
- Aumentar el número de fuentes de IG accesible.
- Interpretar correctamente el significado de los datos provistos por los usuarios.

Buscar, recuperar e integrar información es un proceso de comunicación que involucra al cliente y a uno o varios proveedores. En este proceso de comunicación se espera que el cliente y los proveedores interactúen de forma transparente. En el capítulo 6 se estudiaron

las propuestas basadas en ontologías para resolver los problemas que surgen de conceptualizaciones diferentes, terminología, contexto y pérdida de información considerando la representación formal y explícita del significado de los datos.

7.2. Resultados obtenidos

Para esta sección vamos a retomar los objetivos planteados en el capítulo 1 y analizar a partir de ellos resultados alcanzados. Recordemos que el objetivo principal definido para este trabajo fue:

“Estudiar *los modelos de representación para IG* con el fin de *establecer las necesidades de descripción semántica* que permitiesen avanzar hacia mejorar la disponibilidad de la IG en entornos heterogéneos distribuidos”.

En el camino hacia el objetivo se plantearon las siguientes metas parciales, de las que vamos ir presentando brevemente los aportes surgidos:

1. *Estudiar las particularidades de la IG.*

- En esta parte de la investigación se caracterizaron y explicaron las dos conceptualizaciones posibles para el espacio geográfico: basada en campos y basada en objetos, según las interpretaciones del espacio geográfico discreto o continuo respectivamente (sección 2.2).
- Se definió al fenómeno geográfico como la unidad de modelado en un SIG. Se analizaron sus requerimientos particulares como elemento de abstracción para incidentes geográficos reales: necesidades de atributos descriptivos y atributos espaciales, representación geométrica y relaciones y operaciones espaciales de composición, geométricas y topológicas (secciones 2.2.1 y 3.2).

2. *Estudiar los modelos de representación para IG existentes, organizándolos según su nivel de abstracción.*

- Al inicio de la investigación se recopiló y analizó un número importante de material y publicaciones de estudio sobre modelos de datos espaciales. Como

experiencia personal, para quién aborda por primera vez el tema de investigación “modelos de representación de IG”, se encuentra mucha información pero poco organizada.

Como parte del trabajo de esta tesis se propone y concreta la organización de los modelos recopilados a partir de considerar a los SIG con un tipo de SI que administra *datos persistentes*, con información geo-referenciada y características geométricas. Así se estudiaron y clasificaron los modelos de datos según la organización que ofrece el área de bases de datos: modelos físicos, modelos lógicos y modelos conceptuales (capítulo 2). Ordenar los modelos en grados de abstracción contribuyó en cuanto a organizar, comprender y presentar cada uno de los modelos, concentrando su descripción en su usabilidad y abstrayéndolos de las responsabilidades de los modelos de otro nivel de abstracción.

3. *Identificar la necesidades y dificultades en integración de IG.*

- Los requerimientos de integración e interoperabilidad para entornos heterogéneos incluyen *encontrar* o capturar a los posibles proveedores de la información requerida e *integrar IG* o administrar la información de los proveedores de acuerdo a las necesidades propias. Encontrar y recuperar en forma eficiente IG distribuida es un factor clave en el planeamiento y soporte de decisión para una variedad de dominios.

Para los problemas que administran IG se identificaron los problemas de integración y se organizaron las dificultades en tres niveles: problemas de representación, problemas sintácticos y problemas semánticos (secciones 4.2.1 y 4.3).

4. *Relevar el estado actual del arte en estandarización de IG y las posibilidades de integración basada en estándares.*

- Los requerimientos de estandarización para IG digital, transferencia de datos y servicios resurgen principalmente con la industria IDE y los LBS. Los resultados en estandarización han significado que los entornos web para IG hayan aumentado considerablemente.

Como parte de esta investigación se estudiaron y presentaron resultados en estándares relacionados con los modelos de representación de IG conducidos

por los organismos internacionales OGC e ISO. Desde sus ámbitos y siguiendo un trabajo colaborativo, ambos organismos definen formas normalizadas para servicios, catálogos y metadatos y representación de datos geográficos (capítulo 4).

5. *Relevar el estado actual del arte en definición de ontologías para IG y sus posibilidades de integración semántica.*

- Los catálogos, formatos y metadatos definidos por los estándares mejoran la recuperación e integración de información. Sin embargo, hacer geo-disponible los datos y servicios geográficos requiere de metadatos semánticos interpretables por los clientes. Una forma de incluir información semántica es a través del uso de ontologías. En este capítulo se realizaron las siguientes tareas:
 - Se reunieron las definiciones y uso de ontologías relacionados con IG (sección 6.1).
 - Se caracterizaron las ontologías para SIG según el nivel de generalidad (sección 6.3.1).
 - Se identificaron posibles arquitecturas para sistemas de integración basados en ontología (sección 6.4.2).
 - Se presentaron las consideraciones para el diseño de ontologías de dominio y ontologías de aplicación (sección 6.5).
 - Se elaboró una propuesta para identificar el nivel de cohesión entre una ontología de dominio y la ontología de aplicación de un desarrollo particular (sección 6.5.2).
 - Se evaluaron las posibilidades de considerar ontologías en el proceso de desarrollo de SIG y el impacto sobre integración e interoperabilidad (sección 6.6).

7.3. Publicaciones

Como parte de esta investigación se realizaron las siguientes presentaciones a Congresos y Workshops:

- “Integración de Sistemas de Información Geográfica” [MV10]. Este trabajo presenta resultados en estandarización, plantea problemas semánticos y soluciones basadas en ontologías con incorporación de información semántica.
- “Interoperabilidad y Estándares de Datos para Información Geográfica” [VF08]. Este trabajo se centra en presentar resultados en estandarización para IG de la familia normativa ISO 19100, desarrollados por el Comité Técnico ISO/TC 211 y por Open Geospatial Consortium (OGC) desde la industria y sus usos.
- “Modelos de Datos Espaciales” y “Modelos de Datos para Datos Espaciales” [VCM05, VFC03]. Estos trabajos resumen la recopilación realizada sobre modelos de datos para IG. En [VFC03] se resumen los modelos de representación lógicos, características, y su comportamiento en consultas especiales. En [VCM05] se presenta la taxonomía de modelos por niveles de abstracción.
- “Interfaces en la Visualización de Información” [DLCMT02]. Este trabajo se escribió en el ámbito de un grupo de investigación sobre Visualización. Si bien no se relaciona directamente con los temas de esta tesis, a partir de la investigación en técnicas de visualización, se disparó el interés en conocer los modelos de representación de IG con el fin de estudiar la visualización y la exploración de colecciones de información espacial.

7.4. Trabajo Futuro

La mayor parte del trabajo de investigación realizado en esta tesis consistió en la recopilación y organización de información distribuida en diversos materiales de estudio sobre modelos de representación de información espacial referidos al dominio de la IG, con el objetivo de caracterizar los requerimientos para la representación de información semántica.

Este trabajo partió del estudio de un dominio de aplicación desconocido y del estudio de ontologías para SI y su generalización hacia ontologías para SIG. Como parte de este estudio se elaboraron los siguientes resultados propios que requieren del desarrollo e investigación futura:

- Se identificó la propuesta de integración de información basada en la arquitectura de ontología híbrida como la más flexible y se planteó su personalización a SIG usando ontologías de dominio geográfico y ontologías de aplicación que respeten el estándar GFM de ISO 19109. Se planteo además una propuesta de diseño para las ontologías de esquemas de aplicación de tipos de fenómeno y sus relaciones semánticas con la o las ontologías de dominio. Queda abierto como trabajo futuro experimentar en aplicaciones del dominio geográfico la propuesta de diseño de ontologías de aplicación para tipos de fenómenos.
- La implementación de integración usando la arquitectura de ontología híbrida requiere identificar para una aplicación una o más ontologías de dominio. En este trabajo se propuso una técnica empírica para medir el nivel de interrelación entre una ontología de aplicación y una ontología de dominio. Esta clasificación está basada únicamente en la experiencia personal en el desarrollo y migración de sistemas. Queda propuesto como trabajo futuro profundizar y formalizar la propuesta para poder medir nominalmente la cohesión entre una ontología de dominio y un SIG. Alcanzar resultados en este sentido puede constituir un aporte significativo a la investigación.
- Se analizaron las posibilidades de incluir el uso de modelos ontológicos en las distintas etapas del ciclo de vida de un SIG. Para desarrollos futuros se debería preveer la construcción del sistema considerando las definiciones semánticas de uno o más modelos ontológicos de dominio. Como en el punto anterior, en esta tesis se analizaron y presentaron las distintas posibilidades, junto con sus ventajas y desventajas a partir de la experiencia empírica. Queda propuesto como tema de investigación futuro formalizar para cada una de las actividades de desarrollo su vinculación con los modelos ontológicos y definir lo que sería un modelo de proceso de ingeniería de software incluyendo ontologías.

Apéndice A

Ejemplo GML

En esta sección se transcribe la definición complete en GML de figura A La figura muestra un modelo conceptual para una aplicación que incluye a los fenómenos PARCELA (PARCEL), CONSTRUCCIÓN (BUILDING) y PERSONA (PERSON). PARCELA es un tipo de fenómeno geográfico que está relacionado por asociación de *propiedad* con PERSONA y por *agregación* con el fenómeno geográfico CONSTRUCCIÓN.

A.1. Definición GML del Esquema de Aplicacion Mapa Urbano

```
<?xml version="1.0" encoding="UTF-8"?>
<schema targetNamespace="http://www.someorg.de/example"
xmlns="http://www.w3.org/2001/XMLSchema" xmlns:ex="http://www.someorg.de/example"
xmlns:gml="http://www.opengis.net/gml/3.2" elementFormDefault="qualified" version="1.0">
  <!-- ===== -->
  <import namespace="http://www.opengis.net/gml/3.2" schemaLocation="./gml.xsd"/>
  <import namespace="http://www.w3.org/1999/xlink" schemaLocation="./xlinks.xsd"/>
  <!-- ===== -->

  <!-- Parcela - AbstractFeature -->
  <element name="Parcel" substitutionGroup="gml:AbstractFeature">
    <complexType>
      <complexContent>
        <extension base="gml:AbstractFeatureType">
```

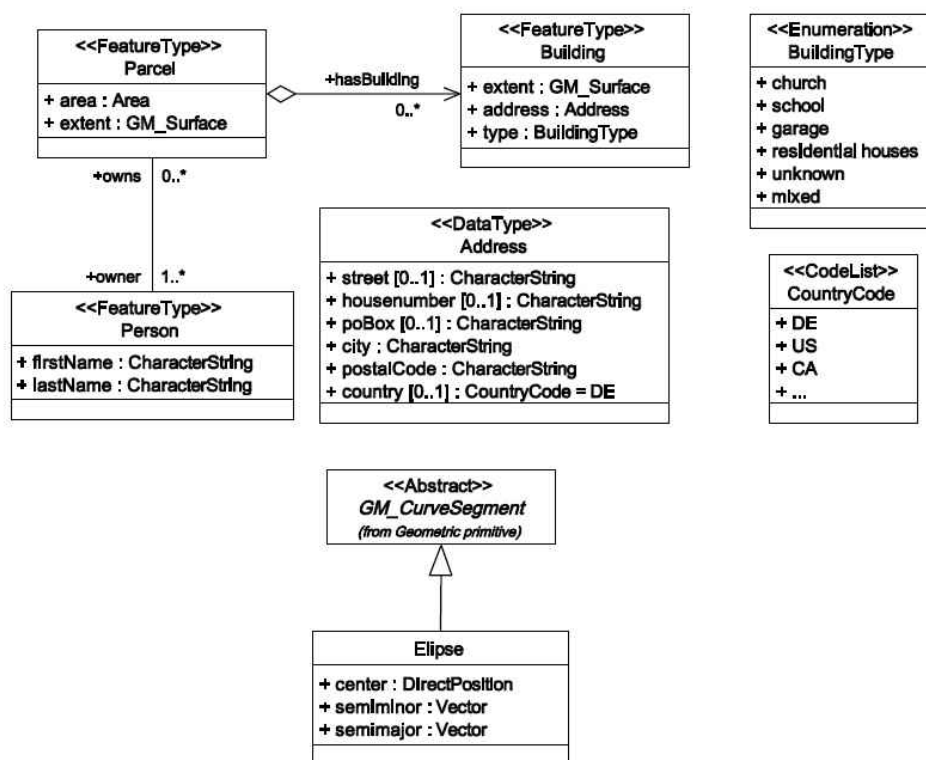


Figura A.1: Ejemplo de esquema de aplicación Mapa Urbano

```

        <sequence>
          <element name="area" type="gml:AreaType"/>
          <element name="extent" type="gml:SurfacePropertyType"/>
          <element name="owner" type="ex:PersonPropertyType" maxOccurs="unbounded">
            <annotation>
              <appinfo><gml:reverseProperty>ex:owns</gml:reverseProperty></appinfo>
            </annotation>
          </element> <element name="hasBuilding" type="ex:BuildingPropertyType" minOccurs="1" maxOccurs="1" />
        </sequence>
      </extension>
    </complexContent>
  </complexType>
</element>
<complexType name="ParcelPropertyType">
  <sequence minOccurs="0">
    <element ref="ex:Parcel"/>
  </sequence>
  <attributeGroup ref="gml:AssociationAttributeGroup"/>
  <attributeGroup ref="gml:OwnershipAttributeGroup" />
</complexType>

<!-- ===== -->

<!-- Construccion - AbstractFeature -->
<element name="Building" substitutionGroup="gml:AbstractFeature">
  <complexType>
    <complexContent>
      <extension base="gml:AbstractFeatureType">
        <sequence>
          <element name="extent" type="gml:SurfacePropertyType"/>
          <element name="address">
            <complexType>
              <sequence>
                <element name="Address" type="ex:AddressType"/>
              </sequence>
            </complexType>
          </element>
          <element name="type" type="ex:BuildingTypeType"/>
        </sequence>
      </extension>
    </complexContent>
  </complexType>

```

```

    </complexType>
</element>
<complexType name="BuildingPropertyType">
  <sequence minOccurs="0">
    <element ref="ex:Building"/>
  </sequence>
  <attributeGroup ref="gml:AssociationAttributeGroup"/>
  <attributeGroup ref="gml:OwnershipAttributeGroup" />
</complexType>

<!-- ===== -->

<!-- Persona - AbstractFeature -->
<element name="Person" substitutionGroup="gml:AbstractFeature">
  <complexType>
    <complexContent>
      <extension base="gml:AbstractFeatureType">
        <sequence>
          <element name="firstName" type="string"/>
          <element name="lastName" type="string"/>
          <element name="owns" type="ex:ParcelPropertyType" minOccurs="0" maxOccurs="unbounded">
            <annotation>
              <appinfo><gml:reverseProperty>ex:owner</gml:reverseProperty></appinfo>
            </annotation>
          </element>
        </sequence>
      </extension>
    </complexContent>
  </complexType>
</element>
<complexType name="PersonPropertyType">
  <sequence minOccurs="0">
    <element ref="ex:Person"/>
  </sequence>
  <attributeGroup ref="gml:AssociationAttributeGroup"/>
  <attributeGroup ref="gml:OwnershipAttributeGroup" />
</complexType>

<!-- ===== -->
<complexType name="AddressType">
  <sequence>

```

```

        <element name="street" type="string" minOccurs="0"/>
        <element name="houseNumber" type="string" minOccurs="0"/>
        <element name="poBox" type="string" minOccurs="0"/>
        <element name="city" type="string"/>
        <element name="postalCode" type="string"/>
        <element name="country" type="ex:CountryCodeType" minOccurs="0" default="DE"/>
    </sequence>
</complexType>
<!-- ===== -->
<simpleType name="BuildingTypeType">
    <restriction base="string">
        <enumeration value="church"/>
        <enumeration value="school"/>
        <enumeration value="garage"/>
        <enumeration value="residential houses"/>
        <enumeration value="unknown"/>
        <enumeration value="mixed"/>
    </restriction>
</simpleType>
<!-- ===== -->
<simpleType name="CountryCodeType">
    <union memberTypes="ex:CountryCodeEnumerationType ex:CountryCodeOtherType"/>
</simpleType>
<simpleType name="CountryCodeEnumerationType">
    <restriction base="string">
        <enumeration value="DE"/>
        <enumeration value="US"/>
        <enumeration value="CA"/>
        <enumeration value="..."/>
    </restriction>
</simpleType>
<simpleType name="CountryCodeOtherType">
    <restriction base="string">
        <pattern value="other: \w{2,}"/>
    </restriction>
</simpleType>
<!-- ===== -->
<element name="Ellipse" type="ex:EllipseType" substitutionGroup="gml:AbstractCurveSegment"/>
<complexType name="EllipseType">
    <complexContent>
        <extension base="gml:AbstractCurveSegmentType">

```

```
<sequence>
  <element name="center" type="gml:DirectPositionType"/>
  <element name="semiminor" type="gml:VectorType"/>
  <element name="semimajor" type="gml:VectorType"/>
</sequence>
</extension>
</complexContent>
</complexType>
</schema>
```

Apéndice B

Abreviaturas y Acrónimos

GML (GML)¹: Lenguaje de Mercado Geográfico.

IDE (SDI): Infraestructuras de Datos Espaciales.

IG (GI): Información Geográfica.

ISO/TC 211: Comité Técnico Internacional para Estadarización de IG.

MER (ERM): Modelo Entidad Relación.

SMBL (LBMS): Sistemas Móviles Basados en Localización.

SBL (LBS): Servicios Basados en la Localización.

MADS: Modeling of Application Data with Spatio-temporal features. Lenguaje de modelado para aplicaciones geográficas.

OMT: Obejet-Oriented Modeling Technique. Lenguaje de modelado.

OMT-G: Object-Oriented Data Model for Geographic Applications. Lenguaje de modelado para aplicaciones geográficas.

SI (IS): Sistema de Información.

SIG (GIS): Sistema de Información Geográfico.

SMBD (DBMS): Sistema de Gestión (Manejo) de Bases de Datos.

SMBDE (SDBMS): Sistema de Gestión (Manejo) de Bases de Datos Espaciales.

¹Acrónimo en *Inglés*

SMBDROO (ORDBMS): Sistema de Gestión (Manejo) de Bases de Datos Relacionales Orientadas a Objetos.

TDAE (SADT): Tipo de Dato Abstracto Espacial.

UML: Lenguaje Unificado de Modelado.

XML: Lenguaje de Marcado Extensible.

Apéndice C

Glosario

En esta sección se incluyen definiciones en lenguaje natural para algunos términos de uso frecuente en el entorno de SIG y con las interpretaciones usadas en el desarrollo de este trabajo de tesis:

A,B,C

Arquitectura de ontología híbrida – define un vocabulario compartido global con los términos del dominio (primitivas). Estos se vinculan con las semánticas de aplicación de cada fuente de información.

Arquitectura de ontología múltiple – cada fuente de información define su ontología local u ontología de aplicación.

Arquitectura de ontología única – define una única ontología global que provee el vocabulario compartido para especificar la semántica de todas las fuentes de información.

Bases de Datos Espaciales – son un tipo de base de datos basada en un modelo de datos espacial.

Ciclo de vida del sistema – define los estadios por los que transcurre el desarrollo de sistema particular, desde la concepción del mismo, pasando por las actividades de análisis, diseño, implementación y mantenimiento, hasta que se deja de usar.

D,E,F

Dato Geográfico – información relativa a una locación en la superficie terrestre.

Dato Espacial – información relativa a una locación en un plano. Un dato geográfico es una especialización para dato espacial.

Esquema conceptual de aplicación – el esquema de base de datos conceptual que surge de del diseño del modelo conceptual representado usando las abstracciones de un lenguaje de representación conceptual.

Esquema lógico de aplicación – esquema de base de datos lógico o discreto. Surge de la traducción de cada elemento del esquema conceptual de aplicación a su correspondiente abstracción en el modelo discreto usando los componentes de un lenguaje de modelado lógico.

Fenómeno (feature) – objeto geográfico distinguible de otros. Por ejemplo, el parque nacional “Los Arrayanes” es un fenómeno del conjunto de fenómenos “parques”. Los fenómenos se describen por atributos tradicionales y atributos espaciales.

G,H,I

General Feature Model GFM – Modelo de Fenómeno General. Definido por ISO 19109. Proporciona la descripción formal para el esquema de datos, la descripción general de los elementos tipo de fenómeno y sus atributos.

Geography Markup Language - GML – lenguaje de marcado basado en XML definido para transportar IG y su definición sintáctica a través de la red. GML cuenta con capacidades para describir objetos geográficos complejos. Fue desarrollado en el marco OGC y está basado en OpenGIS Abstract Specification y la serie ISO 19100 de ISO/TC 211.

GPS – Sistema de posicionamiento global.

Heterogeneidad Estructural – se relaciona con la forma en que dos soluciones representan los modelos mentales usando distintos atributos descriptivos.

Heterogeneidad Semántica – considera el “significado” de cada elemento de información. La heterogeneidad semánticos ocurre cuando en contextos diferentes no se utilizan las mismas interpretaciones. Se requiere de reglas de integración semántica para encontrar las correspondencias o “mapeos” entre los modelos.

Heterogeneidad Sintáctica – cuando dos soluciones almacenan sus datos bajo modelos de representación diferentes. Se refiere al modelo de datos utilizado.

Integración de información – se refiere a la necesidad de compartir datos entre sistemas heterogéneos. Más específicamente información, esto es datos y su significado.

Interoperabilidad – se relaciona con la calidad de ciertos sistemas que cuentan con capacidades para integrar componentes de software que provienen de diferentes proveedores.

ISO/TC 211 – comité técnico dentro de la Organización Internacional de Estándares (ISO) creado para la normalización en Geomática e Información Geográfica. A este comité se le asignó la responsabilidad de definir normativas de referencia para IG digital y transferencia de datos y servicios requeridos por la industria de SIG distribuidos.

J,K,L

LBMS – Sistema de gestión basado en ubicación.

LBS – Servicio basado en ubicación.

M,N,O

Modelo de Clases UML – el modelo o diagrama de clase de UML conceptual visualiza las clases que involucra un sistema y sus relaciones. Un diagrama de clases está compuesto por las *clases* con atributos, métodos y visibilidad, y las *relaciones*, que pueden ser de herencia, composición, agregación, asociación o uso.

Modelo Conceptual Estándar para Esquemas de Aplicación - ISO 19109 – estándar internacional ISO que define las Reglas para crear y documentar Esquemas de Aplicación.

Modelo Conceptual Estándar para Fenómenos Geográficos - ISO 19107 –

estándar internacional ISO que proporciona los elementos conceptuales para describir y manipular las características espaciales asociadas a fenómenos geográficos.

Modelo de Referencia - ISO 19101 – estándar ISO que provee el marco de la serie de estándares ISO 19XXX sobre modelos geoespaciales. Introduce el "Modelo de Referencia de Dominio", que provee la arquitectura de información para SIG distribuidos.

Modelo Entidad Relación (MER) – modelo conceptual de datos para representar la vista estructural de un sistema. Expresa el modelo de datos a través del conjunto de entidades relevantes para un sistema, junto con sus propiedades e interrelaciones.

Modelo de Datos Espaciales – modelo de datos con abstracciones para objetos espaciales. Los principales modelos de datos espaciales son:

- **Raster**– divide el terreno o espacio en celdas iguales y regulares en tamaño y forma. El valor de cualquier celda para un atributo particular se corresponde con el valor del atributo en dicha celda, por ejemplo *nivel sobre el mar*. Cuando se desea tener información sobre otro atributo, por ejemplo temperatura, se define lo que se conoce como capa (en inglés *layer*). La celda se asocia con el pixel
- **Vector**– divide el espacio en elementos discretos, basado en alguna característica geográfica. Se define una jerarquía de tres particiones básicas para el modelo vector:
 1. **Punto**: se representa por un par coordenado (x, y)
 2. **Línea**: es una secuencia de puntos que forma un arco dirigido.
 3. **Polígono**: ciclo simple cerrado limitando a una región

Las estructuras de datos usadas para implementar el modelo vector incluyen K-D-B trees, quadtrees, grid files y R-trees

Modelo espacial basado en campos – asocia a cada punto del espacio con uno o varios valores de atributos definidos como funciones continuas sobre x e y . Por ejemplo, altitud sobre el nivel del mar.

Modelo espacial basado en objetos – asocia sobre el espacio entidades discretas. Cada entidad tiene asociada un componente espacial, que corresponde con la forma y ubicación del objeto en el espacio subyacente, y un conjunto de propiedades descriptivas.

Niveles de Abstracción de Datos – son un medio para ordenar los modelos de datos en sistemas que administran datos con persistencia. Los niveles de abstracción usados incluyen.

- *Nivel conceptual*– modela los objetos del mundo real y sus relaciones sin considerar las limitaciones de representación.
- *Nivel lógico*– provee los constructores para representar los datos dentro del sistema. Oculta detalles de implementación.
- *Nivel físico*– se ocupan de las estructuras para que la gestión (recuperación y guardado) de los datos sea eficiente. Trabajan con estructuras de archivos e índices.

Objetos Geográficos – representa la unidad de información a nivel conceptual. Un tema es una colección de *objetos geográficos*. Un objeto geográfico se corresponde con una entidad del mundo real y tiene dos componentes:

- Una *descripción*, cada objeto se describe por su conjunto de atributos descriptivos. Por ejemplo el nombre y la población de una ciudad.
- Una *componente espacial*, también denominado *objeto espacial* o *extensión espacial*, puede incluir geometría (ubicación en el espacio geométrico subyacente, forma, etc.) y topología (relaciones espaciales con otros objetos). Por ejemplo, una ciudad puede tener como valor geométrico un polígono en el espacio 2D. La componente espacial aislada de un objeto geométrico se denomina *objeto espacial*. Puede darse el caso que se considere separadamente.

OGC Open GeoSpatial Consortium – organismo internacional integrado por representantes del gobierno, de las ciencias de investigación y socios de la industria que tienen por objetivo definir y divulgar estándares para representar y manipular IG que favorezcan la interoperabilidad entre herramientas SIG.

Ontología – en Ciencias de la Computación se define como ontología a la especificación explícita y formal de una conceptualización compartida.

Ontología de Aplicación – clasificación de ontología según generalidad. Captura el conocimiento de un sistema o aplicación. Describen conceptos que dependen del dominio. A menudo se corresponde con roles desarrollados por las entidades mientras realizan una actividad.

Ontología de dominio – clasificación de ontología según generalidad. Define el conocimiento usado por un dominio o actividad específica describiendo su vocabulario. También se denominan ontologías de tarea.

Ontología genéricas (Top-Level) – clasificación de ontología según generalidad. Captura conocimiento de propósito general independiente del dominio de aplicación.

Objetos geométricos – según su dimensión incluyen *punto*, *línea* y *polígono*. Los objetos geométricos se pueden agrupar formando conjuntos de objetos de la misma dimensión (por ejemplo conjunto de puntos) o distinta.

Operador Espacial – operador del lenguaje de consulta que opera sobre datos espaciales. Por ejemplo *join espacial*, *selección espacial*.

Operación Geométrica – operaciones métricas sobre fenómenos geográficos. Incluye distancia, superficie, etc.

Operación Topológica – operaciones que permanecen invariantes ante cambios morfológicos, de escala o de proyección. Incluye vecindad, adyacencia, inclusión y conectividad.

P,Q,R

Proceso ETL – proceso que sigue un modelo lineal para la extracción, transformación y carga de datos.

Relaciones espaciales – son relaciones entre *objetos* del espacio. Se organizan en:

- **Topológica**– se refieren a calidades topológicas objetos espaciales que son invariantes bajo transformaciones topológicas, si un objeto está adyacente a, conectado a, incluido en, disjuntos con,

- **Direccionales**— posición relativa de un objeto con otro.
- **Métricas**— dan por resultado una medida, como puede ser *distancia*.

S,T,U

Sistemas de Bases de Datos Distribuidos Heterogéneos — representa a los entornos con varios sitios o nodos (sean SMBD o SMBDE) con diferentes esquemas e instancias de base de datos y software de base. No se requiere que todos los sitios conozcan de la existencia del resto y es posible que sólo estén dispuestos a compartir parcialmente sus datos.

Sistemas de Información Geográfica (SIG) — sistema con capacidad para recolectar, analizar y presentar información que describe propiedades físicas y lógicas del mundo de la geografía.

Sistema de Manejo de Bases de Datos Espaciales (SMBDE) — SMBD con capacidad para almacenar datos espaciales, así como proveer soporte para operaciones y consultas espaciales.

TDAE — Tipo de datos abstracto espacial.

Tema — la información geoespacial correspondiente a un tópico particular se denomina tema. Un tema tiene un esquema e instancias. Ríos, ciudades y países son ejemplos de tema.

Tesauro — conjunto de términos más formalizados que en un vocabulario. Acuerda una jerarquía de conceptos.

Tipo de fenómeno — unidad que agrupa a los fenómenos de una misma clase. El tipo de fenómeno provee la descripción general para los elementos que pertenecen a él, sus atributos y asociaciones.

V,W,X,Y,Z

Variable cualitativa — tiene asociado un conjunto de valores nominales discretos y cada valor típicamente está vinculado a un término.

Variable cuantitativa – tiene por rango valores numéricos, que pueden ser continuos, como temperatura o precipitación; o discretos como población o cantidad de especies. Las variables cuantitativas pueden ser calculadas. También se pueden agrupar por rangos disjuntos

Vocabulario – es un conjunto de términos relacionados por su significado. Las relaciones básicas que se definen entre términos de un vocabulario incluyen *término más amplio* (o más abarcativo), *término más específico* y *término sinónimo*.

Bibliografía

- [AMT04] ANDRITSOS, P., MILLER, R. J., AND TSAPARAS, P. Information-theoretic tools for mining database structure from large data sets. In *SIGMOD '04: Proceedings of the 2004 ACM SIGMOD international conference on Management of data* (New York, NY, USA, 2004), ACM, pp. 731–742.
- [ASR⁺05] ARPINAR, I. B., SHETH, A., RAMAKRISHNAN, C., USERY, E. L., AZAMI, M., AND KWAN, M. Handbook of geographic information science, 2005.
- [atk03] ATKIS – Objektartenkatalog Basis – DLM http://www.atkis.de/dstinfo/dstinfo2.dst_deckblatt_25?dst_ver=dst. Website, 2003.
- [BCU⁺99] BEDDOE, D., COTTON, P., ULEMAN, R., JOHNSON, S., AND HERRING, J. R. OpenGIS Simple Features Specification For SQL. Tech. rep., OGC, May 1999.
- [BEH⁺03] BERNARD, L., EINSPANIER, U., HAUBROCK, S., HUBNER, S., KUHN, W., LESSING, R., LUTZ, M., AND VISSER, U. Ontologies for intelligent search and semantic translation in spatial data infrastructures, 2003.
- [Bis98] BISHR, Y. Overcoming the semantic and other barriers to gis interoperability. *International Journal of Geographical Information Science* 12, 4 (1998), 299–314.
- [BLD99] BORGES, K. A. V., LAENDER, A. H. F., AND DAVIS, JR., C. A. Spatial data integrity constraints in object oriented geographic data modeling. In *GIS '99: Proceedings of the 7th ACM international symposium on Advances in geographic information systems* (New York, NY, USA, 1999), ACM, pp. 1–6.

- [BM98] BURROUGH, P., AND MCDONNELL, R. *Principles of Geographical Information Systems*. Oxford University Press, 1998.
- [BM03] BILENKO, M., AND MOONEY, R. J. Adaptive duplicate detection using learnable string similarity measures. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, NY, USA, 2003), ACM, pp. 39–48.
- [BO00] BOSSARD, M., J. F., AND OTAHEL, J. Corine land cover technical guide, 2000.
- [BT88] BATTISTA, G. D., AND TAMASSIA, R. Algorithms for plane representations of acyclic digraphs. *Theoretical Computer Science* 61, 2-3 (1988), 175 – 198.
- [Buc09] BUCCELLA, A. *Integración de Sistemas de Información Geográfica*. PhD thesis, Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur. Bahía Blanca. Argentina, 2009.
- [CdA99] CASTANO, S., AND DE ANTONELLIS, V. A schema analysis and reconciliation tool environment for heterogeneous databases. In *IDEAS '99: Proceedings of the 1999 International Symposium on Database Engineering & Applications* (Washington, DC, USA, 1999), IEEE Computer Society, p. 53.
- [CEN07] CEN, I. Cen tc 287 geographic information. inspire, 2007.
- [Che76] CHEN, P. P.-S. The entity-relationship model—toward a unified view of data. *ACM Trans. Database Syst.* 1, 1 (1976), 9–36.
- [Cla86] CLARKE, K. C. Advances in geographic information systems. *Computers, Environment and Urban Systems* 10, 3-4 (1986), 175 – 184.
- [Cli09] The opengis abstract specification. topic 5: Features, 2009.
- [Cod70] CODD, E. F. A relational model of data for large shared data banks. *Commun.* *ACM* 13, 6 (1970), 377–387.
- [Cor00] CORFORD, D. *Geographic Information System*. Computer Science Aston University, <http://www.cs.aston.ac.uk>, October 2000.
- [CYC02] Cyc Knowledge Server. Website, 2002. <http://www.cyc.com/>.

- [dB93] DE BRUIJN, J. Using ontologies - enabling knowledge sharing and reuse on the semantic web. *DERI Technical Report DERI-2003-10-29* (1993).
- [DDH01] DOAN, A., DOMINGOS, P., AND HALEVY, A. Y. Reconciling schemas of disparate data sources: A machine-learning approach. In *SIGMOD Conference* (2001).
- [DGJ98] DI GREGORIO, A., AND JANSEN, L. J. M. Land cover classification system (lccs): Classification concepts and user manual., 1998.
- [DH05] DOAN, A., AND HALEVY, A. Semantic integration research in database community: A brief survey, 2005.
- [DJ06] DEVILLERS, R., AND JEANSOULIN, R. *Fundamentals of Spatial Data Quality (Geographical Information Systems series)*. ISTE, 2006.
- [DLCMT02] DI LUCA, S., CASTRO, S., MARTIG, S., AND TRUTNER, G. AND VITTURINI, M. Interfaces en la visualización de información. *IV Workshop de Investigadores en Ciencias de la Computación (WICC2002)* 1, 1 (2002).
- [DLD⁺04] DHAMANKAR, R., LEE, Y., DOAN, A., HALEVY, A., AND DOMINGOS, P. imap: Discovering complex semantic matches between database schemas, 2004.
- [dSR04] DE SEVILLA RIAZA., T. F. Gml: El lenguaje de marcado extendido (xml) para la ingeniería geográfica. tesis de doctorado en ingeniería geográfica, 2004.
- [EEA01] EEA. Gemet version 2001 (general multilingual environmental thesaurus), 2001.
- [EF91] EGENHOFER, M. J., AND FRANZOSA, R. D. Point-set topological spatial relations. *International Journal of Geographical Information Systems* 5, 2 (1991), 161–174.
- [Ege02] EGENHOFER, M. Toward the semantic geospatial web, 2002.
- [EN00] ELMASRI, R., AND NAVATHE, S. *Fundamentals of Database Systems*. Addison Wesley 3d. Ed., 2000.

- [EN07] ELMASRI, R., AND NAVATHE, S. *Fundamentals of Database Systems (4th Edition)*, 5 ed. Addison Wesley, 2007.
- [ESR08] ESRI. Geologic fault-finding with gis, 2008.
- [FGD07] FGDC, N. National geospatial program standards. the federal geographic data committee, 2007.
- [GHEK00] GOLDBECK, J., HUERTGEN, B., ERNST, S., AND KELCH, L. Lane following combining vision and dgps. *Image and Vision Computing* 18, 5 (2000), 425 – 433.
- [GJM02] GHEZZI, C., JAZAYERI, M., AND MANDRIOLI, D. *Fundamentals of Software Engineering*, 2nd ed. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2002.
- [GMPQ⁺04] GARCIA-MOLINA, H., PAPAKONSTANTINOY, Y., QUASS, D., RAJARAMAN, A., SAGIV, Y., ULLMAN, J., VASSALOS, V., AND WIDOM, J. The tsimmis approach to mediation: Data models and languages. *Journal of Intelligent Information Systems* 8, 2 (2004), 117–132.
- [Goo05] GOOGLE. Google maps web site, 2005.
- [Gru93] GRUBER, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition* (1993).
- [Gua98] GUARINO, N. Formal ontology and information systems. 3–15.
- [Gut95] GUTING, R. H. An introduction to spatial database systems. *GeoInformatica* 3, 3 (1995), 245–268.
- [HS95] HERNANDEZ, M. A., AND STOLFO, S. J. The merge/purge problem for large databases. In *SIGMOD Conference* (1995), pp. 127–138.
- [IIMM99] INFORMATICA, V., INFORMATICA, E. T., MEERSMAN, R., AND MEERSMAN, R. A. Semantic ontology tools in is design, 1999.
- [Ins00] INSTITUTE, T. G. R. Getty vocabularies, 2000.
- [ISO] Geographic information – metadata. international organization for standardization (iso) 19115. Tech. rep., ISO/TC 211.

- [ISO94] ISO/TC 211 Geographic information/Geomatics. Website, 1994. <http://www.isotc211.org/>.
- [ISO02a] Geographic information – core profile of the spatial schema. International Organization for Standardization (ISO) 19137. Tech. rep., ISO/TC 211, 2002.
- [ISO02b] Geographic information – quality principles. International Organization for Standardization (ISO) 19113. Tech. rep., ISO/TC 211, 2002.
- [ISO02c] Geographic information – reference model. International Organization for Standardization (ISO) 19101. Tech. rep., ISO/TC 211, 2002.
- [ISO02d] Geographic information - temporal schema. International Organization for Standardization (ISO) 19108. Tech. rep., ISO/TC 211, 2002.
- [ISO03a] Geographic information – quality evaluation procedures. International Organization for Standardization (ISO) 19114. Tech. rep., ISO/TC 211, 2003.
- [ISO03b] Geographic information - spatial referencing by geographic identifiers. International Organization for Standardization (ISO) 19112. Tech. rep., ISO/TC 211, 2003.
- [ISO03c] Geographic information - spatial schema. International Organization for Standardization (ISO) 19107. Tech. rep., ISO/TC 211, 2003.
- [ISO05a] Geographic information – location-based services – tracking and navigation. International Organization for Standardization (ISO) 19133. Tech. rep., ISO/TC 211, 2005.
- [ISO05b] Geographic information – portrayal. International Organization for Standardization (ISO) 19117. Tech. rep., ISO/TC 211, 2005.
- [ISO05c] Geographic information – rules for application schemas. International Organization for Standardization (ISO) 19109. Tech. rep., ISO/TC 211, 2005.
- [ISO05d] Geographic information - methodology for feature cataloguing. International Organization for Standardization (ISO) 19110. Tech. rep., ISO/TC 211, 2005.

- [ISO05e] Geographic information - web map server interface. International Organization for Standardization (ISO) 19128. Tech. rep., ISO/TC 211, 2005.
- [ISO06] Geographic information – data quality measures. International Organization for Standardization (ISO) 19138. Tech. rep., ISO/TC 211, 2006.
- [ISO07a] Geographic information – data product specifications. International Organization for Standardization (ISO) 19131. Tech. rep., ISO/TC 211, 2007.
- [ISO07b] Geographic information – Geography Markup Language (GML). International Organization for Standardization (ISO) 19136. Tech. rep., ISO/TC 211, 2007.
- [ISO07c] Geographic information – location-based services – multimodal routing and navigation. International Organization for Standardization (ISO) 19134. Tech. rep., ISO/TC 211, 2007.
- [ISO07d] Geographic information – metadata – xml schema implementation . International Organization for Standardization (ISO) 19139. Tech. rep., ISO/TC 211, 2007.
- [ISO07e] Geographic information – spatial referencing by coordinates International Organization for Standardization (ISO) 19111. Tech. rep., ISO/TC 211, 2007.
- [ISO07f] ISO. *Documentation – Guidelines for the establishment and development of monolingual thesauri*. International Organization for Standardization (ISO) 2788:1986. Multiple. Distributed through American National Standards Institute (ANSI), August 2007.
- [KABL01] KARLA A.V. BORGES, C. A. D., AND LAENDER, A. H. Omt-g: An object-oriented data model for geographic applications. *Springer Netherlands GeoInformatica* 5, 3 (2001), 221–260.
- [KH96] KAPLAN, E. D., AND HEGARTY, C. *Understanding GPS Principles and Applications*. Artech-House, Boston London, 1996.
- [Kli08] KLIEN, E. *Semantic Annotation of Geographic Information*. PhD thesis, Institute for Geoinformatics, University of Muenster. Muenster, Germany, 2008.

- [KPS95] KÖSTERS, G., PAGEL, B.-U., AND SIX, H.-W. Object-oriented requirements engineering for gis-applications, 1995.
- [KS96] KASHYAP, V., AND SHETH, A. Semantic and schematic similarities between database objects: a context-based approach. *The VLDB Journal* 5, 4 (1996), 276–304.
- [LB00] LECHNER, W., AND BAUMANN, S. Global navigation satellite systems. *Computers and Electronics in Agriculture* 25, 1-2 (2000), 67 – 85.
- [LBTR04] LAKE, R., BURGGRAF, D., TRNINIC, M., AND RAE, L. *Geography Mark-Up Language: Foundation for the Geo-Web*. John Wiley & Sons, June 2004.
- [LC00] LI, W.-S., AND CLIFTON, C. Semint: a tool for identifying attribute correspondences in heterogeneous databases using neural networks. *Data Knowl. Eng.* 33, 1 (2000), 49–84.
- [LGMR05] LONGLEY, P., GOODCHILD, M., MAGUIRE, D., AND RHIND, D. *Geographic information systems and science*. Wiley, John and Sons, Incorporated, 2005.
- [LRP03] LUTZ, M., RIEDEMAN, C., AND PROBST, F. A classification framework for approaches to achieving semantic interoperability, 2003.
- [LT92] LAURINI, R., AND THOMPSON, D. *Fundamentals of Spatial Information Systems*. Academic Press, 1992.
- [LTM82] LEATHERDALE, D., TIDBURY, G. E., AND MACK, R. *AGROVOC: a multilingual thesaurus of agricultural terminology*. Food and Agriculture Organization of the United Nations. and Commission of the European Communities, english version. ed. Apimondia, by arrangement with the Commission of the European Communities, [S.l.] :, 1982.
- [Lua04] LUACES, M. A. R. *A Generic Architecture for Geographic Information Systems*. PhD thesis, Universidade Da Caruña, 2004.
- [MBR01] MADHAVAN, J., BERNSTEIN, P. A., AND RAHM, E. Generic schema matching with cupid. In *The VLDB Journal* (2001), pp. 49–58.

- [McG03] MCGUINNESS, D. L. Ontologies come of age. In *Spinning the Semantic Web*, D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster, Eds. MIT Press, Cambridge, MA, 2003, ch. 6, pp. 171–196.
- [MDVK] MCCANN, R., DOAN, A., VARADARAJAN, V., AND KRAMNIK, A. Building data integration systems via mass collaboration.
- [ME96] MONGE, A. E., AND ELKAN, C. The field matching problem: Algorithms and applications. In *Knowledge Discovery and Data Mining* (1996), pp. 267–270.
- [ME02] MUMBY, P. J., AND EDWARDS, A. J. Mapping marine environments with ikonos imagery: enhanced spatial resolution can deliver greater thematic accuracy. *Remote Sensing of Environment* 82, 2-3 (2002), 248 – 257.
- [MFW90] MICHAEL F. WORBOYS, HILARY M. HEARNshaw, D. J. M. Object-oriented data modelling for spatial databases. *International Journal of Geographical Information Science* 4, 4 (1990), 369 – 383.
- [MGR99] MAGUIRE, D., GOODCHILD, M., AND RHIND, D. *Geographic Information Systems*. Longman Scientific and Technical, 1999.
- [Mit99a] MITCHELL, A. *The ESRI Guide to GIS Analysis Volume 1: Geographic Patterns and Relationships*. ESRI, 1999.
- [Mit99b] MITCHELL, A. *GIS Analysis. Geographic Patterns and Relationships*. Environmental Systems Research Institute, Inc, 1999.
- [MKS⁺05] MCCANN, R., KRAMNIK, A., SHEN, W., VARADARAJAN, V., SOBULO, O., AND DOAN, A. Integrating data from disparate sources: a mass collaboration approach, 2005.
- [Mol95] MOLENAAR, M. Spatial concepts as implemented in gis, 1995.
- [MSv⁺05] MENKEN, M., STUCKENSCHMIDT, H., (VRIJE, H. W., STOILOS, G., TZOUVARAS, V., HELLAS, T., FENSEL, C. P. D., AND LEGER, C. P. A. Knowledge web consortium, 2005.

- [MV10] MERCEDES VITTURINI, P. F. Integración de sistemas de información geográfica. *XII Workshop de Investigadores en Ciencias de la Computación (WICC2010)* 1, 1 (2010).
- [MZ98] MILO, T., AND ZOHAR, S. Using schema matching to simplify heterogeneous data translation. In *Proc. 24th Int. Conf. Very Large Data Bases, VLDB* (24–27 1998), pp. 122–133.
- [Nac06] NACIONAL, I. G. Prosiga - proyecto sistema de información geográfica nacional de la república argentina. Website, 2006. <http://www.sig.gov.ar/>.
- [OGC94] Open Geospatial Consortium, Inc. (OGC). Website, 1994. <http://www.opengeospatial.org/>.
- [OGC99] The OpenGIS Abstract Specification. Topic 3: Locational Geometry Structures, 1999.
- [OGC05] The OGC Abstract Specification. Open Geospatial Consortium (OGC). Website, 2005. <http://www.opengeospatial.org/>.
- [OGC06] OpenGIS Web Map Server Implementation Specification. Open Geospatial Consortium (OGC). Website, 2006. <http://www.opengeospatial.org/>.
- [OGC07] OpenGIS Geography Markup Language (GML). Open Geospatial Consortium (OGC) . Website, 2007. <http://www.opengeospatial.org/>.
- [OGC08] OGC Reference Model (ORM). Open Geospatial Consortium (OGC). Website, 2008. <http://www.opengeospatial.org/>.
- [oJ09] OF JUSTICE, U. D. Maps: Geospatial tools. Website, 2009. <http://www.ojp.usdoj.gov/nij/maps/tools.htm>.
- [oM09] OF MEDICINE, N. L. Toxmap: Environmental health e-maps. Website, 2009. <http://toxmap.nlm.nih.gov/toxmap/combo/applyLayerOptions.do>.
- [Ope02] OPEN GIS CONSORTIUM INC. Web feature service implementation specification, 2002.

- [OPM97] OLIVEIRA, J. L. D., PIRES, F., AND MEDEIROS, C. B. An environment for modeling and design of geographic applications. *Geoinformatica 1*, 1 (1997), 29–58.
- [ORA] ORACLE. Oracle spatial and oracle locator: Location features for oracle database 11g.
- [Pas08] PASCUAL, A. R. Revista interancional de ciencias de la tierra. instituto geográfico nacional francisco javier ariza lópez. universidad de jaén. enero 2008, 2008.
- [Pau05] The OpenGIS Abstract Specification. Topic 2: Spatial referencing by coordinates, 2005.
- [POS10] POSTGRES. Postgis 1.5.2, 2010.
- [PSZ⁺98] PARENT, C., SPACCAPIETRA, S., ZIMANYI, E., DONINI, P., PLAZANET, C., AND VANGENOT, C. Modeling spatial data in the mads conceptual model. In *in Proceedings of the 8th International Symposium on Spatial Data Handling* (1998), pp. 138–150.
- [PSZ99] PARENT, C., SPACCAPIETRA, S., AND ZIMÁNYI, E. Spatio-temporal conceptual models: data structures + space + time. In *GIS '99: Proceedings of the 7th ACM international symposium on Advances in geographic information systems* (New York, NY, USA, 1999), ACM, pp. 26–33.
- [PSZ06a] PARENT, C., SPACCAPIETRA, S., AND ZIMÁNYI, E. *Conceptual Modeling for Traditional and Spatio-Temporal Applications: The MADS Approach*. Springer-Verlag, 2006.
- [PSZ06b] PARENT, C., SPACCAPIETRA, S., AND ZIMÁNYI, E. The mads data model: Concepts to understand the structure of your spatial and temporal data. *Journal of Informative Modelling for the Architectural Heritage*, 1 (July 2006), 51–64. Proceedings of the MIA 2006 International Workshop.
- [PU98] PALOPOLI, D. S. L., AND URSINO, D. Semi-automatic, semantic discovery of properties from database schemes. In *IDEAS '98: Proceedings of the 1998 International Symposium on Database Engineering & Applications* (Washington, DC, USA, 1998), IEEE Computer Society, p. 244.

- [Raj10] RAJESH, M. Geographic information system and mapping, 2010.
- [RBP⁺91] RUMBAUGH, J., BLAHA, M., PREMERLANI, W., EDDY, F., AND LORENSEN, W. *Object-oriented modeling and design*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1991.
- [RJB98] RUMBAUGH, J., JACOBSON, I., AND BOOCH, G. *The Unified Modeling Language Reference Manual*. Addison-Wesley, 1998.
- [RJB04] RUMBAUGH, J., JACOBSON, I., AND BOOCH, G. *Unified Modeling Language Reference Manual, The (2nd Edition)*. Pearson Higher Education, 2004.
- [RL02] REED, S. L., AND LENAT, D. B. Mapping ontologies into cyc. Tech. rep., Cycorp, Inc., 2002.
- [RSV02] RIGAUX, P., SCHOLL, M., AND VOISARD, A. *Spatial databases with application to GIS*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002.
- [SB02] SARAWAGI, S., AND BHAMIDIPATY, A. Interactive deduplication using active learning. In *KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, NY, USA, 2002), ACM, pp. 269–278.
- [SC03] SHEKHAR, S., AND CHAWLA, S. *Spacial Databases: A Tour*. Prentice Hall, 2003.
- [SCG⁺97] SHEKHAR, S., COYLE, M., GOYAL, B., LIU, D.-R., AND SARKAR, S. Data models in geographic information systems. *Commun. ACM* 40, 4 (1997), 103–111.
- [SHW⁺04] STUCKENSCHMIDT, H., HARMELEN, F. V., WAARD, A. D., SCERRI, T., BHOGAL, R., BUEL, J. V., CROWLESMITH, I., FLUIT, C., KAMPMAN, A., BROEKSTRA, J., AND MULLIGEN, E. V. Exploring large document repositories with rdf technology: The dope project, 2004.
- [SKS10] SILBERSCHATZ, A., KORTH, H., AND SUDERSHAN, S. *Database System Concepts*, 6th ed. McGraw-Hill, Inc. New York, NY, USA, 2010.

- [SMB⁺97] SHEKHAR, S., M.COYLE, B.GOYAL, LIU, D., AND S.SARKAR. Experiences with data models in geographic information systems. *Communications of the ACM* 40, 4 (1997).
- [std85] Documentation – Guidelines for the Establishment and Development of Multilingual Thesauri International Organization for Standardization (ISO) 5964:1985=BS 6723:1985. Tech. rep., ISO, 1985.
- [SWB⁺] SALOKHE, G., WEINHEIMER, J., BOVO, M. G., AGRIMI, M., KEIZER, J., AND KATZ, S. Food and agriculture organization of the united nations, italy.
- [Ter06] TERRASA, A. N. Semantic integration of thematic geographic information in a multimedia context. *Doctorate in Computer Science and Communication Department of Technology. Universitat Pompeu Fabra* (2006).
- [Ull88] ULLMAN, D. *Principles of Database and Knowledge-base Systems*. Computer Science Press, Rockville, 1988.
- [UNS] United Nations Standard Products and Services Code (UNSPSC). Website. <http://www.unspsc.org>.
- [Usc98] USCHOLD, M. Knowledge level modelling: concepts and terminology. *Knowl. Eng. Rev.* 13, 1 (1998), 5–29.
- [USG07] Earthquake Hazards Program. Website, 2007. <http://earthquake.usgs.gov/earthquakes/>.
- [VCM05] VITTURINI, M., CASTRO, S., AND MARTIG, S. Modelos de datos espaciales. *VII Workshop de Investigadores en Ciencias de la Computación (WICC2005)* 1, 1 (2005).
- [VF08] VITTURINI, M., AND FILLOTTRANI, P. Interoperabilidad y estándares de datos para información geográfica. *XIV Congreso Argentino de Ciencias de la Computación (CACIC 2008)* 1, 1 (2008).
- [VFC03] VITTURINI, M., FILLOTTRANI, P., AND CASTRO, S. Modelos de datos para datos espaciales. *V Workshop de Investigadores en Ciencias de la Computación (WICC 2003)* 1, 1 (2003).

- [WD04] WORBOYS, M. F., AND DUCKHAM, M. *GIS: A Computing Perspective*. CRC Press, 2004.
- [wms04] Web Map Service Interface. Open GeoSpatial Consortium (OGC). Tech. rep., OGC, January 2004.
- [WOR98] Wordnet an electronic lexical database, 1998. <http://wordnet.princeton.edu/wordnet/>.
- [WVV⁺01] WACHE, H., VÖGELE, T., VISSER, U., STUCKENSCHMIDT, H., SCHUSTER, G., NEUMANN, H., AND HÜBNER, S. Ontology-based integration of information - a survey of existing approaches. pp. 108–117.
- [XE03] XU, L., AND EMBLEY, D. W. Discovering direct and indirect matches for schema elements. *Database Systems for Advanced Applications, International Conference on D* (2003), 39.
- [Zei99] ZEILER, M. *Modeling our World. The ESRI Guide to Geodatabase Design*. Environmental Systems Research Institute, Inc, 1999.