

## Capítulo 3

# Modelos y estructuras de datos

La función de cualquier programa informático es ejecutar un conjunto de instrucciones, escritas en un lenguaje formal<sup>1</sup> capaces de manipular símbolos que representan algún tipo de situación del mundo real; en el caso de un SIG objetos y variables espaciales. El ser humano es capaz de desenvolverse sin mayores problemas a pesar de que su conocimiento del mundo es ambiguo e incompleto; no ocurre lo mismo con los ordenadores, las descripciones de la realidad que manejan no deben contener ambigüedad alguna. Por tanto los lenguajes formales siguen estrictas reglas lógicas y asumen un conocimiento preciso de las situaciones del mundo real que se simbolizan.

Un buen ejemplo de esto es el de las curvas de nivel cortadas para insertar el texto que describe su altitud (figura ??). Esta opción es adecuada para un mapa en papel que debe interpretar un usuario humano, pero un ordenador interpretara que las curvas se han cortado y que existen otras *curvas pequeñas* que generan patrones extraños. Por otra parte para que el ordenador procese adecuadamente estas curvas, su altitud debe integrarse como una variable enlazada al objeto y no como un rótulo escrito al lado.

En SIG, al margen de asignar a cada punto de la superficie un par de coordenadas siguiendo un sistemas de proyección, debemos definir, formalmente, *estructuras de datos* (conjunto finito de datos discretos y fácilmente manipulables con un ordenador) que representen entidades y variables y las *instrucciones* que utilizará el ordenador para manipular estas estructuras de datos. Se trata en definitiva de un problema de lenguaje. En este tema se tratará la base teórica de la codificación de los diferentes elementos y atributos de la superficie terrestre en formato digital para su tratamiento con un programa SIG.

Por ejemplo, en el Sistema de Información de una empresa, los distintos trabajadores se codificarían mediante estructuras de datos similares a esta:

```
trabajador{
    Nombre (80 caracteres)
    NIF (9 caracteres)
```

---

<sup>1</sup>Los lenguajes naturales permiten expresar ambigüedades e ideas contradictorias, los lenguajes formales (los lenguajes de las computadoras) no

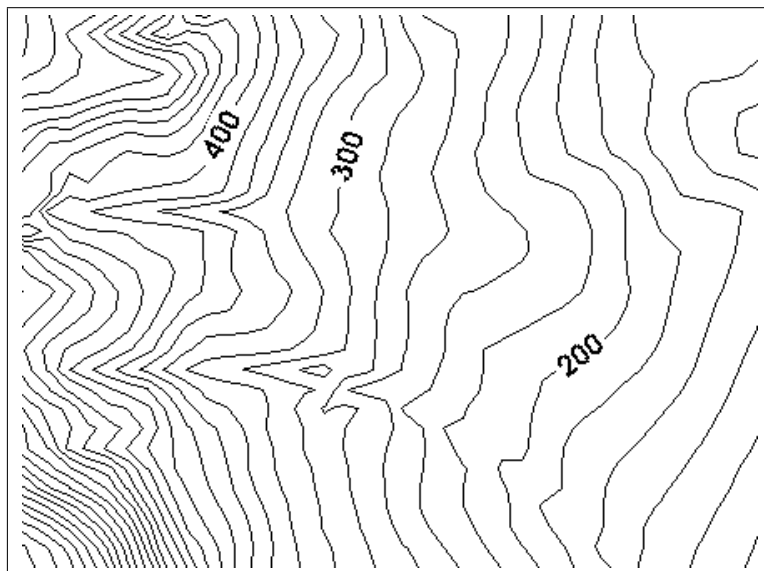


Figura 3.1: Curvas de nivel cortadas para insertar texto

```

Dirección (120 caracteres)
Edad (número entero)
Año de contratación (número entero)
Categoría (20 caracteres)
}

```

de manera que toda la información relevante queda codificada mediante elementos simples (cadenas de caracteres y números enteros). Las instrucciones deberían permitir responder a consultas como obtener el número de trabajadores que llevan más de 20 años en la empresa.

Partiendo de la definición de modelo como representación simplificada de la realidad, que refleja lo fundamental de esta ignorando los detalles accesorios, en informática se denomina *modelo de datos* al conjunto de reglas utilizadas para representar las diferentes entidades que deben almacenarse en la base de datos (trabajadores, proveedores, clientes, etc.) mediante elementos sencillos.

Existe una diferencia fundamental entre el SI de una empresa como el presentado anteriormente y un SIG. Los elementos que deberán integrar un SIG incluyen variables que, como la temperatura, varían constantemente de un punto a otro del espacio y en muchas ocasiones se desconoce su valor preciso, por otro lado se manejan entidades que no tienen límites precisos, por ejemplo la Sierra de Carrascoy, finalmente habrá entidades con límites precisos pero variables con el tiempo.

Codificar toda la información relevante acerca de una porción del territorio en forma de estructura de datos es, por tanto, mucho más complejo que en el caso anterior. Supone un gran salto que se entiende mejor si consideramos la existencia de varios niveles de abstracción:

1. **Realidad perceptible** (montañas, lagos, campos de cultivo, etc.). Nivel propio de los gestores preocupados por problemas de gestión y planificación del espacio.
2. **Modelo conceptual**. Nivel de los científicos (geógrafos, geólogos, ecólogos, etc.) que desarrollan, verifican o aplican teorías e hipótesis sobre variables y procesos que tienen lugar en el espacio. Estos consideran la existencia de dos tipos fundamentales de elementos sobre la superficie terrestre: entidades<sup>2</sup> y variables. Deben ser capaces de tomar un problema abstracto del nivel anterior y determinar las variables implicadas en su resolución.
3. **Modelo lógico**. Nivel de los técnicos en SIG que utilizan las herramientas del sistema para llevar a cabo, en el ordenador, las tareas requeridas por gestores o científicos. En lugar de trabajar con la realidad trabajan con representaciones de la misma que suelen ser de dos tipos: raster y vectorial. Deben decidir cual es la más adecuada para representar las variables obtenidas en el desarrollo del modelo conceptual y cuales son los procedimientos más adecuados para obtenerla con los datos de partida disponibles.
4. **Modelo digital o estructura de datos**. Nivel de informáticos, y desarrolladores de SIG cuya misión es optimizar las estructuras de datos utilizadas para almacenar la información y ampliar el repertorio de herramientas para cumplir en la medida de lo posible las necesidades de los científicos y técnicos en SIG.

Tal como aparece en la figura ??, cada uno de estos pasos está interrelacionado con los demás. La comunicación entre unos y otros se hace cada vez más difícil cuanto más alejados se encuentren en la anterior jerarquía, por tanto es necesaria la existencia de especialistas de los diferentes niveles para el adecuado desarrollo de un proyecto, al mismo tiempo se requiere un diálogo constante entre los mismos. No obstante suele darse el caso de personas que abarcan varios niveles de esta jerarquía, especialmente cuando la introducción de los SIG en una organización está dando sus primeros pasos.

### 3.1. El modelo conceptual. Entidades y variables

La realidad, por ejemplo el trozo de realidad representado en la figura ?? puede entenderse según dos modelos mentales (conceptuales) en principio contradictorios:

1. Como un continuo definido por una serie de *variables* que pueden ser de tipo cualitativo (litología, usos del suelo, etc.) o bien tratarse de *superficies*<sup>3</sup> (elevaciones, precipitación, etc.) (figura ??). De este modo una porción del territorio puede caracterizarse por la superposición de un conjunto de superficies que se consideran como más significativas.

---

<sup>2</sup>también denominadas como objetos

<sup>3</sup>Se trata de variables cuantitativas que varían de forma más o menos suave a través del espacio. A este tipo de variables espaciales se les denomina en algunos manuales de SIG como *campos* por analogía con el campo electromagnético que muestra propiedades similares, se ha utilizado también el término *variable regionalizada* que procede de la geoestadística, aquí se ha preferido el término *superficie*, también muy utilizado, para evitar confusiones cuando se hable de variables en un sentido más general)

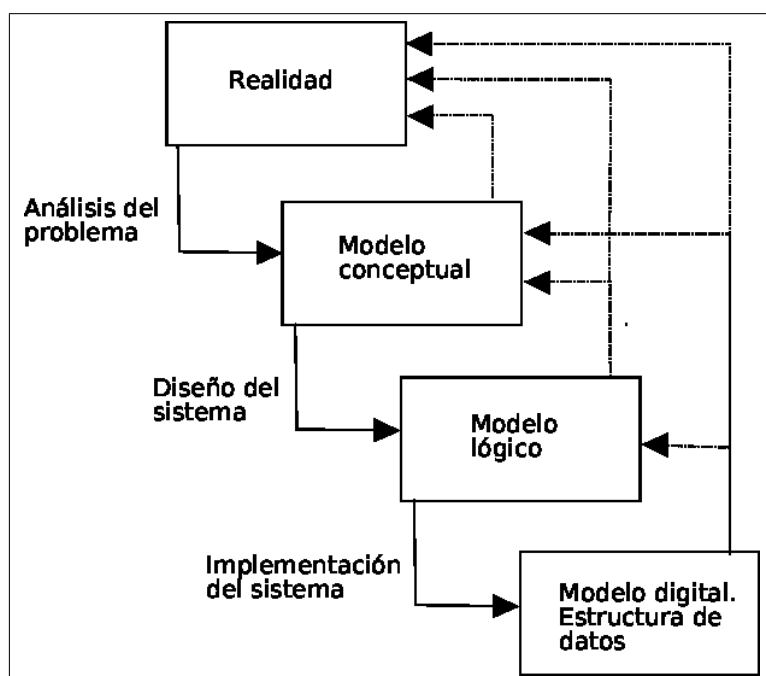


Figura 3.2: Esquema de la modelización de datos en SIG

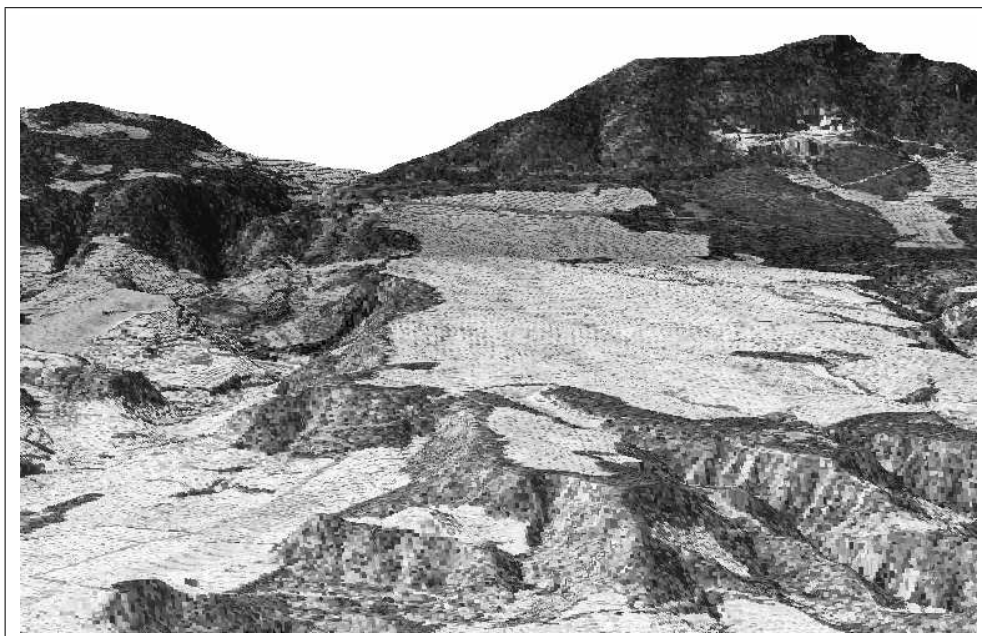


Figura 3.3: Realidad

2. Como la yuxtaposición de *entidades* de límites definidos y con características homogéneas, por ejemplo parcelas de propiedad, núcleos urbanos, carreteras, etc. Cada uno de estos objetos va a tener un identificador único (figuras ?? y ??). Mientras que las variables cubren el espacio de forma completa, una capa formada por un conjunto de objetos puede no hacerlo.

### 3.1.1. Superficies

En el primer caso podemos estudiar la realidad como un conjunto de superficies definidas por variables cuantitativas (altitud, humedad del suelo, precipitación, densidad de vegetación) que adquieren diferentes valores en diferentes puntos del espacio, siendo estos valores más parecidos cuanto más cerca se encuentren los puntos, a esta propiedad se denomina *autocorrelación espacial*<sup>4</sup>.

Las superficies son objetos tridimensionales con dos dimensiones que representan los ejes espaciales y una tercera que representa una tercera variable cuantitativa representada en cada punto del espacio. Este tipo de modelos se suelen denominar como de dos dimensiones topológicas y media (gráficos 2,5D), pues en realidad la tercera dimensión (la Z) no se analiza en su totalidad, no se considera exactamente un hecho volumétrico, sino una superficie (las dos dimensiones) ondulada, levantada en tres dimensiones (la media dimensión). Los gráficos

<sup>4</sup>En el tema 5 se discuten las implicaciones matemáticas del concepto de autocorrelación espacial

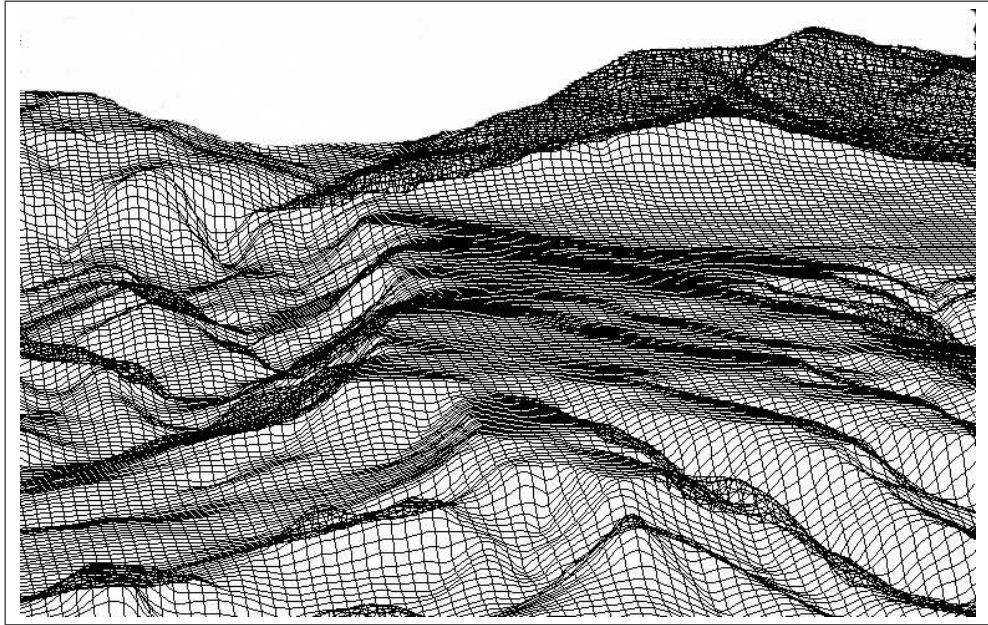


Figura 3.4: Superficie

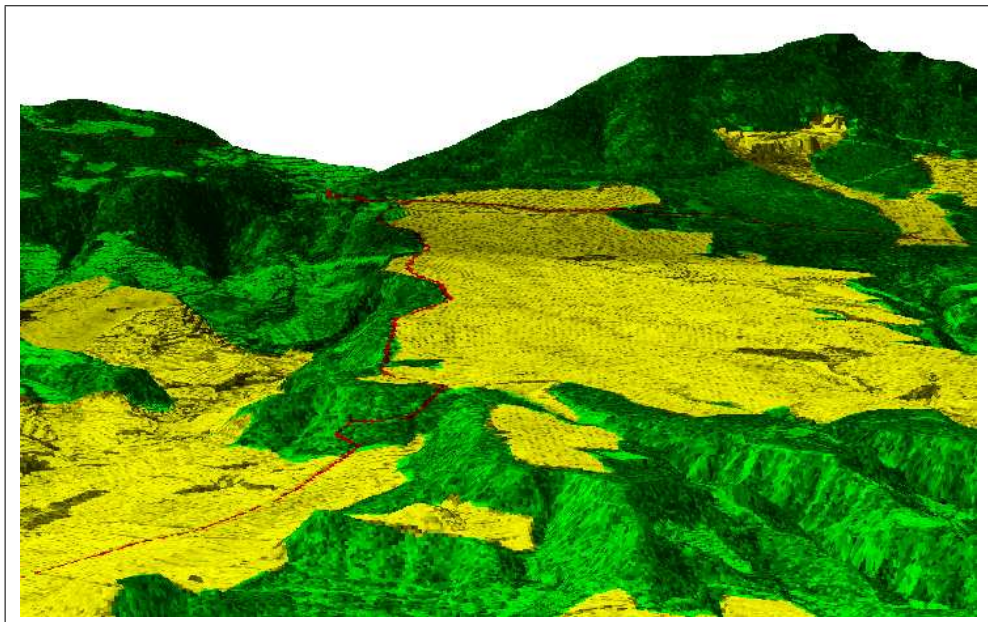


Figura 3.5: Objetos

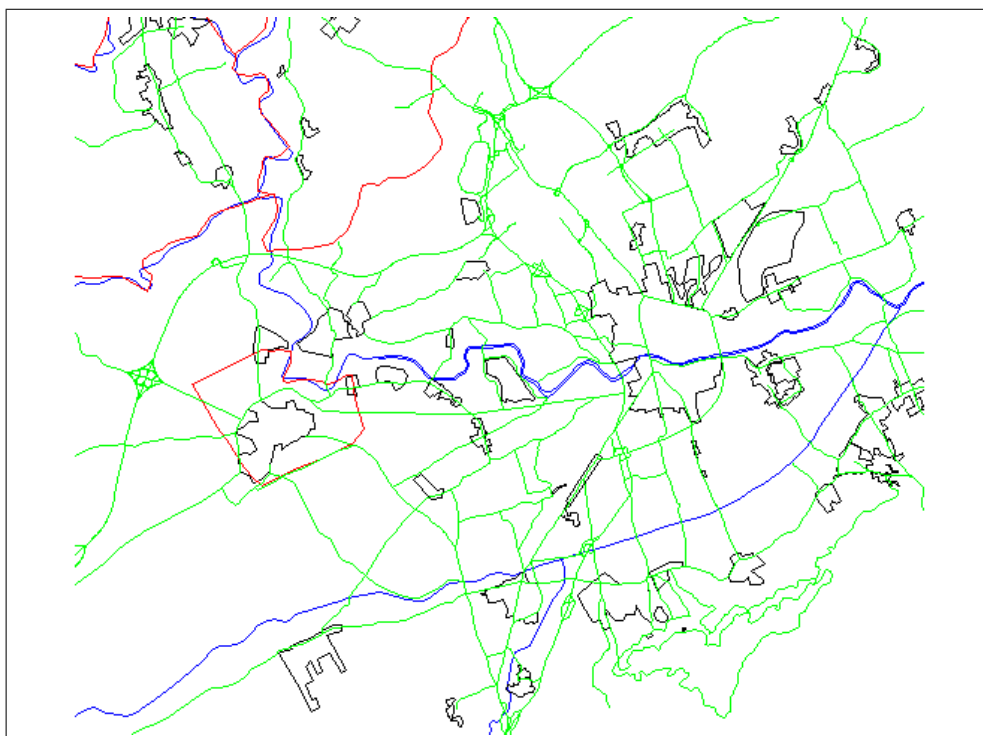


Figura 3.6: Representación de objetos (carreteras, red de drenaje, núcleos urbanos y límites municipales) en un SIG en las cercanías de la ciudad de Murcia



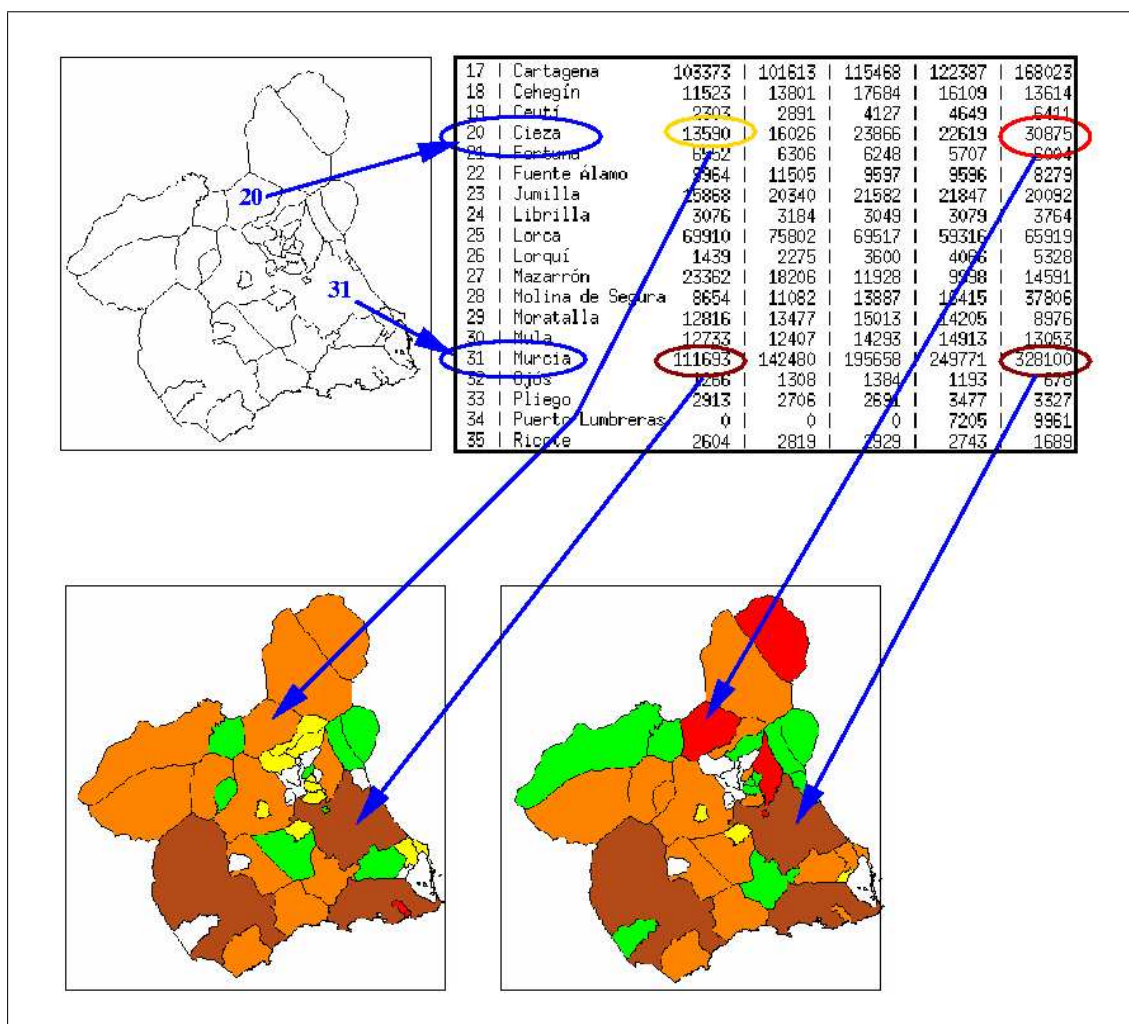


Figura 3.7: Capa de objetos (municipio) enlazada a una base de datos (información censal)



y los análisis verdaderamente 3D necesitan emplear modelos de datos diferentes y bastante más complejos que son de especial utilidad para algunas aplicaciones prácticas como en Geología, Oceanografía o Meteorología.

Mientras que el resto de los objetos son más o menos perceptibles sobre el terreno o utilizando la cartografía apropiada, las superficies suelen ser mucho más difíciles de determinar, ya que en cada punto del territorio los valores son diferentes, al no disponer de un valor para cada punto, es necesario realizar una estimación. Es necesario recurrir a técnicas de interpolación.

El ejemplo más típico de superficie es la elevación sobre el nivel del mar, representada mediante los Modelos Digitales de Elevaciones (MDE)<sup>5</sup>. Se trata de una superficie que representa la topografía del terreno, es decir, las alturas en cada punto de un territorio. Pero, en realidad, se puede crear superficies a partir de cualquier variable que cumpla unas mínimas características, esencialmente la continuidad espacial, sin que existan saltos bruscos en el valor de la variable. De este modo, diversos aspectos físicos naturales, tales como las precipitaciones, las temperaturas, la composición litológica o mineral, la acidez o basicidad de los suelos, etc., o también variables sociales: número de habitantes, densidad de población, etc., se pueden representar y analizar como una superficie.

### 3.1.2. Entidades

Si consideramos la realidad como una yuxtaposición de *objetos*, cualquier entidad que aparezca en el espacio (casas, carreteras, lagos, tipos de roca, etc.) puede modelizarse a la escala adecuada como un objeto extraído de la geometría euclidiana. Pueden ser clasificados en función del número de dimensiones en tres tipos:

- **Objetos puntuales** (figura ??a). Objetos geométricos de dimensión cero, su localización espacial se representa por un par de coordenadas (X,Y).
- **Objetos lineales** (figura ??b). Objetos geométricos de dimensión uno, su localización espacial se representa como una sucesión de pares de coordenadas llamados vértices, salvo el primero y el último que se denominan nodos (en la figura ?? aparecen los vértices en blanco y los nodos en negro).
- **Objetos poligonales**. Objetos geométricos de dimensión dos. Se representan como una línea cerrada (figura ??c) o como una sucesión de líneas denominadas arcos (figura ??d).

El escoger un tipo u otro para representar determinado objeto dependerá en gran manera de la escala y del tipo de abstracción que se pretenda hacer, de forma similar a lo que ocurre en la generalización cartográfica. Así una ciudad puede ser puntual o poligonal y un cauce fluvial lineal o poligonal. Una ciudad sólo tendrá sentido considerarla poligonal en estudios de planificación urbana. Para casi todas las aplicaciones hidrológicas tiene más sentido representar los cauces como objetos lineales y codificar su anchura y profundidad como propiedades espaciales.

Podemos considerar a priori 6 categorías de información que caracterizan a los diferentes entidades que aparecen en el espacio:

---

<sup>5</sup>La razón de la importancia de los MDE estriba tanto en su carácter indispensable como base territorial de un SIG como en la facilidad con que se puede medir la elevación de cualquier punto del territorio en comparación con otras variables regionalizadas como precipitación, humedad del suelo, etc.)

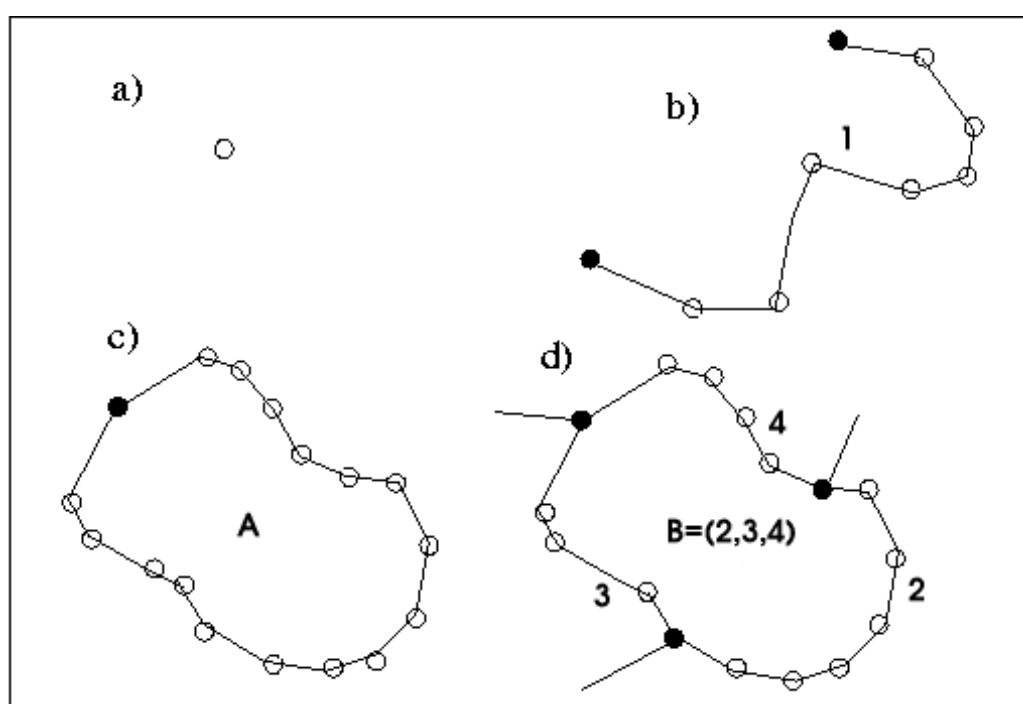


Figura 3.8: Tipos de objetos en formato vectorial a) Punto, b) Línea, c) Polígono en formato OO, d) Polígono en formato Arco-Nodo.

1. **Identificador.** Se trata de una variable cuantitativa que identifica cada objeto dentro de un conjunto de objetos del mismo tipo. Cada entidad recibe por tanto un identificador único.
2. **Posición.** Indica la ubicación del objeto en un espacio, generalmente bidimensional. Implícitamente indica también su dimensión y su forma. De este modo cada tipo de objeto tiene, en función de su número de dimensiones, una serie de propiedades espaciales de tamaño y forma directamente extraíbles de su codificación espacial:
  - Los objetos lineales tienen longitud, sinuosidad y orientación.
  - Los objetos poligonales tienen área, perímetro, elongación máxima y diversos índices de forma directamente calculables a partir de estas.
3. **Propiedades espaciales.** Son variables cuantitativas medidas en magnitudes espaciales y que indican algún aspecto de la extensión espacial de los objetos no representable debido a la escala de trabajo, a tratarse de una magnitud en la tercera dimensión o a la dificultad de representarla por el tipo de abstracción que implica su representación (por ejemplo la profundidad de un cauce o la anchura de una carretera).
4. **Propiedades no espaciales.** Son variables cualitativas o cuantitativas que no tienen nada que ver con el espacio pero que se relacionan con el objeto. Resultan de mediciones simples o de descripciones. Pueden ser constantes o variables en el tiempo. Por ejemplo toda la información relativa a la demografía de un municipio. Existen diversas operaciones que permiten derivar propiedades nuevas a partir de otras ya existentes.
  - Combinación aritmética:  $Densidad = Poblacion / Superficie$
  - Combinación lógica: Si  $Poblacion < x \ \& \ PIB > y \Rightarrow Riqueza = 1$
  - Reclasificación: Si  $Poblacion < 1000 \ \& \ Poblacion > 500 \Rightarrow Recl = 2$
5. **Relaciones con el entorno.** Todos los objetos geográficos tienen unas relaciones con su entorno, es decir con el resto de los objetos del mismo o distinto tipo que aparecen a su alrededor. Estas relaciones pueden ser de tipo puramente topológico (polígonos vecinos) o de tipo físico (cauces tributarios que se conectan al cauce principal). Pueden codificarse de forma explícita en la base de datos asociada al objeto o estar implícita en la codificación de su localización espacial. Estas relaciones pueden dar lugar a la creación de tipos compuestos (redes, mapas de polígonos, etc.).

### 3.2. Modelos lógicos. Formato raster y vectorial

El modelo lógico hace referencia a como se muestrean y organizan las variables y objetos para lograr una representación lo más adecuada posible. En un SIG existen básicamente dos modelos lógicos que se conocen como formato raster y formato vectorial y que dan lugar a los dos grandes tipos de capas de información espacial.

En el **formato raster** se divide el espacio en un conjunto regular de celdillas, cada una de estas celdillas contiene un número que puede ser el identificador de un objeto (si se trata de una capa que contiene objetos) o del valor de una variable (si la capa contiene esta variable).

En el **formato vectorial** los diferentes objetos se representan como puntos, líneas o polígonos (figura ??). La representación de puntos o líneas es inmediata, sin embargo al representar polígonos aparecen dos situaciones diferentes (figura ??):

- Si los polígonos aparecen aislados los unos de los otros, como en el caso de los núcleos urbanos, cada polígono se codifica como una línea cerrada, se trata de un modelo *Orientado al Objeto*, tal como aparece en la figura ??c en el que el polígono *A* se codifica como una única línea.

Si los polígonos se yuxtaponen, como en el caso de los términos municipales, codificar los polígonos como líneas cerradas tiene el problema de que habría que repetir cada una de las líneas interiores; por ejemplo el límite entre Murcia y Librilla (figura ?? debería introducirse dos veces, una al codificar Murcia y otra al codificar Librilla).

- El formato alternativo es el modelo *Arco-Nodo* con el que se codifican las líneas por separado y, posteriormente, se define cada uno de los polígonos a partir del conjunto de líneas que lo componen. Así en la figura ??d el polígono *B* se codifica como la unión de los arcos 2, 3 y 4. La codificación de polígonos con este modelo requiere por tanto dos etapas:

1. **Digitalización**, durante la que se introducen los arcos
2. **Reconstrucción de la topología**, durante el que se definen los polígonos y se crea la tabla que relaciona polígonos con arcos. La reconstrucción de la topología exige que la disposición de los arcos sea topológicamente correcta, así en la figura ??d los nodos iniciales y finales de los tres arcos deben coincidir exactamente.

La mayor virtud del modelo *Arco-Nodo* es ahorrar memoria, facilitar la digitalización y algunas de las operaciones de análisis SIG, siendo hoy día el modelo más utilizado.

### 3.2.1. Representación de superficies, variables cualitativas y entidades

Generalmente se considera que el formato vectorial es más adecuado para la representación de entidades o variables cualitativas y el formato raster para representar superficies. Sin embargo esto no es necesariamente así.

Para representar superficies podemos considerar hasta 4 modelos posibles (figura ??):

- Malla regular de puntos, a cada uno de ellos se asigna el valor de la variable medido en el punto.
- TIN (Red Irregular de Triángulos), los puntos se concentran en aquellas zonas donde la variable representada tiene mayor variabilidad.

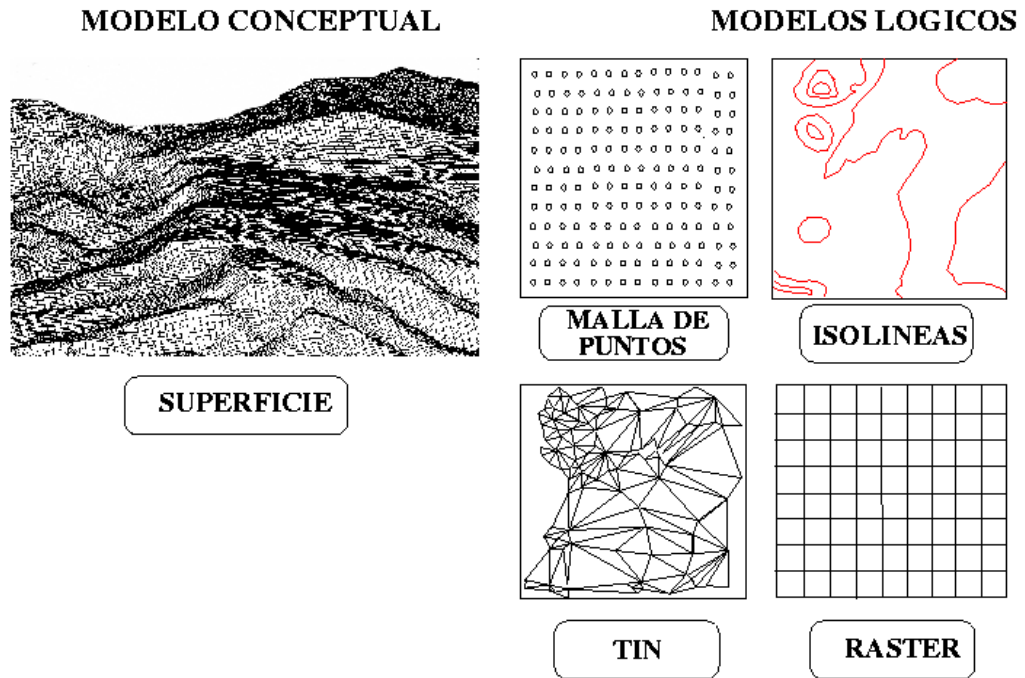


Figura 3.9: Modelos lógicos para representar superficies

- Isolineas (líneas en las que el identificador se sustituye por el valor de la variable).
- Raster, el área de trabajo se divide en celdillas.

Los tres primeros son representaciones en formato vectorial ya que se utilizan puntos o líneas para representar variables regionalizadas. El problema fundamental que plantean es que no representan a la totalidad del espacio, por tanto requieren una interpolación más o menos compleja para saber cual es el valor en un punto concreto. El modelo raster completa el espacio y la obtención del valor en cualquier punto es inmediata.

Para representar variables cualitativas existen dos alternativas (figura ??)

- Formato raster
- Formato vectorial Arco-Nodo
- Formato vectorial orientado a objetos, menos adecuado debido a que se introduce mucha información redundante

Para representar objetos las alternativas son (figura ??):

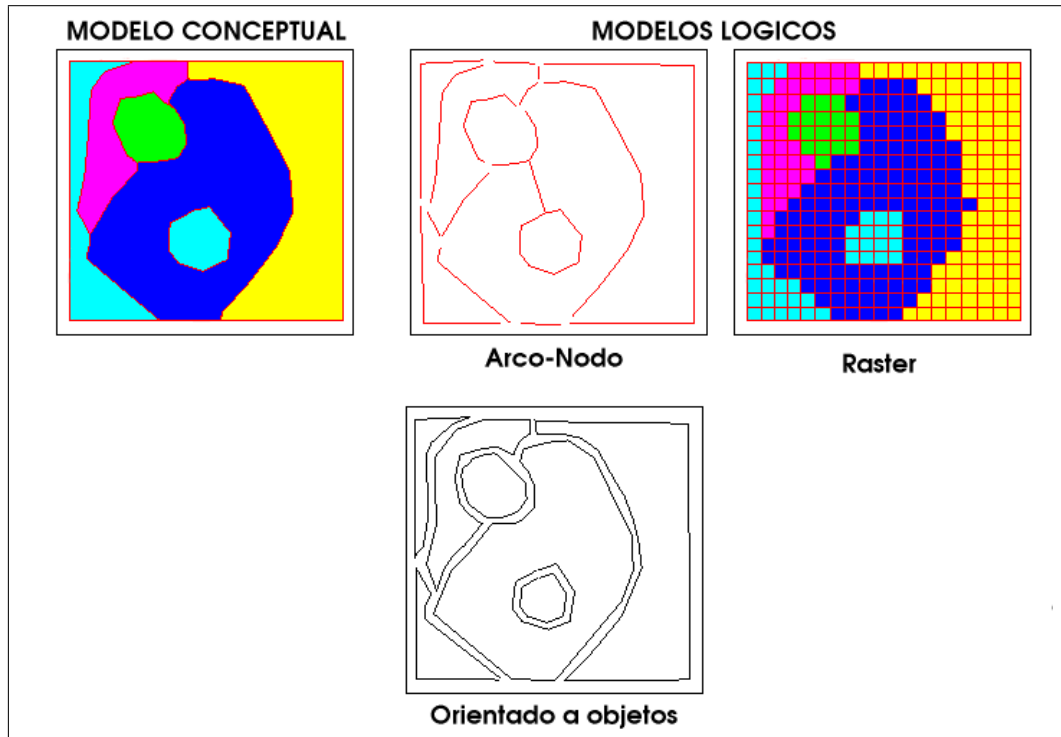


Figura 3.10: Modelos lógicos para representar variables cualitativas

- Formato vectorial Arco-nodo u orientado a objetos, si los polígonos no se tocan es lo mismo
- Formato raster, en una misma capa no pueden aparecer objetos que coinciden en el espacio

En el caso de variables cualitativas y objetos, el formato raster define de forma explícita el interior y de forma implícita el exterior, en el caso del formato vectorial es al revés.

### 3.2.2. Ventajas y desventajas de los formatos raster y vectorial

El debate acerca de la conveniencia de uno u otro modelo ha dado lugar a una abundante bibliografía, la decisión entre un modelo u otro debe, en todo caso, basarse en el tipo de estudio o enfoque que se quiera hacer, pero también del software y fuentes de datos disponibles.

Está claro que las superficies se representan más eficientemente en formato raster y sólo pueden representarse en formato vectorial mediante los modelos híbridos (mallas de puntos, TIN e isocurvas) que no resultan adecuados para la realización de posteriores análisis ya que todas las operaciones que permite el modelo raster resultarían mucho más lentas con el modelo vectorial. En general, cualquier tipo de modelización física de

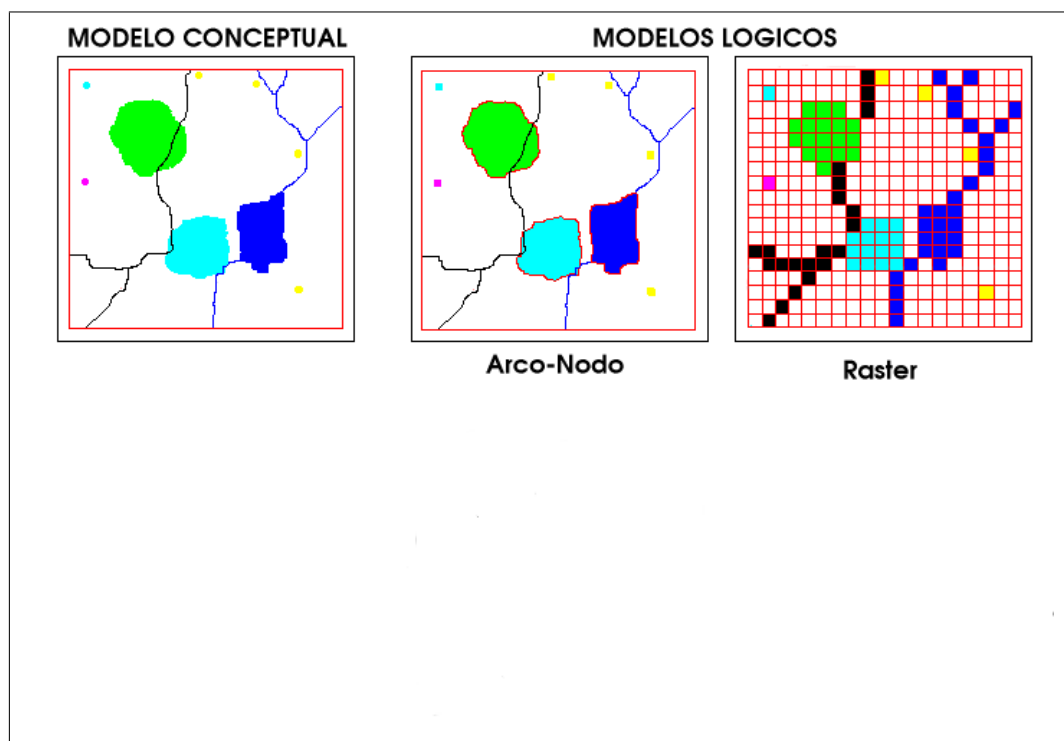


Figura 3.11: Modelos lógicos para representar entidades



procesos naturales que se base en Sistemas de Información Geográfica requiere un modelo de datos de tipo ráster.

Tradicionalmente se ha considerado que para la representación de los objetos resulta más eficiente la utilización de un formato vectorial ya que ocupa menos espacio en disco duro (aunque este último problema puede compensarse mediante diversos sistemas de compresión y en todo caso es cada vez menos relevante debido a la cada vez mayor capacidad de los discos duros) y los ficheros se manejan de forma más rápida si lo que se quiere es simplemente visualizar la capa. Sin embargo el formato vectorial es más lento que el raster para la utilización de herramientas de análisis espacial y consultas acerca de posiciones geográficas concretas.

En el caso de las variables cualitativas estaríamos en un caso intermedio entre los dos anteriores.

Las ventajas del modelo ráster incluyen la simplicidad, la velocidad en la ejecución de los operadores y que es el modelo de datos que utilizan las imágenes de satélite o los modelos digitales de terreno. Entre las desventajas del modelo ráster destaca su inexactitud que depende de la resolución de los datos y la gran cantidad de espacio que requiere para el almacenamiento de los datos. Este último problema puede compensarse mediante diversos sistemas de compresión. Además en muchos casos se confunde precisión y exactitud cuando se trabaja con datos vectoriales de modo que la exactitud en las coordenadas del modelo vectorial es más teórica que real.

Hoy en día se tiende a compaginar ambos modelos facilitada por el aumento en la capacidad de los ordenadores. Se trata de representar los diferentes fenómenos espaciales con el modelo de datos más apropiado en cada caso. En líneas generales se pueden codificar las formas en un modelo vectorial y los procesos con un modelo ráster, para ello se requieren herramientas eficaces de paso de un formato al otro. Resulta sencillo, finalmente, la visualización simultánea de datos en los dos formatos gracias a la capacidad gráfica actual.

Cada uno de los modelos lógicos puede implementarse de diferentes formas dando lugar a diferentes *modelos digitales* (formatos de ficheros) diferentes y por tanto incompatibles entre diferentes programas. Los formatos de ficheros pueden ser públicos (con lo cual resulta más sencillo desarrollar herramientas de importación-exportación) o privados (con lo que para implementar estas herramientas habría que pagar derechos). En los temas 4 y 9 se analizarán en detalle las características de los modelos digitales en formato raster y vectorial respectivamente.

### 3.2.3. Escala y modelos lógicos de datos

El concepto tradicional de escala en cuanto relación entre dos longitudes, no tiene sentido en un SIG. Las herramientas de zoom, disponibles en cualquier programa gráfico, permiten un cambio en la escala de representación en la pantalla o en una salida impresa, sin embargo este cambio de escala de representación no implica un cambio en la escala original de los datos.

En general podemos asumir que la escala de un SIG es la de los mapas que se han utilizado como información de entrada (en realidad la del mapa con escala inferior si se han utilizado varias). Sin embargo no toda la información de entrada procede de mapas (conjuntos de puntos de muestreo, imágenes de satélite, etc.). En el caso de los mapas de puntos la resolución se relaciona, aunque de manera algo difusa, con la distancia media entre los puntos.

En SIG podemos sustituir el concepto de escala por otro concepto, algo difuso, que es el de precisión espacial de los datos. En el caso del formato raster la precisión se relaciona claramente con el tamaño de las celdillas. Si hacemos un zoom excesivo aparecerán los bordes de estas como advertencia de que la profundidad del zoom es excesiva, en el caso del formato vectorial no tenemos este mecanismo de advertencia y en muchos casos se fuerzan los zoom para obtener una precisión completamente ilusoria.

### 3.3. Exactitud, precisión y error

En cualquier actividad científica o técnica es inevitable la existencia de errores. Puede tratarse de errores de medición o errores de transcripción, errores de posicionamiento o errores en la variable medida.

Tan absurdo resultaría negar el error como invertir esfuerzos más allá de lo razonable para conseguir pequeñas reducciones del mismo una vez que se ha conseguido una exactitud adecuada para los objetivos del trabajo. La actitud más razonable es tratar de estimar la magnitud de los errores cometidos, comunicarlo y tenerlo en cuenta para no pedirle a los resultados una precisión mayor de la que realmente pueden ofrecer.

En primer lugar hay que distinguir entre *precisión* y *exactitud*. Precisión es el detalle con el que un instrumento o procedimiento puede medir una variable mientras que exactitud es lo que se acerca esta medición al valor real, en SIG el concepto de precisión se relaciona con el de resolución en formato raster. El *error* es la diferencia entre el valor real y el medido, sin embargo puesto que el valor real nunca se conoce realmente, el error siempre debe estimarse.

Por ejemplo, una regla tiene una precisión de milímetro mientras que un metro de electricista tiene una precisión de centímetro. Sin embargo será más exacto medir un muro con un metro que con una regla ya que el instrumento es más apropiado.

En ocasiones los errores proceden de la utilización de ordenadores con precisión finita para manejar el espacio que es un continuo. Los errores en las operaciones en coma flotante de los ordenadores pueden afectar a los resultados, por ello es recomendable utilizar siempre números enteros. Por ejemplo, las coordenadas X e Y deberían expresarse siempre en metros nunca en kilómetros, los valores de las variables regionalizadas deberían expresarse de forma que no incluyeran decimales (altitud en centímetros, precipitación en décimas de milímetro, etc.). De esta manera además de prevenir errores conseguiríamos que los ficheros ocuparan menos en el disco duro.

### 3.4. Bibliografía

Gatrell, A.C. (1991) "Concepts of space and geographical data" en Maguire, D.J.; Goodchild, M.F. Rhind, D.W. (Eds.) *Geographical Information Systems: Principles and Applications*. John Wiley & sons pp. 119-134 ([www.wiley.co.uk/wileychi/gis/resources.html](http://www.wiley.co.uk/wileychi/gis/resources.html))

Moolenaar, M. (1998) *An introduction to the theory of Spatial Object Modelling for GIS*, Taylor & Francis, Londres, 246 pp.

