

Un Modelo para la Revisión de la Calidad de la Información Geográfica Investigación

Dr. Enrique Luna Ramírez
Instituto Tecnológico El Llano Ags.
El Llano, Ags., México
elunaram@hotmail.com

M. C. J. Trinidad Carrillo Mtz.
INEGI – Edificio Sede
Aguascalientes, Ags., México
trinidad.carrillo@inegi.gob.mx

Dr. Francisco J. Álvarez Rodríguez
Universidad Autónoma de Ags.
Aguascalientes, Ags., México
fjalvar@correo.uaa.mx

Dr. Marco Villalobos Abarca
Universidad de Tarapacá
Arica, Chile
mvillalo@uta.cl

C. Dr. Edgar A. Taya Acosta
U. Nal. Jorge Basadre Grohmann
Tacna, Perú
edgartaya76@hotmail.com

Resumen

Existen diversas herramientas de software que pueden ser utilizadas como apoyo durante el proceso de toma de decisiones en una organización. Los sistemas de gestión de información geográfica, los sistemas que ofrecen servicios OLAP (On Line Analytical Processing) y, en general, los sistemas gestores de bases de datos son ejemplos representativos de este tipo de herramientas. Particularmente, los sistemas de información geográfica han cobrado una importancia significativa en dicho proceso, ya que cerca del 80% de la información tratada por las organizaciones en el mundo tiene en alguna medida relación con información de tipo geográfico [1], lo que hace evidente la necesidad de tener un control adecuado de la calidad, exactitud y actualidad de este tipo de información. Así, teniendo como motivación la importancia que tiene la información geográfica durante el proceso de toma de decisiones, en este artículo se plantea el diseño de un modelo para la revisión de la calidad de la misma. Una de las bondades principales del modelo propuesto es su universalidad, ya que su aplicación no está circunscrita a un tipo específico de información geográfica, ni es exclusiva de una normatividad específica.

Palabras clave

Sistemas de información geográfica, calidad de la información geográfica

Abstract

In today's world, there exists a variety of software tools that support the decision making process in an organization. Some representative examples of this type of tools are the geographic information management systems, the OLAP (On Line Analytical Processing) systems and the database management systems. Particularly, the geographic information management systems have become more and more important in such a process due

to the common relationship between organizational data and geographical data, making it necessary to have a reliable control of the quality, accurateness and actuality of this type of data. Thus, considering the importance of the geographic information in the decision making process of an organization, a model for the revision of its quality is proposed in this paper. One of the main features of our model is its universality, since it is not designed for a specific data type, neither for a specific standard.

Key words

Geographic information management systems, geographic information quality

Introducción

Todas las civilizaciones invariablemente han generado información geográfica, que han utilizado para conocer con más detalle su entorno y resolver diversos problemas económicos y sociales. De esta manera, la cartografía siempre ha coadyuvado al progreso de las sociedades, siendo en la actualidad una parte fundamental para el desarrollo de las naciones. En México, el Instituto Nacional de Estadística, Geografía e Informática (INEGI) es el encargado de generar información geográfica para un gran número de temas (topografía, geología, uso de suelo, hidrología, etc.), que dan soporte a todos los sectores del país. Además, la Ley de Información Estadística y Geográfica (LIEG) [2] le confiere al INEGI las atribuciones de normar, evaluar, integrar y difundir la información estadística y geográfica que generan las Unidades Productoras de Información en la Administración Pública Federal, así como de regular conjuntamente con las autoridades federales y estatales la producción de información estadística y geográfica para su integración a los Sistemas Nacionales Estadístico y de Información Geográfica.

Hasta 1993 el INEGI generaba la mayoría de su información geográfica de manera analógica, es decir, en papel. Sin embargo, a partir de dicho año el INEGI ha venido generando este tipo de información de manera digital, la cual ha servido como base para la gran mayoría de las dependencias de gobierno, instituciones de educación y empresas particulares, mismas que requieren la información geográfica para generar información del tema de su competencia o para facilitar la toma de decisiones. No obstante, a pesar del avance tecnológico que actualmente posee el INEGI para la producción de información geográfica digital, aún existen grandes áreas de oportunidad para mejorar la calidad de la información producida. Si bien es cierto que cada línea de producción cuenta con mecanismos de control de la calidad, también es cierto que se ha detectado que en general la información no cumple al 100% la normatividad del propio Instituto. Esto obliga a realizar procesos de limpieza y edición de la información a fin de estar en condiciones de realizar procesos de análisis.

Así, la información geográfica producida por el INEGI presenta problemas en el cumplimiento de la normatividad establecida por el propio Instituto, lo que hace necesario un mayor aseguramiento de la calidad de este tipo de información para que pueda ser utilizada de manera confiable. Con este propósito, en este trabajo de investigación se propone el diseño de un modelo de revisión de calidad.

Fundamentos teóricos

A continuación se describen los problemas o errores más comunes ocurridos en el cumplimiento de la normatividad establecida por el INEGI para la calidad de la información geográfica, misma que puede ser consultada en los documentos titulados Modelo de Datos Vectoriales [3] y Diccionarios de Datos [4], publicados en el portal institucional del INEGI [5]. Como se verá, estos errores están relacionados con la geometría de rasgos geográficos, relaciones espaciales de entidades y atributos de objetos.

Errores de geometría

El documento de Modelo de Datos Vectoriales establece que la representación geométrica del componente espacial de un rasgo geográfico puede ser:

- Punto*: Se constituye por un solo par de coordenadas (x,y).
- Línea*: Se constituye por una serie de dos o más pares distintos de coordenadas, ligadas secuencialmente.
- Área*: Se delimita por una línea cerrada o una serie de líneas que cierran.

Lo anterior significa que todos los rasgos geográficos

que formen parte de la información geográfica que genera el INEGI deben ser modelados con base en estos tres tipos de representaciones. Por ejemplo, la entidad caseta de peaje puede ser modelada con una representación de punto, una carretera con una representación de línea y un cuerpo de agua con una representación de área.

Así, un error típico de geometría son los polígonos que no se cierran correctamente. Esta situación es crítica, ya que el usuario regularmente no tiene los elementos para poder agregar las líneas faltantes y cerrar el polígono a fin de poder utilizar la información en algún proceso de análisis espacial.

Errores en las relaciones espaciales

La normatividad para la generación de la información geográfica de un tema específico se encuentra en el correspondiente Diccionario de Datos, en el cual están especificadas las relaciones espaciales que debe cumplir cada entidad (carretera, localidad, etc.). Una relación espacial es una asociación entre entidades y, de acuerdo a la normatividad, las posibles relaciones espaciales son conectar y compartir, definidas a continuación.

•*Relación conectar*: Existe una relación de conectar entre ocurrencias de entidades cuando exista una intersección planimétrica entre los rasgos geográficos involucrados. Una relación de conexión implica la terminación de la representación geométrica de las entidades presentes en el punto de conexión. Es decir, las entidades serán segmentadas en el punto de conexión, con lo cual se garantiza que no existan excesos o defectos en la unión de la geometría de las entidades, apegándose así la relación a una realidad física.

•*Relación compartir*: Existe una relación de compartir entre ocurrencias de entidades cuando los rasgos geográficos percibidos como líneas o áreas sean parcial o totalmente contiguos o coincidentes. Una relación de compartición implica la conexión de todas las líneas que terminen en los extremos de la línea compartida.

Errores en los atributos

Los atributos representan las cualidades y características de un objeto del mundo real. Dicho de otra manera, cada rasgo está compuesto tanto por su geometría, como por sus atributos. Para la gestión de atributos, la normatividad especifica lo que un área productora de información geográfica debe tener en cuenta; especifica los atributos que deben existir, el tipo de cada uno de ellos (numérico o alfanumérico) y el dominio de valores válido para cada atributo; los errores comunes son los siguientes:

- Existen entidades con más o menos atributos de los especificados en su diccionario de datos.

- Existen atributos con otro tipo de datos diferente al indicado en el diccionario de datos. Por ejemplo, el atributo “número de carretera” de la entidad “carretera” está definido de tipo numérico en el diccionario de datos; sin embargo, en los datos dicho atributo puede encontrarse definido de tipo alfanumérico.
- Para algunos atributos, no todos sus valores están dentro del dominio de valores. Por ejemplo, el atributo “jurisdicción” de la entidad “carretera” se especifica en el diccionario de datos con un dominio de valores que va de 01 al 33, donde el rango 01- 32 corresponde a las claves de las entidades federativas y el 33 es utilizado para especificar una carretera con jurisdicción “federal”; sin embargo, en los datos dicho atributo se puede encontrar con valores de cero o valores mayores a 33.

Los errores en los atributos suceden comúnmente debido a cambios necesarios durante la gestión de la información geográfica, que debieron haber sido legitimados en la normatividad, pero no se hizo así; esto genera naturalmente inconsistencias y falta de completitud en los atributos.

Modelo propuesto

Antes de proceder al planteamiento de nuestra propuesta, es importante señalar que la problemática discutida en la sección anterior no es exclusiva del INEGI, sino más bien, es una problemática propia de cualquier productor de información geográfica, en mayor o menor grado. Es decir, los errores de geometría, los errores en las relaciones espaciales y los errores en los atributos pueden suceder en cualquier contexto donde se produzca información geográfica. Esto sugiere una propuesta de solución universal, más que específica. En este sentido, nuestra propuesta es el diseño de un modelo universal que permita la revisión de la calidad de la información geográfica, capaz de detectar al 100% los errores existentes en un conjunto de datos geográficos, sin importar su origen. Así, la problemática discutida en torno a la calidad de la información geográfica producida tanto en el INEGI, como en otras organizaciones, puede ser resuelta mediante un modelo que permita detectar los errores ocurridos en la misma. En la Figura 1 se muestra un esquema genérico de nuestra propuesta de solución.

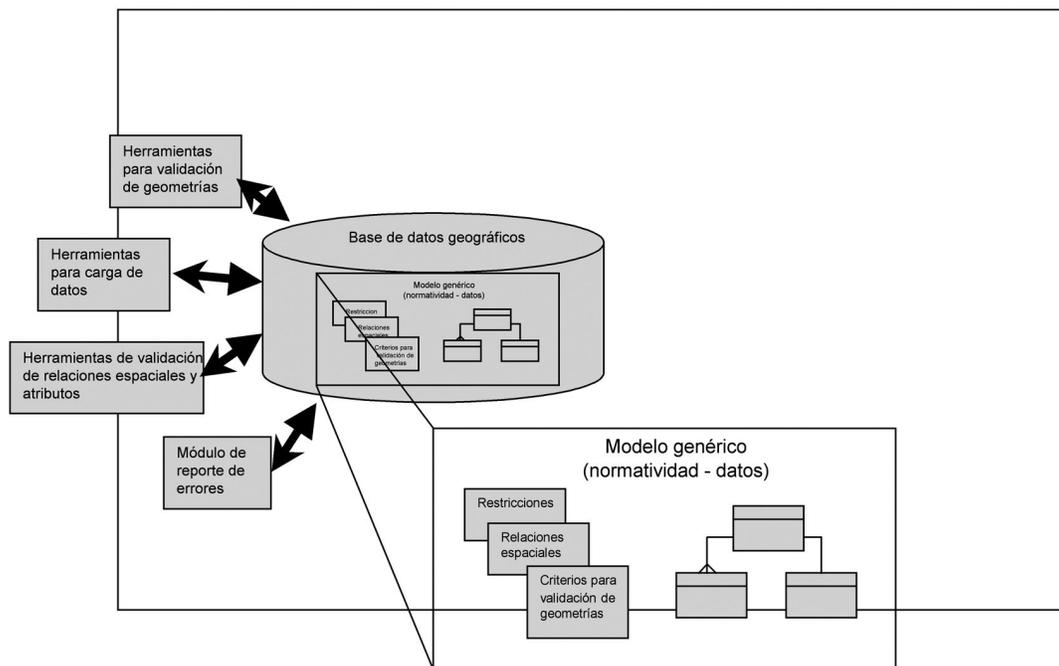


Figura 1. Modelo genérico para la revisión de la calidad de la información geográfica.

El modelo que se propone está integrado por las siguientes partes:

- Un modelo genérico subyacente en el repositorio que almacena tanto la parte normativa que deben cumplir los datos (restricciones, relaciones

espaciales, atributos, tipo de los atributos, dominio de valores de los atributos, etc.), como los datos mismos.

- Un conjunto de herramientas para la validación de geometrías
- Un conjunto de herramientas para la carga de datos en el repositorio.
- Herramientas para la validación de relaciones espaciales y atributos.
- Un conjunto de herramientas para la generación de reportes de errores.

Con la finalidad de diseñar un modelo sustentando los requerimientos de los usuarios, se utilizó el método

de desarrollo orientado a objetos que propone Craig Larman [6], el cual no fija una metodología estricta, sino que define una serie de actividades que pueden realizarse en cada fase, las cuales deben adaptarse según las condiciones del proyecto. La notación que fue utilizada para la diagramación fue la conocida como UML, que se ha convertido en el estándar de facto en la notación orientada a objetos [7].

Implementación del modelo

Una vez que se han detectado las entidades, relaciones, atributos y valores, se obtuvo el modelo lógico mostrado en la Figura 2.

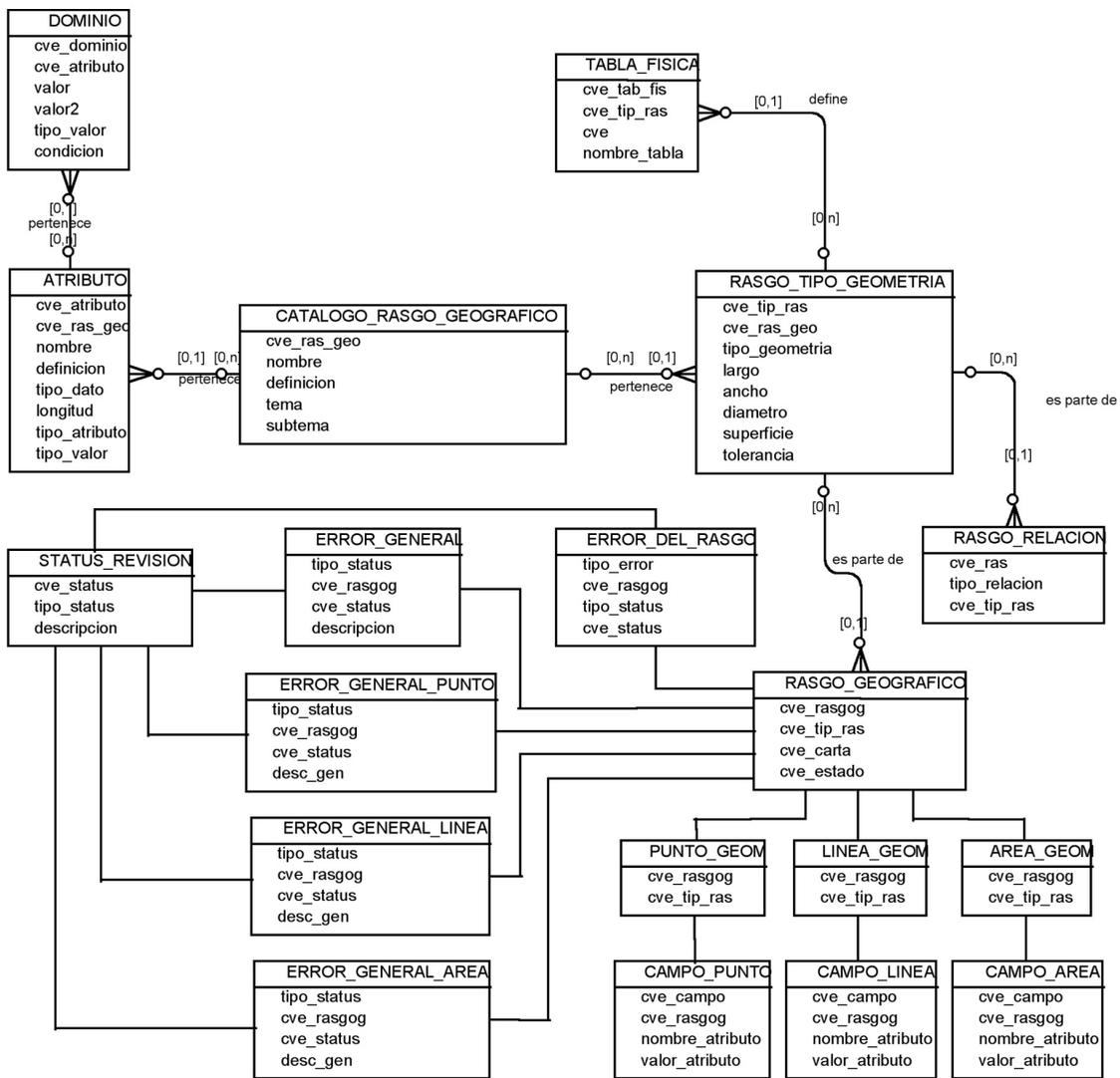


Figura 2. Modelo lógico de revisión de la calidad de la información geográfica.

En el diseño de este modelo se consideraron aquellas entidades y relaciones necesarias para la representación tanto de la información geográfica, como de la información normativa de una organización, teniendo en cuenta lo siguiente:

- independencia de sus aplicaciones individuales,
- independencia del sistema de bases de datos,
- independencia del hardware utilizado,
- independencia de la forma de almacenamiento.

Así, para el diseño del modelo fue necesario identificar los conceptos, las asociaciones y los atributos relacionados con el tipo de información mencionada. En primera instancia, la identificación de conceptos se llevo a cabo con base en los documentos de *especificación de requisitos*, que incluyen el documento descriptivo del sistema, los diagramas de casos de uso, los diagramas de actividades y el diseño de la interfaz inicial del sistema, entre otros, quedando identificados de manera definitiva los conceptos *rasgo_geográfico*, *rasgo_relación*, *geometría_rasgo*, *catalogo_rasgo_geográfico*, *error_rasgo*, *atributo*, *campo*, *dominio* y *error_general*. Estos conceptos, vistos como entidades bajo un enfoque relacional, tienen las asociaciones y los atributos mostrados en la Figura 2, que se identificaron con base en lo siguiente:

- Cada entidad debe tener al menos un atributo, de otra manera no se justificaría. También debe tener al menos un tipo de representación geográfica.
- Las entidades geográficas pueden o no tener relaciones espaciales con otras entidades de otro tipo de representación.
- Los atributos deben tener al menos un valor.
- Los rasgos geográficos deben tener al menos un atributo.
- Los rasgos geográficos deben ser almacenados (tabla *rasgo_geográfico*) para poder ejecutar posteriormente los procesos de revisión.
- En la revisión de los rasgos geográficos contra la norma pueden o no haber errores de completitud y de consistencia lógica.

Cabe señalar que el modelo lógico contempla todas las observaciones que se hicieron durante la fase de validación del diseño conceptual, quedando implementado y operando satisfactoriamente en la detección de errores y análisis espacial para la validación de relaciones espaciales.

Resultados

A continuación se muestran algunos resultados obtenidos al aplicar las herramientas de verificación de la calidad de la información geográfica, para lo cual se tomaron algunos reportes de error generados al verificar algunos conjuntos de datos de la información topográfica escala 1:50,000.

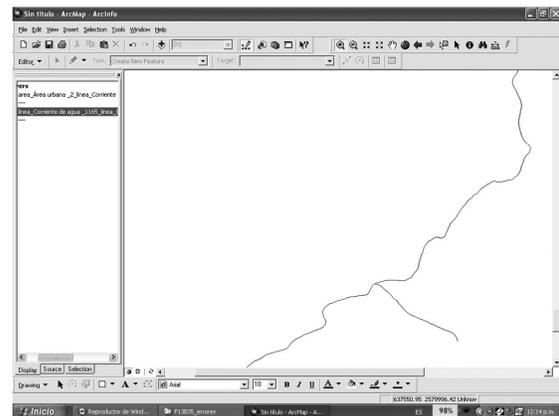


Figura 3. Corrientes de agua aparentemente conectadas.

En la Figura 3 se muestran un par de corrientes de agua aparentemente conectadas; debido a la escala de visualización utilizada no es posible apreciar claramente la continuidad entre los dos rasgos.

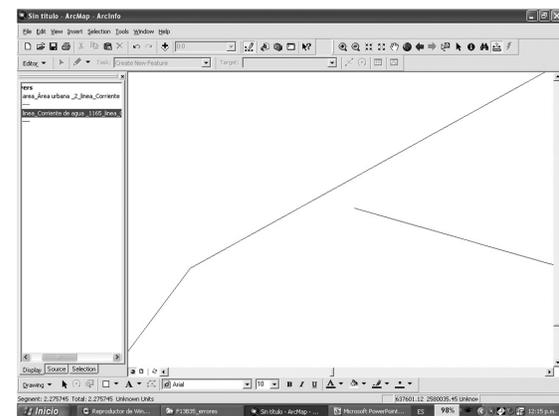


Figura 4. Acercamiento a corrientes de agua no conectadas.

En la Figura 4 se muestra un acercamiento a la zona donde las dos corrientes de agua deben conectarse, en la cual se puede apreciar fácilmente que los rasgos no se conectan. De esta manera se determina que existe un error de conexión que debe ser corregido por el área productora.

Conclusiones

En este artículo se discutió de manera resumida el trabajo realizado para lograr nuestro propósito de diseñar un modelo para la revisión de la calidad de la información geográfica, mismo que permite evaluar la información geográfica generada por las áreas

productoras con base en la normatividad establecida por el INEGI.

Como parte del diseño de nuestro modelo se propuso un modelo de base de datos para almacenar cualquier tipo de información geográfica con las características de representación geométrica ya especificadas, capaz de contener la normatividad establecida en el Diccionario de Datos respecto a los atributos, restricciones, relaciones espaciales, tipo de los atributos, dominio de valores de los atributos y las relaciones espaciales válidas para cada entidad del tema a revisar. A partir del modelo de base de datos se desarrollaron un conjunto de herramientas que permiten detectar los errores e inconsistencias en la información geográfica. Así, una vez que las herramientas fueron probadas en la validación de la información topográfica escala 1:250,000 serie III producida por el INEGI, se pusieron en marcha de manera definitiva en el INEGI con resultados satisfactorios, lográndose así el propósito de ofrecer información geográfica de mejor calidad.

Referencias

- [1] Portal del Instituto de Investigación de Recursos Biológicos Alexander Von Humboldt (2007), <http://www.humboldt.org.co>.
- [2] Ley de Información Estadística y Geográfica (1980), "Artículos 2, 13, 16 y 21", *Talleres Gráficos del Instituto Nacional de Estadística, Geografía e Informática*, Aguascalientes, Ags., México.
- [3] Instituto Nacional de Estadística, Geografía e Informática (1993), "Modelo de Datos Vectoriales", *Talleres Gráficos del Instituto Nacional de Estadística, Geografía e Informática*, Aguascalientes, Ags., México.
- [4] Instituto Nacional de Estadística, Geografía e Informática (1996), "Diccionario de Datos de la Información Topográfica Escala 1:50,000", *Talleres Gráficos del Instituto Nacional de Estadística, Geografía e Informática*, Aguascalientes, Ags., México.
- [5] Portal Institucional del Instituto Nacional de Estadística, Geografía e Informática (2007), <http://www.inegi.gob.mx>.
- [6] Larman, C. (1998), *Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design*, Prentice-Hall, Inc.
- [7] Lethbridge, T.C. & R. Laganière (2007), *Object-Oriented Software Engineering: Practical software development using UML and Java*, McGraw Hill.

Agradecimientos: Trabajo colaborativo entre miembros de la Red Latinoamericana de Tecnologías de la Información.

Artículo Recibido: 8 de octubre del 2007

Aceptado para publicación: 6 de diciembre del 2007