

Calidad de bases de datos geográficos digitales

Lourdes Castillo Villanueva

Antonio Iturbe Posadas

loucasti@correo.uqroo.mx, aiturbe@correo.uqroo.mx

Departamento de Ciencias, Universidad de Quintana Roo
Boulevard Bahía s/n esq. Ignacio Comonfort, Col. del Bosque
Chetumal, Quintana Roo, México C.P. 77019

RESUMEN

La calidad en las bases de datos geográficas digitales (BDGD) juega un papel por demás importante en la realización de proyectos de sistemas de información geográfica; en función de la calidad de los datos será la confiabilidad de los resultados derivados de procesos de manejo y análisis espacial, que por lo general, conllevan a tomas de decisión que impactan en el territorio. El presente trabajo describe el fundamento conceptual de los elementos que determinan la calidad de los datos geográficos; con base en lo anterior, se procedió al análisis de diversas bases de datos geográficas digitales en formato vectorial SIG disponibles en internet con cobertura geográfica al territorio de Quintana Roo y posteriormente determinar de manera general la calidad de los mismos. Los resultados indican que la mayor parte de las BDGD tienen algún inconveniente con base en los elementos que definen la calidad, lo que resulta en importantes implicaciones para la realización de análisis espaciales y generación de productos SIG para coadyuvar a la toma de decisiones territorial. Por lo anterior, es necesario que aquellos usuarios que usen BDGD provenientes de Internet para realizar análisis SIG, apliquen un proceso de evaluación y determinen si los datos cumplen con los requerimientos que dictan los objetivos del proyecto SIG, más aún que, se reitera, de la calidad de los datos depende la calidad de los resultados.

Palabras clave: sistemas de información geográfica – bases de datos geográficas digitales – ordenamiento territorial – calidad.

INTRODUCCIÓN

Los Sistemas de Información Geográfica (SIG) se constituyen hoy en día

como importantes herramientas para la realización de procesos de manejo, análisis y modelación de datos geográficos para la resolución de problemas

Bases de datos geográficos

territoriales. Los SIG se encuentran actualmente en un estado de pleno desarrollo y evolución tecnológica y han llegado a ser prácticamente indispensables en aquellas organizaciones que realizan actividades de manejo de recursos naturales, transporte, seguridad, educación, desarrollo urbano, catastro, gestión inteligente de redes de telecomunicaciones, de energía eléctrica, etc., y en todas aquellas actividades donde sea necesaria la resolución de problemas con un carácter territorial.

Los SIG, vistos desde una perspectiva de soluciones para la resolución de problemas territoriales, se integran a partir de cinco componentes que a saber son *hardware*, *software*, personal, procedimientos y datos. Al parecer, es un común denominador en la realización de proyectos SIG prestar mayor atención a los elementos de *hardware* y *software*, quizá por el uso extendido de que un SIG es fundamentalmente una solución informática. Sin embargo, *hardware* y *software* representan precisamente el componente informático y de menor importancia; si bien los *software* de SIG por lo general pueden ser costosos, resalta el hecho que existen soluciones informáticas SIG con excelentes capacidades para el manejo, análisis y modelación sin costo económico alguno, denominados *open source*.

El éxito de un proyecto SIG depende en gran medida de la calidad de los elementos referentes al personal capacitado en los diversos tópicos de especialización que demanda el tipo de proyecto SIG, la forma en como son aplicados los procedimientos explíci-

tamente definidos para el manejo, análisis y modelación de los datos y, por supuesto, los datos geográficos.

En el desarrollo de un proyecto SIG, la creación de la BDGD significa entre 60% y 80% del costo total del proyecto (Bernhardsen, 1999). Los datos deben ser vistos con suma importancia no sólo a razón del tiempo y dinero que conlleva su creación, sino también porque en función de la calidad de los datos será la calidad de la información resultante que sirve para la realización de tomas de decisión que impactan en la mayoría de las ocasiones en el mundo real (Aronoff, 1993).

En nuestro país, es notable apreciar un número cada vez mayor de proyectos SIG e incluso de la generación de proyectos encaminados exclusivamente a la generación de bancos de datos geográficos digitales para su análisis en sistemas de información geográfica. Sin embargo, en muchas ocasiones tienen una calidad cuestionable que hace incluso imposible un uso adecuado y un riesgo considerable para la toma de decisiones.

En el presente trabajo se realiza una revisión conceptual de los elementos que describen y determinan la calidad de las bases de datos geográficas digitales (BDGD); posteriormente, fueron recopilados diversos bancos de datos en formato vectorial SIG con cubrimiento total o parcial al estado de Quintana Roo y se llevó al cabo una evaluación de la calidad de los mismos. Al final, se generan algunas conclusiones en relación con la calidad encontrada en las BDGD.

LA CALIDAD EN BASES DE DATOS GEOGRÁFICAS DIGITALES

Una Base de Datos Geográfica Digital es un conjunto de capas de datos geográficos que se encuentran en formato digital. Cada capa de datos representa elementos de la superficie terrestre delimitadas en tiempo, territorio y descripciones cuantitativas y cualitativas a partir del objetivo del proyecto SIG. Es a partir de éstas o bien, un conjunto de datos geográficos, que se aplican diversos procesos de manejo, análisis y modelación para resolver un problema determinado. Por ejemplo, las carreteras de Quintana Roo pueden servir como datos de soporte para la obtención de rutas óptimas que permiten disminuir costos y tiempos en la distribución de productos perecederos en las localidades de este estado.

Para estar en capacidad de aplicar procesos de manejo y análisis sobre las BDGD se deben cumplir una serie de requisitos. La base de datos debe ser construida bajo un determinado modelo de datos geográfico SIG (ej. vector), una estructura particular (ej. estructura georrelacional) y una serie de consideraciones geográficas y cartográficas. Estos elementos son los que definen cómo debe ser construida una base de datos geográfica digital.

Cuando las BDGD no se construyen bajo premisas y especificaciones que dicta la lógica geográfica, la tecnología de sistemas de información geográfica y ciertas especificaciones cartográficas, derivan en la imposibilidad de realizar análisis espaciales básicos, complejos y procesos de modelación.

Además, se tiene el riesgo de obtener resultados imprecisos y erróneos y los problemas potenciales cuando esos resultados son empleados para resolver un problema territorial; conviene en este punto recordar la frase popular *garbage in-garbage out*.¹

En lo que respecta a la construcción de bases de datos geográficas digitales vectoriales,² Zeiler (1999) y Newel *et al.* (1999); señalan que existen tres estructuras: *i)* CAD o de diseño asistido por computadora. Los elementos geométricos relativos a puntos, líneas y polígonos no guardan reglas topológicas y su creación obedece más a cómo se dibujarán de forma manual, *ii)* georrelacional, en la cual los puntos, líneas y polígonos guardan una lógica geográfica y se tiene una base de datos de atributos asociados interrelacionada y *iii)* geodatabase, que tiene como característica recrear con la mayor semejanza posible la realidad y dinámica geográfica; los elementos geométricos y sus atributos asociados están almacenados íntegramente en un sistema manejador de bases de datos.

El modelo georrelacional es el más extendido y usado actualmente; su concepto se basa en dos componentes bien diferenciados: la parte geométrica o que define la forma y posición de los ele-

¹ Basura entra-basura sale. Esta frase es muy popular en el ámbito de los sistemas de información. Si los datos de entrada son sujetos a análisis y su calidad es pobre, los resultados de igual forma serán pobres en calidad.

² Los datos vectoriales son aquellos que emplean puntos, líneas y polígonos para representar elementos de la realidad geográfica. Por ejemplo, una carretera a escala a:1250 000 es una línea.

Bases de datos geográficos

mentos geográficos por medio del uso de puntos, líneas y polígonos ubicados en un contexto cartográfico y el componente de atributos asociados o descriptivo; ambos se encuentran ligados por apuntadores o identificadores en común y corresponden a un periodo de tiempo definido. Desde que este modelo de datos es el que impera en el medio de la cartografía automatizada y los sistemas de información geográfica, así como el hecho de que la mayoría de los datos SIG disponibles hoy en día están contruidos bajo este modelo, que se toma como punto de partida para la evaluación y análisis de la calidad de las bases de datos geográficas digitales.

La estructura georrelacional dicta requisitos específicos para la construcción de capas de datos geográficos (Aronoff, 1993; Bernhardsen, 1999; Domínguez, *et al.*, 1998). Además, es importante recalcar el hecho de que la construcción de una BDGD para su manejo y análisis SIG, debe recrear las leyes y la lógica geográfica existente en la realidad y respetar las reglas de conectividad y compartición entre los objetos geográficos. Por ejemplo, cuando se digitaliza un río a una escala del orden 1:250 000,³ este debe conectar tal cual en la realidad con otro río; además, se tiene que el río puede constituir un límite compartido entre los distintos tipos de vegetación que hay al lado del fluvio a lo largo de su trayecto. Un río, debe coincidir con la topografía en el sentido de que sólo debe fluir de las partes altas a las partes bajas.

³ Esta escala significa que una unidad de medida en el mapa equivale a 250 mil unidades de medida en la superficie terrestre.

Estas consideraciones forman parte del concepto de la *calidad de los datos*. Son numerosos los autores que han abordado este tema y, particularmente, Chrisman (1997) señala que la calidad debe ser considerada desde las etapas iniciales en la construcción de un proyecto SIG. La calidad de los datos geográficos es un tema tan importante, que la escuela americana de SIG le ha dedicado especial énfasis y forma parte actualmente de las estrategias y compromisos del Instituto Nacional de Estándares y Tecnología (*National Institute of Standards and Technology* de ese país (<http://www.nist.gov/>)).

Con base en la fuente anterior, se reconocen cinco elementos que deben ser considerados en la construcción de una base de datos geográfica digital, mismos que regulan y garantizan la calidad del juego de datos:

- *Linaje*. Concepto que permite conocer a profundidad características de los datos fuente y de los cuales se derivará el nuevo juego de datos. El linaje permite conocer cuáles son las operaciones y transformaciones que se aplicaron y que repercuten tanto en el componente geométrico como en los atributos asociados. Por ejemplo, el conocimiento del linaje permite saber si el origen de un dato es de una adecuada calidad o bien, si los datos fueron adecuadamente procesados y son válidos para derivar resultados confiables. La forma como se puede evaluar el linaje es a través de

los metadatos que deben formar parte de la BDGD.

- *Exactitud posicional.* Aplica al componente geométrico de una BDGD. Permite evaluar las medidas y posición de los objetos respecto a la realidad geográfica. La exactitud posicional permite conocer cuál es la distancia promedio en metros u otra unidad de los puntos, líneas o polígonos que describen elementos geográficos en relación con la realidad geográfica; por lo general, se llega a la definición de una cifra que resulta del cálculo del error medio cuadrático (RMS) entre las posiciones de los elementos del mapa contenidos en el juego de datos y puntos de referencia con alto grado de exactitud. Es importante mencionar, que existen criterios cartográficos generales que determinan valores máximos. Por ejemplo, la exactitud posicional de los objetos en el mapa no debe sobrepasar un error de 0.05mm en equivalencia con las unidades de escala de los datos fuente.
- *Exactitud de los atributos.* Elemento relativo a la correspondencia entre los atributos descriptivos asociados a la geometría y la realidad geográfica. La evaluación puede ser llevada a cabo de muy diversas formas, desde aplicar una revisión básica que permita apreciar si el juego de datos tiene lógica (ej. las lagunas deben tener atributos correspondientes a cuerpos de agua) hasta revisiones hechas en

campo aplicando métodos de muestreo y determinación de la correspondencia en porcentaje.

- *Consistencia lógica.* Este concepto hace referencia a las relaciones topológicas que debe tener el conjunto de datos, no sólo al interior de una capa, sino entre todas las capas de datos a las que temáticamente tienen que respetarse las relaciones de contigüidad, adyacencia y contenido. Este elemento de la calidad permite dar congruencia y una lógica geográfica a los datos, permitiendo por ejemplo, recrear los flujos de los ríos (líneas interconectadas) y que se integran de forma adecuada con el relieve (el comportamiento de la sinuosidad de las curvas de nivel corresponde al paso de los fluvios). La evaluación por lo general se realiza de forma visual, aplicando funciones de análisis espacial como rutas óptimas sobre un conjunto de datos de carreteras para ver si existe conectividad y a través de las funciones de revisión topológica presentes en los diferentes programas de cómputo SIG.
- *Conclusión de los datos.* Elemento que permite conocer cómo es la calidad de los datos en cuanto a la omisión o completitud de rasgos geométricos y atributos asociados. La revisión con las fuentes de datos originales es básica y el objetivo es ver si los datos son completos en su totalidad.

Bases de datos geográficos

Existe un problema significativo en la creación de las BDGD y es que, prácticamente la mayoría de las organizaciones que generan datos para su manejo y análisis SIG en México no realizan los respectivos metadatos. Los metadatos, describen a detalle justamente los datos geográficos digitales y permiten al usuario saber a detalle cuál es la calidad de estos; el sistema de proyección geográfica empleado, y las restricciones legales para el uso de los datos, entre otros aspectos. Si bien existen iniciativas importantes para la generación de metadatos a partir de estándares, generación de programas para la gestión de metadatos y portales en internet para su difusión como es el caso de la Red Sigpy (www.sigpy.uqroo.mx), faltará tiempo para que sea común encontrar datos geográficos que integren como parte de los mismos los correspondientes metadatos.

EVALUACIÓN DE LA CALIDAD DE BASES DE DATOS GEOGRÁFICAS DIGITALES

Se llevó a cabo una búsqueda en internet de BDGD con cobertura territorial al estado de Quintana Roo. Como resultado, se tiene la existencia de varias bases de datos destacando que la mayoría comprende variables de tipo topográfico (carreteras, localidades, hidrología superficial, infraestructura, localidades, entre otras); las escalas de estas fuentes varían desde 1:1 000 000 que cubren todo el territorio estatal hasta BDGD a escala 1:20 000 y cubren porciones de la entidad.



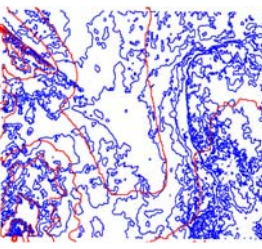

Posteriormente, se realizó un proceso de evaluación con base en los conceptos de *linaje*, *exactitud posicional* y de *atributos*, *consistencia lógica* y *completitud*. Este proceso de evaluación de la calidad de los datos geográficos debe concebirse como un proceso importante, dado que permite el valorar el grado de confiabilidad que poseen los datos y conocer las limitaciones y alcances de los datos para la realización de análisis territoriales.

Desde que la mayor parte de los datos geográficos digitales carecen de metadatos, esta evaluación reviste de importancia para conocer hasta dónde los usuarios pueden obtener resultados aceptables de los mismos.

Como resultado de la evaluación de la calidad de los datos topográficos digitales obtenidos, se mencionan a continuación una serie de problemáticas que de forma repetitiva se encuentran y que repercuten en el uso potencial para la realización de análisis espacial en sistemas de información geográfica (tabla 1).


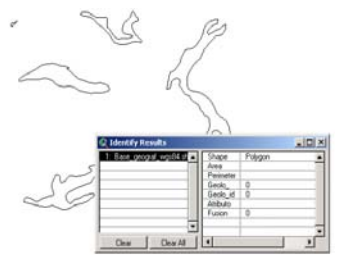
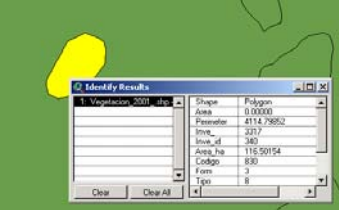
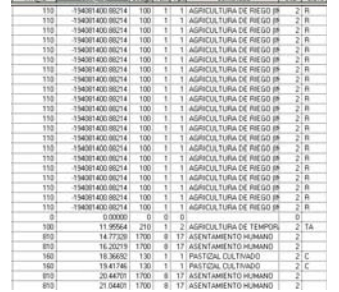
La mayor parte de los datos analizados presentan problemas en uno o varios elementos que determinan la calidad de los datos geográficos. Se menciona una vez más el hecho de que los datos no poseen metadatos, por lo que se desconocen muchos aspectos esenciales para un adecuado uso de los datos por parte de usuarios SIG. Los resultados ponen en claro que los datos no cuentan con la calidad adecuada para ser manejados y analizados desde una perspectiva de sistema de información geográfica y, en gran medida, para efectos de tomas de decisión de tipo territorial.

Tabla 1. Problemas de calidad presentes en la BDGD analizadas

Ejemplo gráfico del problema	Tipo y descripción	Limitaciones
	<ul style="list-style-type: none"> • <i>Exactitud posicional y consistencia lógica inadecuados.</i> Curvas de nivel sin una forma congruente con la realidad geográfica. No existen curvas de nivel en forma de cruz. No pueden representarse curvas de nivel como líneas aisladas y líneas perpendiculares que intersectan a los polígonos que representan líneas con igual altitud. 	<ul style="list-style-type: none"> • Imposibilidad de hacer mapas hipsométricos. • Obtención imprecisa de áreas y perímetros. • Generación de representaciones cartográficas inadecuadas. • Cualquier análisis derivado como modelos digitales del terreno, mapas de pendientes, aspecto, cuencas, etc., resultarán inadecuados para tomas de decisión.
	<ul style="list-style-type: none"> • <i>Exactitud posicional inadecuada.</i> Curvas de nivel digitalizadas de forma deficiente, que no corresponden a la realidad geográfica, dado que no hay para el caso de Quintana Roo curvas de nivel con segmentos cuadriculados. Imagen irreal del terreno. 	<ul style="list-style-type: none"> • Obtención imprecisa de áreas y perímetros. • Generación de representaciones cartográficas inadecuadas. • Imposibilidad de realizar modelos digitales del terreno y otros productos derivados.
	<ul style="list-style-type: none"> • <i>Exactitud posicional inadecuada.</i> Las líneas rojas corresponden a curvas de nivel precisas, mismas que no corresponden con las curvas de nivel azules. Existen diferencias altitudinales de hasta 100mts que para el caso del contexto geomorfológico del estado de Quintana Roo es un error excesivo. 	<ul style="list-style-type: none"> • Imposibilidad de realizar mediciones confiables, derivar mapas precisos de pendientes, aspecto, cuencas, entre otros. • Inadecuado para tomas de decisión territorial donde participe como dato relevante el relieve.
	<ul style="list-style-type: none"> • <i>Exactitud posicional y completitud inadecuada.</i> Las líneas en color azul corresponden a curvas de nivel precisas. Las líneas en color rojo son curvas de nivel de datos geográficos digitales en internet. El resultado de esta sobreposición gráfica indica que existen problemas de posición (hasta 300mts) y una falta considerable de elementos o bien, el juego de datos es incompleto. 	<ul style="list-style-type: none"> • Imposibilidad de realizar mediciones confiables, derivar mapas precisos de pendientes, aspecto, cuencas, entre otros. • Inadecuado para tomas de decisión territorial donde participe como dato relevante el relieve.

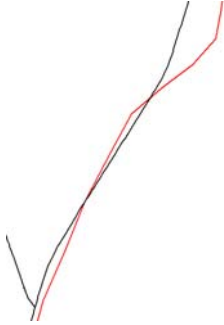
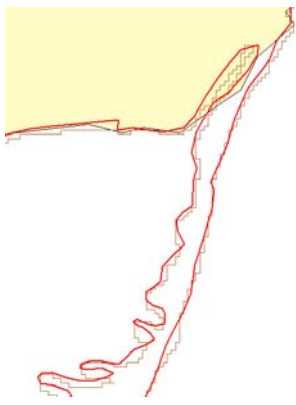
Continúa...

Tabla 1 (continuación)

Ejemplo gráfico del problema	Tipo y descripción	Limitaciones
	<ul style="list-style-type: none"> Exactitud de atributos inadecuada. La base de datos correspondiente a atributos asociados de un mapa de geología presenta problemas relativos a una inadecuada representación de unidades; desde que no se brindan metadatos el usuario está imposibilitado en saber el significado de cada campo, el linaje, unidades de medida, entre otros aspectos. 	<ul style="list-style-type: none"> El usuario no puede explotar de forma adecuada los datos geográficos. La generación de resultados no son confiables, dado que se desconoce el significado de los campos y las unidades de valores numéricos.
	<ul style="list-style-type: none"> Exactitud posicional, exactitud de atributos y consistencia lógica inadecuados. Para el caso de datos geográficos correspondientes a cuerpos de agua, se tienen desplazamientos de hasta 250mts; no existen atributos asociados para estos elementos geométricos y la representación geométrica es inadecuada, dado que en lugar de líneas deben ser polígonos los elementos geométricos a utilizar para representar lagunas. 	<ul style="list-style-type: none"> No es posible obtener áreas de cuerpos de agua. Existen problemas de desplazamiento por lo que carreteras, localidades y otros elementos de otras capas de datos se encuentran "dentro del agua" con base en los elementos de cuerpos de agua contenidos en esta base de datos geográfica digital.
	<ul style="list-style-type: none"> Exactitud de atributos inadecuada. Existen elementos poligonales que corresponden a tipos de vegetación con área igual a 0.0, lo que supone datos inadecuados. Se desconoce el significado de campos. 	<ul style="list-style-type: none"> Desde que este problema aplica a un 10% de todos los datos, los usuarios evidentemente tienen problemas para efectuar tareas de cuantificación.
	<ul style="list-style-type: none"> Exactitud de atributos inadecuada. Existen numerosos polígonos con áreas negativas y valores de 0. Nótese que muchos registros (polígonos) tienen áreas iguales para polígonos de diferentes tamaños y existen polígonos sin atributos. Se desconoce el significado de campos. 	<ul style="list-style-type: none"> Imposibilidad de aplicar métodos de análisis y derivar tomas de decisión confiables.

Continúa...

Tabla 1 (continuación)

<i>Ejemplo gráfico del problema</i>	<i>Tipo y descripción</i>	<i>Limitaciones</i>
	<ul style="list-style-type: none"> • <i>Consistencia lógica inadecuada.</i> No existe una integridad topológica entre coberturas; por ejemplo, se tiene que segmentos de carretera (líneas en color rojo) que se encuentran mar adentro en más de 300mts con respecto a la línea de costa (línea en color negro). 	<ul style="list-style-type: none"> • Imposibilidad de aplicar métodos de análisis y derivar tomas de decisión confiables como por ejemplo, responder cuál es la infraestructura carretera vulnerable a inundaciones.
	<ul style="list-style-type: none"> • <i>Exactitud posicional, consistencia lógica y completitud inadecuados.</i> No existe integridad topológica entre coberturas. Ningún mapa coincide con algún otro. Obsérvese que el mapa de municipios (línea negra) no corresponde con el mapa de referencia geográfica (línea roja) y con las curvas de nivel (línea café). Estos errores van desde 200 hasta 900 metros lo que significa que a escala 1:250,000 de impresión estos errores son más que evidentes. Obsérvese que al mapa de municipios le hace falta una prolongación considerable de tierra. 	<ul style="list-style-type: none"> • Imposibilidad de aplicar métodos de análisis y derivar tomas de decisión confiables.

Es así, que la mayor parte de los datos geográficos digitales existentes que se encuentran disponibles en Internet con cobertura total o parcial al estado de Quintana Roo, presentan una serie de problemáticas que limitan la capacidad de realizar análisis espaciales y obtener resultados adecuados. Rhind (1998), señala que para el caso particular de los datos topográficos, en la actualidad deben ser generados con la mayor calidad posible no sólo por

los costos que esto implica, sino por la utilidad de los mismos en la toma de decisiones.

Es un hecho que los datos topográficos sirven de marco de referencia para una gran cantidad de proyectos SIG y su contenido es intensamente aplicado para la resolución de variados tipos de problemas territoriales.

La generación de bases de datos geográficas digitales demanda la adopción de estándares y la aplicación de

Bases de datos geográficos

adecuados procedimientos que garanticen al final una BDGD con una calidad adecuada y sobre todo, con las características y propiedades que permitan derivar información útil y confiable para la toma de decisiones territorial.

CONCLUSIONES

Si bien existen bancos de datos geográficos digitales en formato SIG disponibles en Internet y con cobertura al estado de Quintana Roo, estos presentan problemas en uno o varios elementos que determinan la calidad de los mismos. En virtud de la calidad de algunos juegos de datos, se tiene que su uso no es recomendable, incluso con fines de representaciones cartográficas impresas.

Se recomienda a aquellos usuarios de datos geográficos digitales, que realicen una revisión del conjunto de datos geográfico digital y que determinen si los datos son de su utilidad; esta revisión es indispensable ya que no se encontraron bases de datos geográficas con sus respectivos metadatos y por ende, el detalle de construcción, calidad y otra serie de datos importantes para un adecuado uso de los datos en sistemas de información geográfica.

Existen iniciativas con carácter peninsular (www.sigpy.uqroo.mx) para la adopción de estándares, tanto en la creación de bases de datos geográficas digitales vectoriales como en la generación de metadatos. En este sentido, se recomienda su consulta y en su caso, adopción. Por lo planteado

en este artículo, se puede señalar que es necesario contar con estrategias de formación de personal altamente capacitado en materia de sistemas de información geográfica. Ya que sólo de esta forma se podrá contar con datos geográficos útiles en la toma de decisiones de tipo territorial.

BIBLIOGRAFÍA

- Aronoff, Stan (1993), *Geographic Information Systems: A management perspective*, WDL Publications, Canadá, 294 pp.
- Bernhardsen, Tor (1999), *Geographic Information Systems. An Introduction*, 2a. ed., John Wiley & Sons, Inc. Noruega, 372 pp.
- Chrisman, Nicholas (1997), *Exploring Geographic Information Systems*, John & Wiley Sons, Inc. University of Washington, EE.UU., 298 pp.
- Domínguez Tejeda, E., A. Iturbe Posadas, y F. Reyna Sáenz (1998), "Sistema de Información Geográfica para el inventario y análisis de los recursos bióticos del Estado de México", tesis profesional, UAEM, 230 pp.
- Newel, Richard y Tom Sancha (1999), "The Difference Between CAD and GIS", en www.smallworld.co.uk/technology/tech/tech_cadgis.asp
- Rhind, David (1998), "National Mapping as a Business-Like Enterprise", en D. R. Fraser Taylor, *Policy Issues in Modern Cartography*, Pergamon, Ottawa, pp. 1-18.
- Zeiler, Michael (1999), *Modelling Our World*, ESRI Press, California, 232 pp.