CONACYT

CENTROGEO
Centro Público de Investigación
CONACYT

Diplomado en Análisis de Información Geoespacial

Pre-procesamiento de datos vectoriales

Autor:
M. en G. Luis A. Castellanos Fajardo

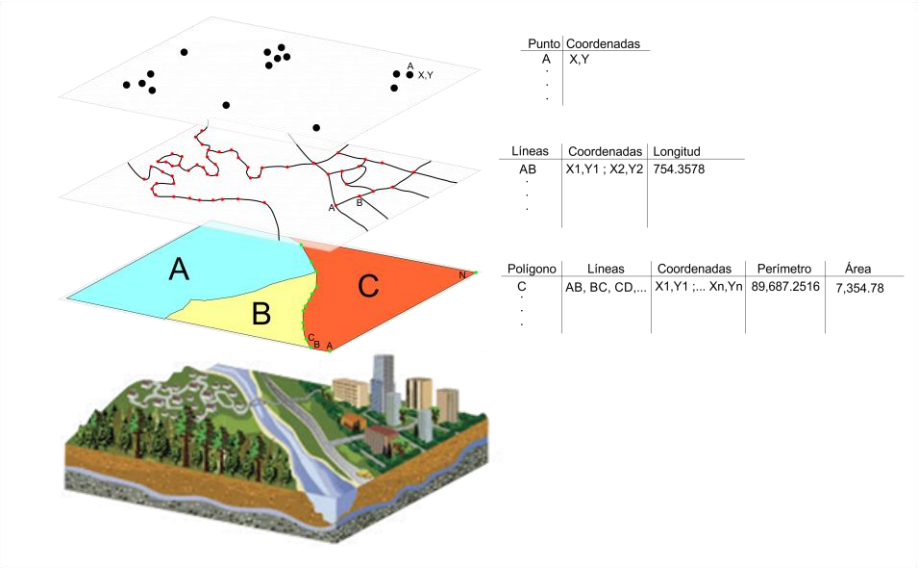
Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

Contoy 137 Esq. Chemax, Col. Lomas de Padierna, Delegación Tlalpan, CP 14240, México, CDMX.

Datos Vectoriales y pre–procesos

La construcción de una base de datos geográfica implica un proceso de abstracción para pasar de la complejidad del mundo real a una representación simplificada que pueda ser procesada con las actuales herramientas de cómputo.

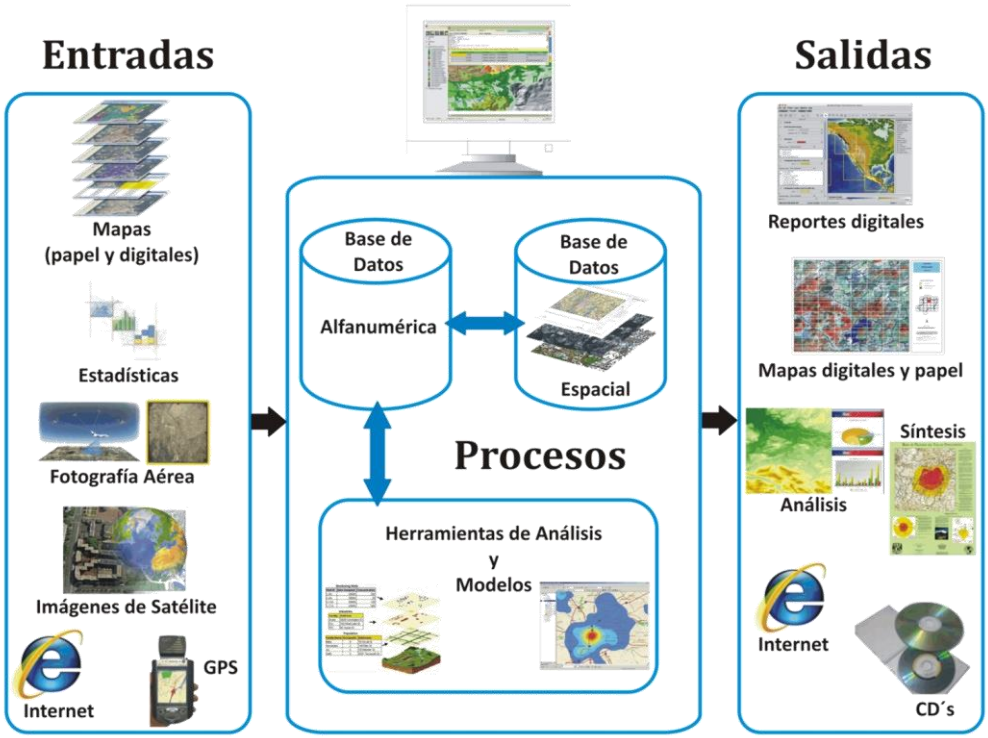
El modelo de datos vectorial utiliza tres elementos geométricos para modelar digitalmente las entidades del mundo real, el **Punto**, la **Línea** y el **Polígono**.



Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

Contoy 137 Esq. Chemax, Col. Lomas de Padierna, Delegación Tlalpan, CP 14240, México, CDMX.

Desde el punto de vista de los sistemas, los Sistemas de Información Geográfica (SIG) tienen tres elementos esenciales **Entradas**, **Procesos** y **Salidas**, que deben cumplir las siguientes funcionalidades: **Entradas** (adquisición de datos y pre-procesamiento), **Procesos** (manejador de base de datos alfanumérica y espacial, funciones de análisis y modelos) y **Salidas** (generación de productos).

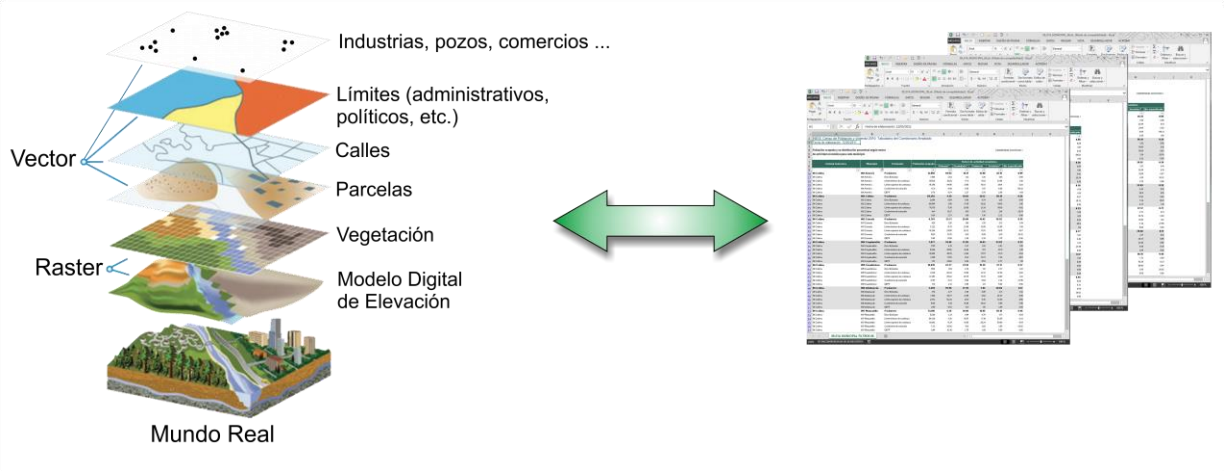


La adquisición de datos se refiere al proceso de identificación y recopilación de los datos específicos que requiere el sistema o la aplicación. Aparte de estas tareas hay otras dos que son el pre-procesamiento y el ingreso al sistema.

Se deben ingresar dos tipos diferentes de datos al SIG: referencias geográficas y atributos.

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

Contoy 137 Esq. Chemax, Col. Lomas de Padierna, Delegación Tlalpan, CP 14240, México, CDMX.



Frecuentemente los datos requieren un pre-procesamiento, manual o automatizado, antes de incorporar los datos al sistema. Por ejemplo, los tabulados necesitan ser organizados y capturados en la base de datos del SIG y relacionada a su geometría, la cartografía digital y las imágenes de satélite necesitan ser geo-referenciadas o corregirse geométricamente, además de poner todos los archivos geográficos en la proyección cartográfica que se va a manejar en el sistema. La cartografía en papel necesita ser digitalizada y georeferenciada. En otras ocasiones se tendrán que elaborar procedimientos para el escaneo de mapas en papel, fotografías aéreas o imágenes de satélite no digitales. También se pueden requerir procedimientos para cambios de formato entre archivos de software diferentes.

Todas las actividades y procesos que se hagan para que los datos puedan ser incorporados al sistema formaran parte de las tareas de pre-proceso.

La preparación del adecuado conjunto de datos es vital para poder llevar a cabo con éxito cualquier proyecto SIG.

La calidad de un dato espacial depende de muchos factores. Las características que proveen dicha calidad al dato espacial son diversas, pues el dato espacial es en sí complejo, y cada una de estas características es susceptible de agregar errores y por

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

tanto de implicar una pérdida de calidad. A continuación se presentan algunos puntos importantes sobre la calidad del dato espacial (Pepijn van Oort, 2005):

- *Exactitud de posición*. Todo dato espacial tiene asociada una posición geográfica. La precisión con la que se toma esta condiciona la calidad del dato. Esta precisión puede considerarse únicamente en los ejes "*x*" y *"y"*, o también en el eje "*z*" (elevación). Si se trabaja en un SIG bidimensional, esta última, puede considerarse como un atributo, y tratarse de la misma forma que cualquier otra variable de similar índole sin significado espacial, tal como la precipitación en el punto (*x, y*) en cuestión.

- *Exactitud en los atributos*. Si la componente espacial puede tener errores, estos también pueden aparecer en la componente temática. Los valores asociados a una coordenada u objeto espacial pueden haber sido medidos con mayor o menor exactitud, o presentar valores incorrectos por muy diversas causas. Cuando el atributo en cuestión es de tipo categórico, puede existir un error de clasificación (se asocia la entidad espacial a una categoría equivocada), mientras que en el caso de atributos no categóricos pueden sencillamente aparecer valores mayores o menores que los reales.

- *Consistencia lógica y coherencia topológica*. Los datos espaciales no son elementos independientes, sino que existen relaciones entre ellos. Un dato de calidad debe almacenar fielmente estas relaciones, siendo la topología la encargada de reflejar este tipo de información. Por ello, debe existir una coherencia topológica en el dato espacial. Además de la coherencia de las relaciones, existe una coherencia implícita en todo atributo o valor recogido, de forma que resulte lógico. Estos atributos y valores han de ser coherentes con las escalas de medida o el tipo de valor que se espera, entre otros. Así un valor de elevación no puede ser igual a "*AgH*", ni un valor de precipitación igual a *-98*.

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

- *Completes*. El dato espacial no recoge todo lo que existe en una zona dada. Algunos elementos pueden no haberse recogido por cuestiones de escala (menores de un tamaño mínimo), pero también pueden incluirse o excluirse en función de otros criterios, en especial para el caso de mapas temáticos. Estos criterios deben conocerse para saber por qué un dato espacial contiene una serie de valores o elementos y no otros.

- *Calidad temporal*. Aunque los datos espaciales son (representaciones) estáticas de la realidad, el tiempo es muy importante, pues puede afectar directamente a su calidad. La realidad que representa un dato geográfico es una realidad que varía con el paso del tiempo, y por tanto este paso del tiempo puede degradar la calidad del dato espacial en mayor o menor medida. Llega a indicar la existencia de rasgos que quizá ya no existen, por ejemplo *"Bosques"* donde hace mucho que ya no existe o su área es mucho menor. Al ver que el mapa nos presenta la presencia de bosques podemos cometer el error de preparar planes de manejo de dichos "Bosques" cuando estos ya no existen.

- *Procedencia*. Un dato espacial puede provenir de una fuente más o menos confiable, o haber sido generado a través de uno o varios procesos, en cada uno de los cuales se puede haber introducido algún tipo de error. Conocer la procedencia de un dato y los procesos que se han empleado en su elaboración es necesario para poder evaluar su calidad.

Es importante ser consciente de los errores que contienen los datos y de la posible aparición de estos a medida que realizamos tareas con ellos, con objeto de minimizar dicha aparición y limitar la presencia e influencia de los errores en los resultados. Es por eso que cobra suma importancia el que la información que consultemos tenga

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

metadatos que nos indiquen el origen y los procesos realizados al conjunto de datos que queremos utilizar.

Las fuentes de datos son muy variadas, y a la hora de elaborar un proyecto SIG podemos obtener datos de muchas fuentes distintas. Conocer todas estas fuentes de datos es importante para elaborar una base de datos geográfica que permita obtener los mejores resultados posibles, pero también lo es el conocer la forma en que esos datos pueden obtenerse. Los datos geográficos se van a almacenar en archivos, existiendo muchos formatos de archivo distintos para recoger un mismo conjunto de datos.

Vector

Sin ser tan abundantes como los formatos para datos raster, existe también un buen número de formatos de archivo para datos vectoriales:

Shape File (shp). Propuesto por la empresa *ESRI*, es el formato más utilizado en la actualidad, convertido en un estándar de facto. No soporta topología y se compone de diversos archivos, cada uno de los cuales contiene distintos elementos del dato espacial (geometrías, atributos, índices espaciales, etc.)

MapInfo TAB. Desarrollado con el software *MapInfo*, de la empresa del mismo nombre.

Los formatos tipo CAD también son muy habituales, destacando los siguientes:

DWG. El formato nativo de *AutoCAD*, la aplicación más popular en el diseño asistido por computadora. Permite almacenar datos bidimensionales y tridimensionales, así como metadatos. Se trata de un formato cerrado, y la lectura y escritura de este

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

requiere el uso de librerías facilitadas en condiciones determinadas por la empresa *Autodesk*, fabricante de *Autocad* y creador del formato. Aun así, son muchas las aplicaciones que pueden leer archivos DWG.

DXF. También derivado de Autocad, surge para impulsar la interoperabilidad en el campo del CAD. En este caso no se trata de un formato cerrado y existe gran cantidad de datos cartográficos en este formato.

DGN. El formato nativo de otra aplicación CAD muy extendida: *MicroStation*, de la empresa *Intergraph*.

Varias instituciones nacionales e internacionales ofrecen datos en formato vectorial y de atributos para distintos temas, censales, económicos, de recursos naturales, geofísicos, etc. Esto a través de sus portales de internet o a través de las Infraestructuras de Datos Espaciales (***IDE***). Según *Wikipedia*, una ***IDE*** integra datos, metadatos, servicios e información de tipo geográfico para promover su uso y está formada por el conjunto de "tecnologías, políticas, estándares y recursos humanos para adquirir, procesar, almacenar, distribuir y mejorar la utilización de la información geográfica".

Las *IDEs* están compuestas por tres tipos diferentes de información: ***datos espaciales, servicios y metadatos***. Estos tres componentes ordenados y estructurados mediante estándares, reglas y marco normativo, como la Directiva INSPIRE, conforman la red de *Infraestructuras de Datos Espaciales*[1].

---

[1] *IDE Andalucía.* *http://www.ideandalucia.es/portal/web/ideandalucia/ides/conceptos;jsessionid=C149DF0AE81B9C7AC56B649A2C8EC4CE*

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

Datos Espaciales: los datos espaciales digitales pueden ser puramente topográficos o temáticos.

Servicios: Son las funcionalidades accesibles mediante tecnologías telemáticas interoperables que una IDE ofrece al usuario para interactuar con los datos geográficos. Estas funcionalidades se organizan en cinco tipos de servicios: búsqueda, visualización, descarga, transformación y acceso. Los estándares de servicios geográficos más usuales son:

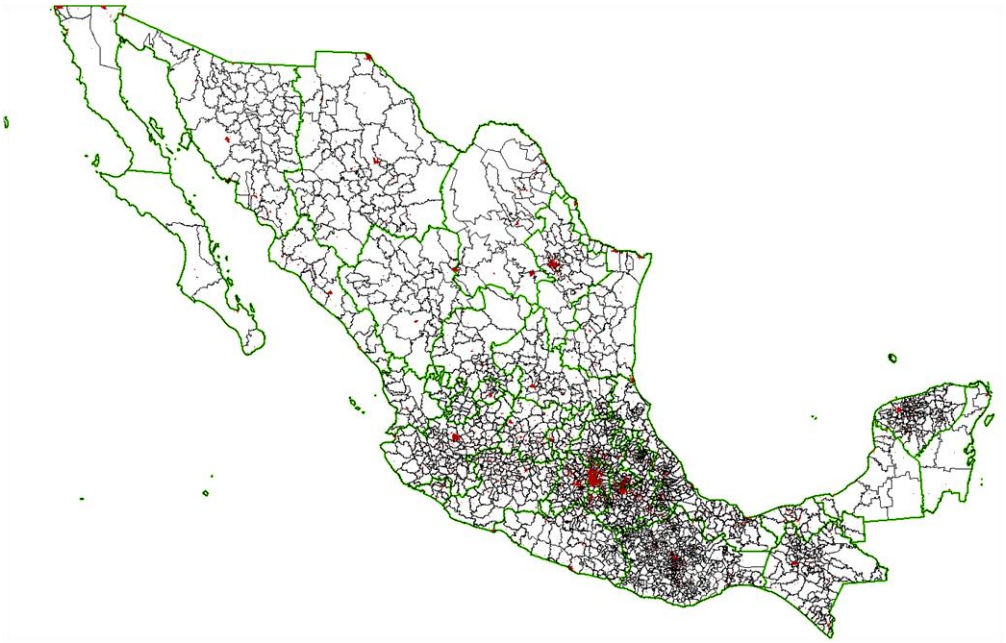
a. ***WMS (Servicio de mapas):*** La información se muestra como imágenes compuestas por una o varias capas y generadas a partir de información espacial vectorial o raster.
b. ***WFS (Servicios de objetos):*** Ofrece la información en capas de datos en las que las que ésta está compuesta por objetos vectoriales. El acceso se produce mediante el formato GML.
c. ***WCS (Servicios de coberturas):*** Este servicio es similar al WFS pero la información que se muestra es de tipo raster.
d. ***CSW (Servicio de catálogos):*** Permite buscar información geográfica a partir de consultas basadas en los metadatos.
e. ***WFS-G (Servicio de Nomenclator o Gazetter):*** Permite localizar entidades geográficas a partir de consultas por su nombre. Relaciona el lugar con las coordenadas que la definen.
f.

Metadatos: Los metadatos son datos asociados a la información producida, que recogen las características fundamentalmente descriptivas de éstos (fecha, origen, autor, sistema de coordenadas, etc.). Los metadatos representan un elemento indispensable en cualquier IDE, debiendo ser obligatoria su inclusión, puesto que resulta esencial conocer las características de la información con la que se va a

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

trabajar. El estándar internacional que regula los metadatos es la ISO 19115 “Geographic Information Metadata”.

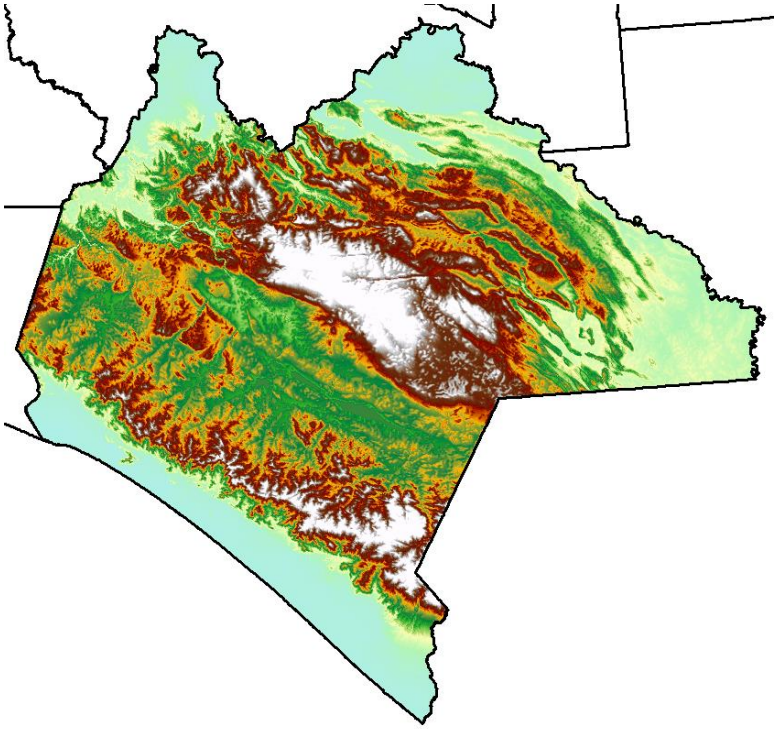
En México el *INEGI, SEMARNAT, CONABIO, SAGARPA*, y algunas instituciones de educación superior como la *UNAM y el Colegio de Postgraduados de Chapingo* entre otras, ofrecen en sus páginas datos en formato vectorial y en algunos casos raster, para ser utilizados por cualquier persona. Ofrecen servicios para visualizar y para obtener los datos en un formato estándar en el caso de los datos vectoriales generalmente en formato ***SHP***.



Datos vectoriales del límite estatal, límite municipal y zonas urbanas.
Fuente INEGI (2014): Recuperado de http://www.inegi.org.mx/geo/contenidos/geoestadistica.

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

Contoy 137 Esq. Chemax, Col. Lomas de Padierna, Delegación Tlalpan, CP 14240, México, CDMX.



Fuente INEGI (2014): Datos raster. Modelo de Elevación del estado de Chiapas.
http://www.inegi.org.mx/geo/contenidos/datosrelieve/continental/default.aspx

Raster.

Los formatos de archivo para datos raster son muy abundantes, existiendo numerosas alternativas con diferencias en ocasiones notables entre sí.

Imágenes de satélite, radar, LiDAR, fotografía aérea digital, mapas escaneados entre otros son ejemplos de fuentes de datos raster.

Los usos de estas fuentes son variados, por ejemplo, las imágenes de satélite frecuentemente son utilizadas para inventario y monitoreo de recursos naturales, para trabajos de ubicación, efectos y control de desastres como inundaciones, impactos de tsunamis, o terremotos.

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

Figura. Imágenes Spot 5 de la región de Aceh, Sumatra, Indonesia. Al comparar la imagen del 8 de diciembre de 2004 (izquierda) con la imagen de 29 de diciembre de 2004 (derecha), se observa la amplitud de la zona afectada por el maremoto en color magenta en esta última. Fuente: CRISP, Singapore. (2004). Imagen Spot 5 de Indonesia. Recuperado el 01 de diciembre de 2014 de http://www.crisp.nus.edu.sg/tsunami/tsunami.html

Los formatos de archivo para imágenes son adecuados para recoger los colores de las imágenes, pero esto no es suficiente a la hora de almacenar otros valores (por ejemplo, valores decimales) o bien cuando son necesarios un número más elevado de bandas, como en el caso de imágenes híper-espectrales.

Una imagen en blanco y negro o en escala de grises contiene una banda. Una imagen en color contiene tres, ya que los colores se expresan como una terna de colores básicos: rojo, verde y azul. Este es el fundamento del modelo de color RGB, en el cual todo color es la combinación de distintas intensidades de los anteriores colores básicos. Las intensidades de cada banda (o las intensidades de la única banda en el caso de una imagen en escala de grises) se expresan habitualmente con valores entre 0 y 255, un rango que resulta insuficiente para el manejo de otras variables tales como las variables físicas que pueden emplearse en un SIG, ya que estas presentan valores continuos.

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

En estos casos, los formatos de imágenes no son adecuados en su forma original, y deben o bien adaptarse o bien emplearse formatos más específicos que tengan en cuenta el tipo particular de imágenes que se almacenan.

Otro problema es la presencia de celdas sin datos. La existencia de celdas sin datos es un hecho que no contemplan los formatos de imágenes. A estas celdas se les asigna un valor establecido por defecto, el cual ha de definirse en el propio archivo para que después sea reconocido por el SIG, pero muchos formatos de imagen no pueden almacenarlo. Una posible solución es la utilización de formatos que permitan transparencia. En estos, se puede especificar un color como transparente, que a efectos de su utilización en un SIG puede considerarse como indicación de la ausencia de datos. Estos formatos, no obstante, no son los más adecuados para datos SIG, y esta solución no resuelve por completo esta deficiencia.

Esta es una breve lista de los principales formatos existentes para raster, los cuales suelen encontrarse con frecuencia a lo largo del desarrollo de un proyecto SIG.

Tagged Image File Format (TIF). Se trata de un formato complejo y altamente flexible, con muchas variantes distintas. Puede incorporar tanto compresión con pérdidas como sin pérdidas, en función del algoritmo que se utilice. Se utiliza habitualmente tanto en el ámbito del tratamiento de imágenes como en el ámbito SIG. En este último, permite también el almacenamiento de valores decimales, siendo apto para almacenar capas que no representen imágenes como tal. Es un formato habitualmente generado por los escáneres, con lo cual es frecuente su utilización al trabajar con cartografía escaneada. Existe una variante denominada GeoTIFF, que permite incorporar en el propio fichero la georreferencia de la imagen, haciendo innecesario el uso de un *World File* asociado.

Enhanced Compression Wavelet (ecw). Formato desarrollado por *Earth Resource Mapping*. Está especialmente preparado para almacenar imágenes de gran tamaño, ya

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

que las imágenes aéreas o de satélite en general tienen tamaños mayores que las imágenes de uso genérico para las que están pensados los formatos como TIFF o JPEG. Se trata de un formato de compresión con pérdidas, y su grado de compresión es alto.

Multi-resolution Seamless Image Database (MrSID) (sid). El formato *MrSID* es un formato cerrado, pero sus características son similares: alta compresión, preparado para imágenes de gran volumen y con posibilidad de descompresión selectiva.
También hay otros formatos muy comunes como el ***IMG*** o el ***GRID*** utilizados por ERDAS y ESRI para el manejo de datos raster.

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.

Bibliografía

1. Anselin, L. (1989). What is special about spatial data? Alternative perspectives on spatial data analysis. In: Symposium on Spatial Statistics, Past, Present and Future. Department of Geography, Syracuse University.
2. Antenucci, John C.; Brown, Kay; Croswell, Peter L.; Kevany Michel &Archer Hugh. (1991). Geographic Information Systems, A Guide to the Technology. Van Nostrand Reinhold, New York.
3. Aronoff, Stan. (1991). Geographic Information Systems: A management perspective. WDL Publications, Ottawa (Canadá).
4. Bosque, Joaquín (1992). Sistemas de Información Geográfica. Ediciones RIALP, SA, Madrid España.
5. C. A. Davis, F. T. Fonseca, and K. A. V. Borges (2003). A exible addressing system for approximate geocoding. In V Brazilian Symposium on GeoInformatics, 2003. http://www.geoinfo.info.
6. Marble, DUane F. (1984). Geographic Information Systems an Overview. Proceedings, Pecora 9 Conference, Siux Falls, S. D.
7. M.J. Jackson and P.A. Woodsford. (1991). Geographical Information Systems: Principles and Applications, chapter GIS data capture hardware and software, pages 239–249. Longman.
8. Nyerges, T.L. (1991). Analytical Map Use. Cartography and GIS, 18.
9. Olaya, Víctor. (2010). Sistemas de Información Geográfica. OSGeo
10. Peuquet, Donna J. & Marble, Duane F. (1990). Introductory Readings in Geographic Information Systems. Taylor and Francis.
11. Pepijn van Oort. (2005) Spatial data quality: from description to application. Netherlands Geodetic Commission..

Derechos Reservados © Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", A.C.