

MODELOS DE INFORMACIÓN GEOGRÁFICA

POR

JUAN A. CEBRIÁN

Modelos de base de datos

En la actualidad, para su implementación y mantenimiento, un Sistema de Información Geográfica (SIG) depende —como cualquier otro tipo de sistema informático— de la tecnología de manejo de *bases de datos* computadorizados. Hasta tal punto es esto cierto, que se puede afirmar abiertamente que *un SIG es una base de datos georeferenciados*.

Empezamos estas líneas con una descripción breve, pero completa, de los distintos paradigmas de organización de bases de datos que se han sucedido en el tiempo. Esta exposición servirá, cuando menos, para explicar los términos de la siguiente expresión, tan divulgada. «Desde los años sesenta hasta la actualidad hemos visto sucederse cuatro modelos de base de datos primordiales, a saber, el modelo jerárquico, el modelo reticular (*network*), el modelo relacional y el modelo objetual (*object-oriented*)». Dependiendo de cuál fuera el paradigma vigente en cada momento, se han diseñado SIG jerárquicos o reticulares, o relacionales u objetuales más recientemente.

Tsichritzis y Lochovsky (1982) son referencia obligada a la hora de explicar el alcance de los tres primeros modelos. Para presentar el cuarto,

Juan A. Cebrián. Departamento de Geografía. CSIC.

Estudios Geográficos
Tomo LV, n.º 215, abril-junio 1994

aunque existen otras muchas fuentes, vamos a utilizar una comunicación al *IGIS Symposium: The Research Agenda* (Egenhofer y Frank, 1987).

Una primera consideración acerca de estos cuatro modelos es que en todos ellos se recurre a la *tabla* como estructura básica de organización de la información. En el modelo jerárquico, y en el modelo reticular, las tablas representan las propiedades de las entidades de la base de datos. Dichas tablas (nodos), y las relaciones entre las entidades (arcos), constituyen el esqueleto (grafo) de la base de datos. El modelo relacional, por definición, sólo contempla tablas (que representan características y/o relaciones de las entidades de la base de datos). El modelo objetual, junto a otras peculiaridades que analizaremos más adelante, utiliza de nuevo tablas y arcos para representar propiedades y relaciones de los individuos de la base de datos.

El modelo jerárquico considera distintos conjuntos de información, definidos, cada uno, por un registro tipo (que equivale a la cabecera de una tabla de atributos y valores). Posteriormente se definen, si es el caso, las relaciones entre dichos conjuntos. Para definir una relación es preciso nombrarla, señalar qué dos conjuntos enlaza y determinar su sentido: qué conjunto es el origen (propietario) de la relación y qué conjunto es su término (miembro).

La relación autonomía-provincia en España, por ejemplo, puede ser inmediatamente representada en un sistema jerárquico porque no existe ninguna provincia que pertenezca a más de una autonomía. En cambio, el modelo jerárquico no permite, el modelo reticular tampoco, definir relaciones en las que un individuo del conjunto término pueda estar relacionado con más de un individuo del conjunto origen. Esta restricción es una consecuencia directa del propio concepto de jerarquía, en el que las relaciones entre los individuos de un nivel (origen de la relación) y los individuos del nivel inmediatamente inferior (destino de la relación) tienden a ser relaciones del tipo 1-n (ver Notarel).¹

Todo sistema jerárquico de base de datos requiere que la estructura de

¹ CEBRIÁN, J. A.: «El significado del término relación en la bibliografía sobre SIG», *Estudios Geográficos*, n. 213. ID.: «La matriz geográfica, casi treinta años más tarde», *Estudios Geográficos*, n. 214; ID.: «Sobre los conceptos de clasificación y regionalización», en este mismo número.

las relaciones ente conjuntos de elementos sea un árbol —la estructura de datos jerárquica por antonomasia (ver figura 1)— ordenado. En un árbol, por definición, sólo hay un camino entre dos nodos. Esta propiedad facilita enormemente cualquier tipo de búsqueda. En un árbol ordenado, además, se conoce siempre qué lugar, entre las diversas ramas (relaciones) descendientes de un nodo —ascendentes sólo hay una— ocupa una determinada relación.

El criterio de orden de las ramas descendentes de los nodos puede diferir de un nivel a otro, pero no dentro de un mismo nivel. El criterio puede ser zonal, para facilitar las búsquedas espaciales; alfabético, para acelerar las búsquedas nominales; categórico, para agilizar las búsquedas por tipos y propiedades de los datos, etc. Si se conoce el criterio de ordenación de un archivo, no es necesario escudriñarlo exhaustivamente en cada búsqueda, ya que lo habitual es llegar antes al punto del archivo donde se debe encontrar lo que nos interesa. Si se encuentra, hemos dado con lo que buscábamos; si no, es imposible que se encuentre en el resto del archivo no consultado todavía. La ordenación del árbol mejora la eficacia media de los procesos de recuperación de datos.

En la figura 1, a las ramas emergentes del nodo raíz (España) se les ha asignado un rango alfabético. De esta manera las demandas de información que se refieran al espacio andaluz serán atendidas recorriendo solamente la primera rama del árbol: el procedimiento de búsqueda no continuará indagando, si no encuentra alguna referencia a Andalucía en ese sector. En el segundo nivel del árbol los nodos tienen uno o dos descendientes, dependiendo de si alguna de las autonomías implicadas está organizada de manera diferente a la provincial. Cuando así ocurre, la primera rama aglutina las provincias y la segunda las divisiones administrativas equivalentes (comarcas, en nuestro ejemplo). En el tercer nivel, cada nodo del árbol emite cuatro arcos, que se corresponden respectivamente con los cuatro cuadrantes que definen el paralelo y el meridiano de Madrid. Esta ordenación de tipo espacial evitará, por ejemplo, que el sistema busque al sur de Madrid lo que se sabe que está al norte.

Un sistema jerárquico ofrece a su usuario las siguientes herramientas básicas. Una función de creación de registros tipo (tablas de atributos y valores) para almacenar por clases las propiedades de los elementos de información. Otra función de definición de relaciones, que requiere el nombre específico de la relación, el nombre del conjunto origen y el del

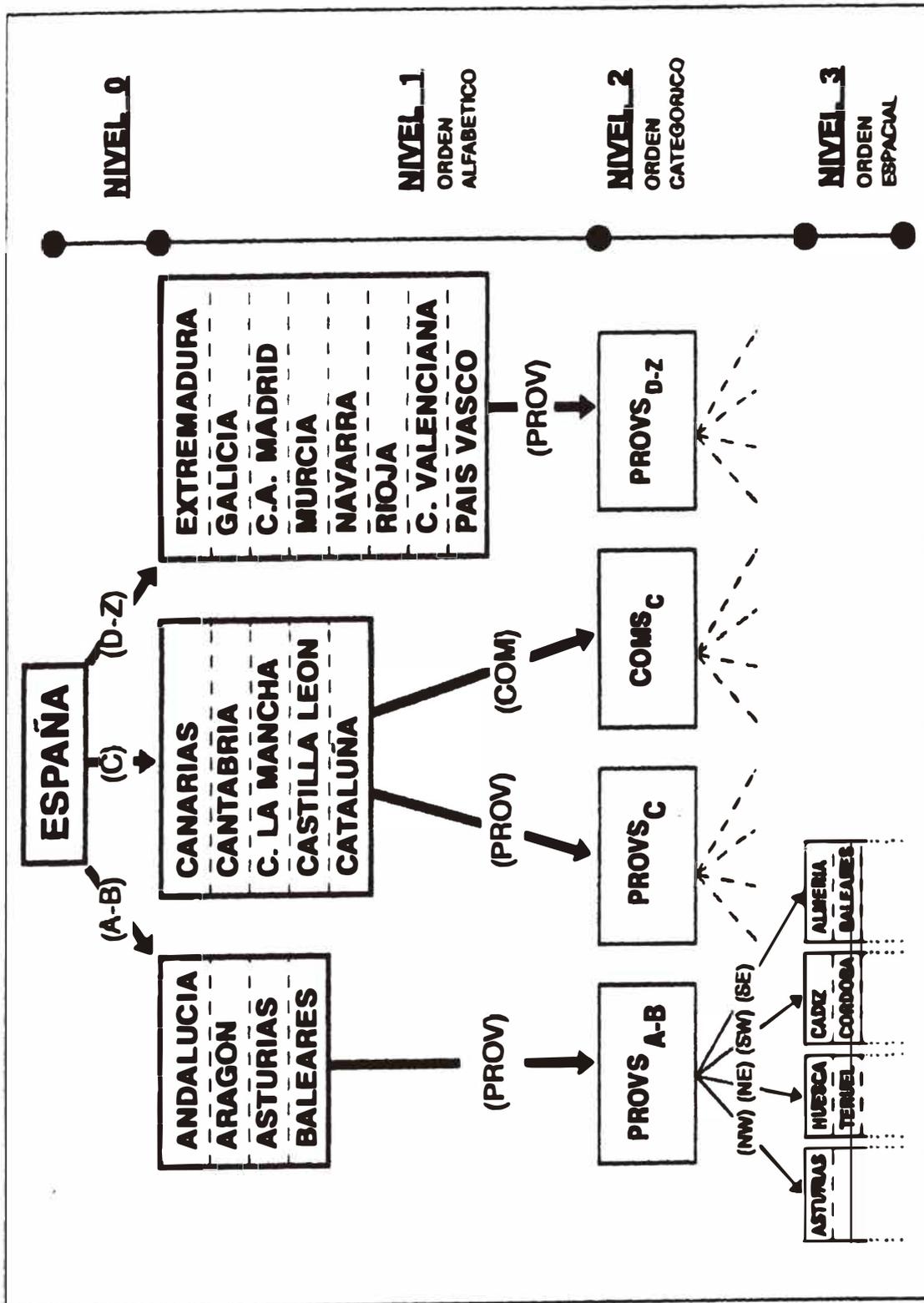


FIGURA 1.—La jerarquía administrativa territorial

conjunto destino. Una función de búsqueda de elementos, y de sus propiedades, en una tabla de nombre conocido. Una función de búsqueda de elementos, y sus propiedades, por su relación, conocida, con elementos de otra tabla ya identificada.

El modelo reticular contempla también conjuntos de información relacionados, definidos cada uno por un registro tipo. La versión más difundida de este modelo es la que se conoce con el nombre de modelo reticular DBTG (*Data Base Task Group*), formulado por primera vez en 1971 en la *Conference on Data Systems Languages (CODASYL)*. Este modelo permite un tratamiento de la información más flexible que el que ofrece el modelo jerárquico. la estructura de nodos (tablas) y arcos (relaciones) no necesariamente ha de ser un árbol, mucho menos un árbol ordenado. En la figura 2, por ejemplo, el modelo reticular permite manejar simultáneamente dos divisiones territoriales concurrentes, sin que haya que duplicar la información de base, que es la misma en ambos casos.

Además, el modelo reticular permite la representación de relaciones del tipo m-n, que se producen con cierta frecuencia en la realidad (ver Notarel). En la figura 3 presentamos un esquema simple de cómo se representaría en un modelo reticular la relación de abastecimiento entre proveedores e hipermercados. Para ello hay que crear una tabla intermedia, que en este caso es la que almacena la información necesaria para reconstruir las múltiples relaciones que pueden surgir entre proveedores e hipermercados. Acabamos de decir que una tabla almacena información referente a las relaciones entre elementos de dos conjuntos: estamos muy cerca, así es, de la formulación del modelo relacional.

El modelo relacional, dada su popularidad, requiere un tratamiento específico. Por ello ha sido abordado en una nota que aparece también en este número (ver Notarel). No obstante, en lo esencial, lo describimos a continuación. El modelo relacional sólo utiliza tablas, para representar las entidades del sistema y sus atributos, y las relaciones entre entidades y, cuando existen, los atributos de estas relaciones.

Las relaciones entre pares de entidades, pertenecientes a dos conjuntos distintos, pueden registrarse en una tabla en la que se reserven dos columnas para especificar los identificadores (nombres) de los individuos que estén propiamente relacionados. El resto de las columnas de la tabla puede dedicarse a los atributos de una de las entidades del par y/o de la

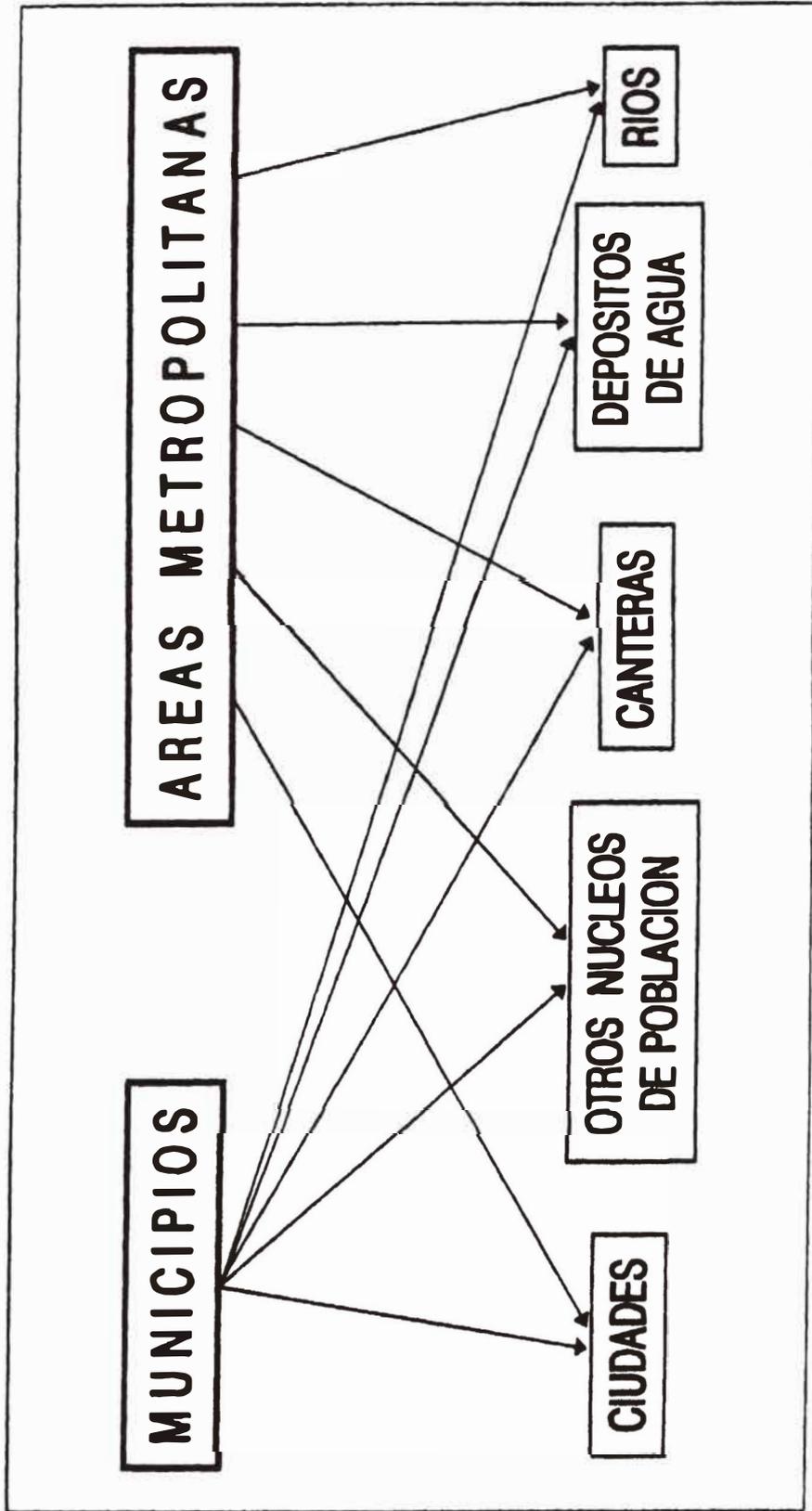


FIGURA 2.—Divisiones administrativas concurrentes

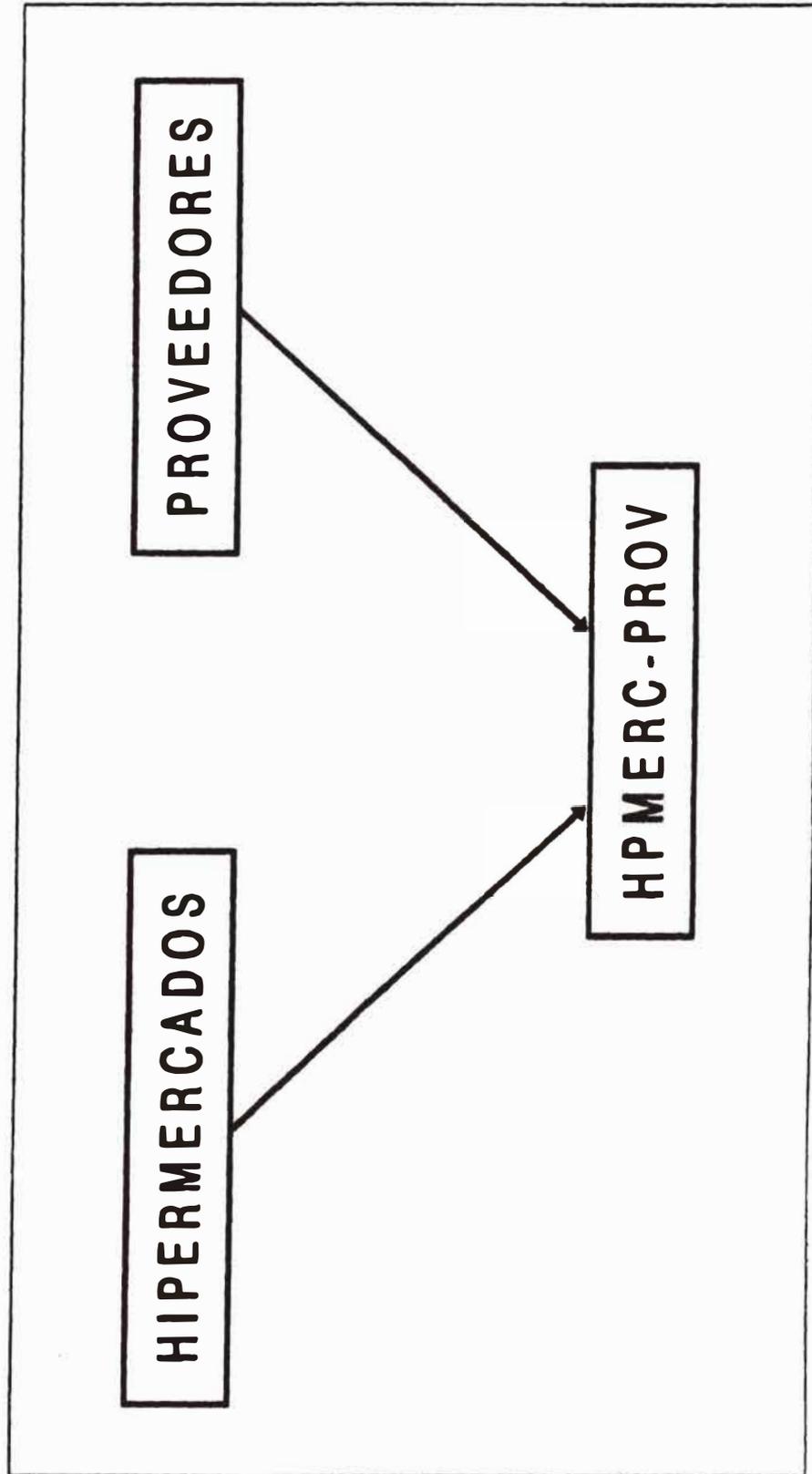


FIGURA 3.—*Hipermercados y sus proveedores. Un caso típico de relación m-n*

relación. Existen normas de diseño de bases de datos relacionales que aconsejan constreñir de alguna manera esta representación tan flexible (ver Notarel).

En la misma nota a que se alude en el párrafo anterior también se aborda la descripción del *modelo objetual*. A pesar de ello, vamos ahora a subrayar alguno de sus aspectos más interesantes. La célula básica de los sistemas objetuales es el objeto tipo, que es una tabla más compleja que las que hemos mencionado hasta ahora. De igual manera que una tabla es un conjunto de filas y de columnas, un objeto tipo también es un conjunto de filas o columnas (lo habitual es que se trate de conjuntos de filas), que representan objetos individuales semejantes.

Pero en la tabla de un objeto tipo no sólo se recoge la descripción de los objetos individuales, sino también información exhaustiva sobre su «comportamiento». En la jerga objetual, en lugar de hablar de llamadas a procedimientos con un determinado objeto como parámetro, se utiliza normalmente la expresión «hacer llegar un mensaje a» dicho objeto. De igual manera, los resultados de la ejecución del procedimiento invocado se consideran «reacciones» del objeto. Las tablas de objetos tipo se componen de una tabla de atributos y valores, y de una lista de los nombres y de las direcciones de todas las operaciones en que un objeto de ese tipo puede participar.²

Las relaciones entre distintos objetos individuales, pertenecientes a diferentes objetos tipo, se representan de forma parecida a cómo lo hemos visto hacer en los sistemas jerárquicos o reticulares. No ha de extrañar que con frecuencia se hable de «canales», en lugar de relaciones, en este tipo de entornos, ya que las relaciones son siempre vías de comunicación entre los individuos de una base de datos.

Dos relaciones genéricas tienen una importancia singular en la definición de una base de datos objetual: la relación *es_un/a* y la relación *parte_de*. La primera es la relación entre la especie y el género: la generalización (cuya inversa es la especificación). La segunda describe las relaciones que hay entre un objeto compuesto y sus partes elementales. Esta

² Por ejemplo, el objeto tipo punto, incluirá en su definición los nombres de las operaciones translación (p), proyección (p), vecindad (p), distancia (p,*), en_dirección (p,ϕ), etc., y sus direcciones, es decir, los puntos de arranque de las correspondientes instrucciones ejecutables por la máquina.

relación se repite en grados de complejidad creciente: por ejemplo, una línea quebrada está compuesta por una serie de puntos; un tramo de carretera es una línea quebrada entre dos intersecciones y una lista de sus atributos: tipo de vía (autopista, autovía, carretera nacional, etc.), perfil altitudinal, estado del firme, etc.; una carretera es un conjunto de tramos de carretera y una serie de atributos tales como su nombre o nombres, régimen y calendario de mantenimiento, régimen de explotación, etc.

En un sistema en el que no se hayan definido previamente el retículo de clases, que registra las relaciones de tipo «es_un/a», y el grafo de composición,³ que resume las relaciones de tipo «parte_de», no se puede introducir ningún dato. Este hecho tiene ventajas e inconvenientes. Las ventajas se comprenden fácilmente: no puede introducirse en la base de datos nada que no haya sido perfectamente conceptualizado —de acuerdo con los criterios de diseño del sistema—. No se puede llenar de contenido la tabla de un objeto tipo sin que previamente se haya situado dicho objeto en el retículo de clases. En esa operación el objeto tipo heredará todas las propiedades de los objetos de niveles superiores a los que está directa o indirectamente conectado (el objeto arroyo heredará automáticamente las propiedades, características y comportamiento, del objeto más genérico corriente de agua). El modelo objetual es, ante todo, un modelo estructurante.

Pero precisamente por ese carácter estructurante hay personas que lo acaban rechazando, porque prefieren trabajar con sus datos en un marco flexible que permita fácilmente corregir los errores previos y modificar las estructuras que ya no tienen vigencia.

Los sistemas objetuales aparecen como respuesta a una demanda de modelos de mayor capacidad semántica, mayor que la de una línea (sea fila, sea columna) de escritura dividida en campos. Así, estos sistemas tienden a ofrecer un tipo de diálogo en el que predomina la manipulación en pantalla, que permite expresar operaciones mediante cambios de configuración de ésta y de los iconos que representan a los objetos definidos. Este diálogo resulta mucho más expresivo que el que era habitual hasta

³ Si nos enfrentamos a una jerarquía estricta, aquella en la que cada clase sólo tiene una superclase inmediata, la estructura resultante es un árbol. Pero la realidad, con frecuencia, suele ser más compleja. Más común es que las relaciones de generalización y composición den lugar a otras estructuras, simplemente jerárquicas.

hace poco, caracterizado por los comandos línea, los formularios, los menús, las tablas, etc.

Modelos de datos

En el apartado anterior hemos analizado los diferentes modelos (o tipos generales) de bases de datos; algo así como los modelos de modelos de datos: los metamodelos, como suele decirse hoy en día. Ahora descendemos un escalón (un nivel de abstracción) para describir cómo construir modelos adecuados a espacios geográficos problema.

En primer lugar, nos permitimos un par de digresiones introductorias.

Resulta muy ilustrativo subrayar que los SIG, reflejando la mentalidad de sus usuarios, pueden crecer de lo particular a lo general, o a la inversa. En el primer caso, constructivista, se parte de unidades que se asume son elementales, simples. Pensemos, por ejemplo, en el individuo, en la unidad familiar, en la manzana, en la sección censal, en la estación climatológica, en la unidad geomorfológica, en el tramo fluvial, en las manchas de vegetación razonablemente homogéneas, en los puntos de venta de una cadena comercial, en los aparcamientos urbanos y en un largo etcétera. Con posterioridad a la definición de los elementos del sistema, se sondan sus características más representativas, que se introducen en tablas de atributos y valores. A continuación, y en base a toda esa documentación recogida, se inicia un proceso de agrupamiento o clasificación de los individuos introducidos en el SIG. Como resultado se obtienen clases espaciales de cierta coherencia interna, que pueden ser la base de la interpretación de un paisaje. Los elementos se agrupan a veces, a veces se combinan (componen, agregan...), dando lugar a nuevos elementos en espacios de complejidad creciente. En estos casos nos parece muy apropiado hablar de *SIG sintéticos*.

En sentido inverso se despliegan los *SIG analíticos* de cariz discriminante. Su descripción del espacio se hace por sucesivas aproximaciones. Partiendo de una primera zonificación, o esquema corológico, se lleva a cabo un progresivo refinamiento del cuarteado espacial, a medida que se va conociendo el espacio en cuestión.

Nada impide que un mismo SIG se construya siguiendo a la vez, en paralelo, ambas orientaciones. Las mismas funciones permiten construir

SIG sintéticos y SIG analíticos, porque la diferencia no se encuentra en la herramienta, sino en el uso que se hace de ella. En ambos casos, es más que conveniente disponer de un modelo de datos previo. No se debe comenzar la materialización de un SIG, en el que vamos a depositar gran parte del componente empírico de nuestras investigaciones, sin una idea clara de su estructura básica. Además, en caso de que el modelo definido en primera instancia no cumpliera su cometido adecuadamente, su adaptación siempre sería una tarea relativamente sencilla.

Llegamos ahora a la segunda digresión, ésta sobre la toma de datos geográficos. ¿Cómo se puede describir el espacio geográfico? En nuestra opinión, de dos formas: *directamente por medio de sus propiedades*; *indirectamente por sus elementos*, o partes, que a su vez serán descritos por sus propiedades y relaciones. La recogida tradicional de datos era fundamentalmente del segundo tipo. La aparición de la percepción remota digital ha abierto, de hecho, la primera posibilidad: la descripción de vastas zonas por la variación de características espaciales significativas. Todos los SIG de última generación contemplan ambos modos de descripción y disponen de procedimientos, más o menos eficaces, más o menos automáticos, de conversión de un modo a otro.

Terminadas ya las digresiones introductorias, vamos ahora a considerar las características propias de los modelos de datos de SIG concretos. Todo modelo de esta índole es una estructura que refleja las propiedades estáticas y dinámicas (atributos y procedimientos en los que participan) de los elementos del sistema, agrupados en clases, y las relaciones potenciales intraclases e interclases. Si se trata de un SIG que contiene, total o parcialmente, información espacial «en estado bruto» (conjuntos de valores pertenecientes a variables espaciales, imágenes digitales, etc.), el modelo de datos debe tener una estructura dual. Si ése es el caso, una de sus dos ramas aparecerá claramente articulada (como grafo de elementos, y sus propiedades, relacionados y categorizados). La estructura de la otra rama, en cambio, será tan flexible como se requiera, para asimilar la información obtenida directamente del paisaje.

Una comparación puede servirnos para entender el porqué de esta dualidad, cada vez más frecuente. Pensemos en el resultado de una semana de trabajo en el campo: unas mediciones precisas, un cuaderno de notas y una buena colección de fotografías. Las mediciones precisas podrán incorporarse inmediatamente a la base de datos del SIG, a su parcela

más estructurada. Como resulta obvio, no se pueden realizar mediciones precisas de propiedades de elementos que no hayan sido previamente conceptualizados. No resulta exagerado afirmar que a cada medición precisa le corresponde un hueco en las tablas de atributos y valores de los elementos del SIG correspondiente. Las notas y las fotografías se incorporarán a la otra rama de la base de datos del SIG, mínimamente estructuradas: por ejemplo, clasificadas por su localización aproximada. Es muy conveniente que un SIG contemple la posibilidad de incluir información escasamente interpretada, pero con el compromiso de que a los nodos ricos en información bruta se los vaya paulatinamente despojando de esa condición, para así incrementar el bagaje de conocimientos (tipos, propiedades y relaciones) sobre los elementos del paisaje.

Se podía pensar que el modelo de datos a que nos estamos refiriendo no es ni más ni menos que el archiconocido diccionario de datos de cualquier sistema, pero no es cierto. Aunque comparten una amplia banda de información, el diccionario de datos, por definición, no recoge información acerca de los procedimientos (que se especifican siempre en el diccionario correspondiente). El modelo de datos, en cambio, incluye esa información. Además, mientras que el diccionario de datos es fundamentalmente una descripción verbal de las propiedades de los mismos, un modelo de datos utiliza casi exclusivamente diagramas.

A continuación repasamos cuáles son las herramientas más adecuadas para la especificación de modelos de datos. Asumimos que un modelo de datos es una descripción diagramática de los siguientes conjuntos:

$O = \{\text{Clase de Objetos}_i, i = 1, n\}$

$A = \{\text{Atributos (Clase de Objetos}_i), i = 1, n\}$

$R = \{\text{Relaciones } (O_i, O_j), i \text{ y } j = 1, n\}$

$F = \{\text{Funciones } (O_i, *), i = 1, n\}$

$\text{MODELO DE DATOS} = \{O, A, R, F\}$

En el conjunto de relaciones destacan dos tipos fundamentales: las relaciones de generalización (técnicamente conocidas por relaciones «es_un/a») y las relaciones de agregación (relaciones «parte_de»).

La herramienta básica de diseño de modelos de datos es el *diagrama*

de entidades y relaciones, propuesto por primera vez por Chen en 1976, y modificado y parafraseado en innumerables ocasiones (ver Notarel).

El diagrama original reconoce tres tipos de bloques para su construcción: las entidades, sus atributos y las relaciones entre entidades. Cada tipo tiene una rúbrica concreta y expresiva. Las entidades se representan mediante rectángulos etiquetados con su nombre. Las relaciones aparecen como rombos unidos por trazos lineales a uno, dos, o más rectángulos, aprovechándose su interior para situar el nombre de la relación. Finalmente, los atributos de entidades y de relaciones se presentan como elipses etiquetadas, unidas por un trazo al rectángulo o al rombo correspondiente.

La inclusión de los procedimientos que afectan a las entidades de una base de datos puede llevarse a cabo ampliando los rectángulos que se refieren a las entidades, para que quepan los nombres de tales procedimientos. Otra posibilidad sería definir un nuevo símbolo para los procedimientos. Estos símbolos tendrían una etiqueta, su nombre, y estarían unidos a símbolos de entidades por dos tipos de trazos: los que se refieren al input de los procedimientos y los que se refieren a su output. Si la representación de todos los procedimientos es nociva para la comprensión del diagrama, se debe proceder a una selección de los mismos. Esta selección puede explotar las relaciones de generalización («es_un/a») existentes entre los conjuntos de datos del problema. Pongamos un ejemplo. Todas las entidades de un SIG deben incluir la referencia al test: «¿Se solapa espacialmente esta entidad con otra?» Si existe un grafo que registra todos los tipos de entidades y sus relaciones de generalización, en cuya raíz se encuentra la entidad «objeto geográfico», sólo es necesario hacer una anotación en el rectángulo correspondiente a esta última. Esta propiedad, como muchas otras, será heredada por los tipos de entidad más específicos.

Cuatro párrafos más arriba hacíamos mención a dos tipos de relación especialmente interesantes: las relaciones de generalización («es_un/a») y las de agregación («parte_de»). ¿Queda justificado, por su relevancia, utilizar algún signo distinto en el diagrama —distinto del rombo, se entiende— para representarlas? En nuestra opinión, sí. Consideremos primero la relación «es_un/a». En un proceso de generalización conceptual, lo normal es que el concepto inferior quede totalmente incluido en el superior, aunque también puede ocurrir que el concepto inferior (el más

específico) sea el resultado de la intersección de varios conceptos generales. En el primero de los casos, podemos representar la inclusión conceptual mediante la inclusión de rectángulos del diagrama, cuyo efecto visual es manifiesto. El mismo tipo de representación sería aplicable al segundo de los casos, situando el rectángulo del concepto más específico en la zona de intersección de los conceptos más generales. Nuestra propuesta, por tanto, recurre al diagrama de Venn, sacando partido de su expresividad y su casi universal aceptación.

Las relaciones de agregación («parte_de») pueden ser definidas de dos maneras. Para algunos autores, todas las relaciones, con excepción de las relaciones de generalización, son relaciones de agregación, o de composición. Para otros, entre los que nos encontramos, sólo son relaciones de agregación las que conectan un todo, que subsiste, y sus partes constitutivas. No son relaciones de agregación, sino relaciones meramente asociativas, los enlaces de individuos, de la misma o distinta clase.⁴ De acuerdo con esta distinción proponemos que las relaciones de agregación se representen mediante rombos incluidos dentro de rectángulos, para reflejar su carácter híbrido, relacional y sustantivo a la vez (ver figura 4).

La figura 4 recoge los elementos fundamentales para la construcción de diagramas de entidades y relaciones. Además, ilustra la existencia de tres tipos básicos de relaciones, que, siguiendo a Smith y Smith (1977), hemos representado como dimensiones ortogonales entre sí (ortogonalidad e independencia son conceptos algebraicos intercambiables). Al representar ortogonalmente los procesos de generalización, agregación y asociación estamos considerándolos independientes uno del otro, como así es. Si no, que se intente representar un número razonable de relaciones de distinto tipo en un mismo plano de dibujo, sin crear un caos visual. Smith y Smith sólo consideran las relaciones de generalización y de agregación. Para ellos, la dirección vertical representa las relaciones de agregación, y la horizontal, las de generalización. Bouillé (1978), sin utilizar exactamente estos términos, introduce en sus diagramas (estructuras) de datos una componente vertical, que él denomina jerárquica. *Mutatis mutandis*, esta dimensión coincide con los procesos de generalización definidos por Smith

⁴ No hace mucho tiempo (Cebrián, 1992, p. 45) desestimábamos este tipo de relación, por no disponer en ese momento de un marco conceptual tan sólido como el que respalda la redacción de estas páginas.

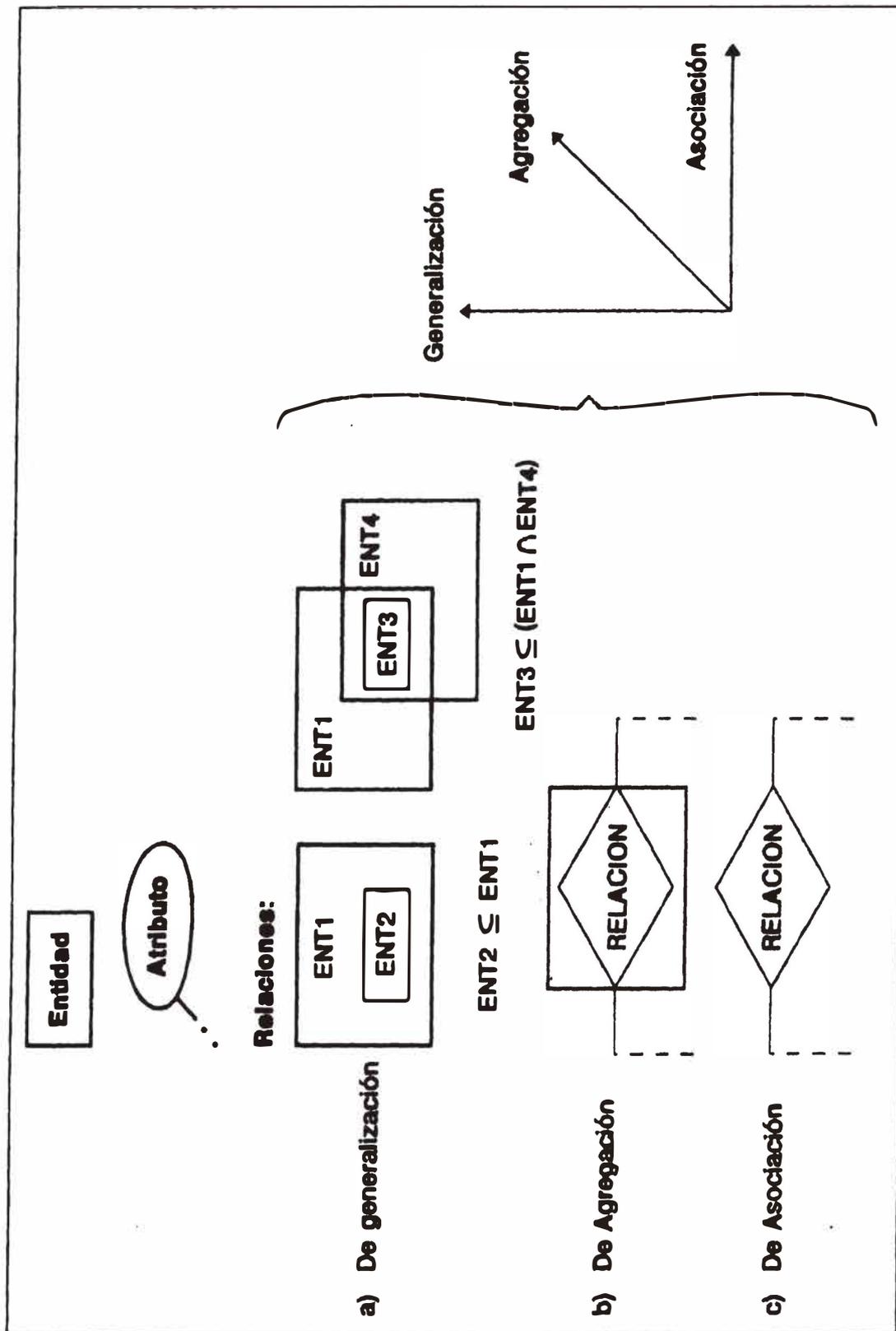


FIGURA 4.—Entidades, atributos y relaciones en un SIG

y Smith (1977). A continuación, en planos horizontales, Bouillé describe las relaciones de agregación, y también las asociaciones. Aun cuando asignen distintos ejes a los distintos tipos relacionales, las propuestas de Bouillé y Smith y Smith convergen, del conjunto de relaciones de un SIG, de cualquier base de datos suficientemente estructurada.

¿Son las relaciones espaciales otra, cuarta, dimensión del problema? En nuestra opinión, no. Las relaciones espaciales son un tipo de relación, más aún, en un SIG son el tipo de relación por antonomasia. Pero la categoría espacial no es una categoría independiente de las otras tres. Piénsese, en la existencia de relaciones espaciales de generalización (simplificación de formas), de agregación (de las comarcas que constituyen una determinada región, por ejemplo, o de los tramos de una determinada carretera) y de asociación (proximidad espacial, por citar alguna). Una relación es espacial porque espacial es, al menos, uno de sus términos.

En la tesitura de bosquejar un modelo de datos, ¿qué alternativas se nos ofrecen? La primera: representar en un solo plano de dibujo todas sus entidades, relaciones y atributos, recurriendo a la simbología de la figura 4. Cualquiera otra posibilidad lleva consigo la fracción del modelo de datos en dos o más submodelos. Por ejemplo, se pueden definir dos series de planos horizontales vistos en perspectiva, unos para la representación de las relaciones de generalización en dirección vertical y de las relaciones de asociación en los planos horizontales, y otros para la representación de las relaciones de agregación y asociación.

En todos los casos se observará una regla primaria de composición, que es la de ordenar entidades y relaciones en función de su complejidad. Además, a no ser que exista una imposibilidad de hacerlo así, se mantendrán las convenciones «naturales» por las que los humanos consideramos que los sentidos positivos de un sistema tridimensional cartesiano son respectivamente de abajo arriba, de izquierda a derecha y de atrás adelante. De acuerdo con ello las relaciones y las entidades de mayor complejidad se encontrarán siempre, o en el borde superior, o en el derecho, o en el fondo, de los diagramas de entidades y relaciones.

Otra posibilidad, que ya hemos expuesto con antelación (cf. Cebrián, 1992), es la de esquematizar las relaciones de generalización por una parte, recurriendo a un grafo orientado (retículo de clases) cuyo nodo más genérico sea el objeto geográfico, y las relaciones de agregación y asociación por otro (grafo de relaciones).

El grafo de relaciones puede estructurarse en dos ramas, correspondientes a las relaciones espaciales una de ellas, y a las no espaciales la otra. Subdividiéndose, a su vez, cada rama en otras dos, para representar asociaciones y agregaciones. El grafo de relaciones queda así estructurado en cuatro grandes conjuntos: asociaciones espaciales, agregaciones espaciales, asociaciones aespaciales y agregaciones aespaciales.

Todos los grafos de relaciones son importantes en un SIG, porque cada uno de ellos es un índice, una guía eficiente de búsqueda de información por criterios de proximidad (de proximidad espacial o de proximidad temática). No obstante lo cual no es imprescindible que en el SIG estén registradas explícitamente todas las relaciones posibles entre entidades, con tal de que lo estén las más consultadas (retículo de clases, relaciones de agregación y vecindad espacial, entre otras) y las que sea muy costoso recalcular, como es el caso de la mayoría de las relaciones del tipo m-n. Las demás relaciones se computarán a partir de los atributos de los individuos de la base de datos, cuando sea necesario.

Cuando se introduce información en un SIG, se pueden calcular las descripciones explícitas de las vecindades de todas las entidades espaciales a partir de sus localizaciones, es decir, a partir de sus coordenadas.⁵ El atributo localización, expresado mediante coordenadas, es la columna vertebral de todo SIG, porque es el que hace imposible la síntesis geocartográfica. No importa cual sea el sistema de coordenadas empleado, siempre que sea correcto, ya que todo SIG dispone de un módulo básico de transformación de coordenadas entre los sistemas más difundidos.

Conocidos ya los elementos de diseño de diagramas de entidades y relaciones, se impone un breve examen de cuáles son los pasos a dar, y en qué orden, para construir el primer borrador de un modelo de datos determinado (ver Notarel).

Lo primero es «arriesgarse» a definir las entidades, las entidades tipo, se entiende: las clases (en sentido lato) de entidades que el modelo va a

⁵ Nótese que no se debe introducir información espacial en un SIG sin registrarla previamente. Técnicamente hablando, el registro es una operación que combina dos conjuntos de datos espaciales (el primero respecto del segundo), para modificar el primero y hacerlo superponible (compatible) con el segundo. no se puede trabajar con dos planos temáticos que cubran una misma zona si éstos no han sido previamente registrados (uno respecto del otro) o ambos respecto de un tercer plano.

considerar. Para algunos, la palabra «arriesgarse» parecerá fuera de lugar. En cambio, las personas con cierta experiencia de trabajo en este campo la encontrarán claramente justificada. Y es que definir las entidades tipo de un modelo supone un riesgo, a veces más subjetivo que objetivo, porque trabajando con procedimientos informáticos es bastante más fácil reestructurar lo que no resultó acertado. Pero, en cambio, no se puede ignorar que el uso de sistemas de bases de datos computarizados puede generar un perfeccionismo estéril: pues se puede dar el caso de que se pierda mucho tiempo en afinar el instrumento, sin llegar a producir ningún sonido agradable a la concurrencia. En este caso como en todos es muy saludable asumir la estrategia de las aproximaciones sucesivas o la del *trial and error*.

La definición de entidades tipo es (propia, o impropriamente, depende de los casos) una clasificación de los individuos del modelo (ver Notaclas). Pero toda clasificación es modificable, bien porque sea incorrecta, o porque cambia el criterio de clasificación (porque cambia el punto de vista; porque cambia la escala del trabajo que da lugar a generalizaciones o a disecciones, etc.), o porque entran nuevos individuos en juego.

Una vez definidas las entidades hay que establecer qué atributos las caracterizan. También en este caso hay que olvidarse de la exhaustividad. Ya se completarán las tablas de atributos y valores a medida que se conozca mejor a los elementos del problema.

Los atributos de las entidades pueden ser conceptualizados como variables estadísticas; los atributos de entidades espaciales como variables estadísticoespaciales. Si las entidades definidas previamente marcan la pauta del muestreo, la variable se sondeará en los lugares ocupados por estas entidades. En estos casos, a cada una de las entidades del conjunto se le asigna un valor de los muchos que la variable puede adoptar. Esta operación se repite con cada uno de los atributos a registrar. De esta manera, a la tabla de atributos y valores se añade una nueva columna por cada atributo que se considere. Esta tabla recibe el nombre de matriz geográfica de datos (ver Notaest).

Pero también puede ocurrir, de hecho ocurre, que los atributos (variables espaciales) se sondeen en localizaciones que no tienen ninguna relación con las de un conjunto de entidades. Las localizaciones ocupan posiciones elegidas aleatoriamente, uniformemente, o según un patrón de

densidad variable (mayor número de muestras en el área donde el atributo es más inestable). Si la variable se ha sondeado en vértices de una malla ortogonal, su presentación es una matriz de números, que de ninguna manera hay que confundir con las matrices geográficas de datos. Para distinguirlas, las denominaremos variables espaciales bidimensionales (ver Notaest).

Una vez generadas las matrices geográficas y/o las variables espaciales bidimensionales que se estimen oportunas, se debe proceder a incorporar al sistema toda la información digitalizada acerca de las entidades en cuestión y del espacio en que se localizan. Ello llevará consigo «la importación» de listas de coordenadas asociadas a tablas de atributos y valores, de variables espaciales bidimensionales, de imágenes digitales sin interpretar.

Terminamos este apartado con una aclaración que creemos necesaria. Está muy difundida la idea de que existe una correspondencia total entre la descripción indirecta de las propiedades del espacio mediante características de las entidades espaciales (es decir, por matrices geográficas) y el así llamado modelo vectorial por una parte, y por la descripción inmediata de las propiedades del espacio por variables bidimensionales y el modelo raster por otro. Pero no es cierto. Por ejemplo, una descripción mediante matrices geográficas puede utilizar la metodología raster, basta que cada fila de cada tabla tenga una etiqueta que comparta con todas las celdas de un fichero raster a las que afecte la entidad correspondiente. O también, los puntos muestrales de una variable bidimensional, distribuidos regular o irregularmente, pueden estar descritos por un par de coordenadas, seguidas del valor oportuno, dando lugar a un fichero vectorial.

Funciones, operaciones, procedimientos, herramientas

A sabiendas de que no es factible ofrecer una relación exhaustiva de las funciones de un SIG, por haberlo intentado ya otras veces (Cebrián, 1987, 1990, 1992; Cebrián y Mark, 1986), nos parece oportuno dedicar un apartado de este artículo a enmarcarlas de nuevo, con el afán de que esta vez la explicación sea más rigurosa y, también, más eficaz. Como sugiere el título de esta sección, vamos a considerar intercambiables (sinónimos no son) los siguientes términos: funciones, operaciones, procedimientos y herramientas. Todos ellos tienen en común un significado agente sobre los datos de un SIG.

Las primeras herramientas que analizamos son las que permiten estructurar la información del sistema. En la mayoría de los campos se trata de un módulo de creación y manejo de tablas, que es el procedimiento más popular (más sencillo e intuitivo) de organizar las características de los elementos de un conjunto geográficamente significativo.

Junto a ese módulo de gestión de tablas, el SIG dispondrá de un gestor de información textual (un procesador de textos, básicamente) georeferenciada o georeferenciable. En la mayoría de los casos, si no en todos, los textos incluidos en un SIG describen lugares geográficos determinados. El gestor de textos del SIG estará preparado para atender las peticiones de información textual sobre lugares incluidos en un dominio geográfico, definido por coordenadas de algún tipo y/o por topónimos.

Gráficos e imágenes son otros tipos de información que los SIG debe estar preparados para gestionar. Al llegar a este punto es necesario llamar la atención sobre el hecho de que los SIG, como muchos otros sistemas de información, pueden manipular gráficos como si se tratara de meras reproducciones fotográficas (gráficos impresos o fotos aéreas y convencionales, fotos aéreas verticales y oblicuas, ortofotos, etc.), confiando al lector, así es el caso, la tarea de su interpretación. Pero además, los gráficos que maneja un SIG pueden ser ya el resultado de una interpretación de la realidad, por simplificación de la información almacenada en la base de datos (que es un modelo más preciso de la realidad pero, por ello, más difícil de manejar). Nos referimos a los mapas de más diverso tipo: una imagen digital procesada, un mapa de entidades de población de más de 2.000 habitantes, un mapa de carreteras comarcales y provinciales, etc. Una característica común a todos estos mapas es que sus partes son reconocibles por el sistema, porque han sido identificadas previamente, siendo precisamente el mapa una plasmación de ese conocimiento sobre el ámbito en cuestión.

Una vez introducidos en el SIG los datos, el usuario del sistema necesita un mecanismo sencillo para establecer relaciones entre los objetos de estudio o análisis. Entre las relaciones geográficas más populares, e interesantes, se encuentran los flujos espaciales de toda índole, físicos o económicos, visibles o invisibles, de personas o de mercancías, flujos de capital, flujos telefónicos y un largo etcétera. Lo habitual será definir estas relaciones mediante tablas que registren todos los pares de nodos conectados, o mediante matrices cuadradas que almacenan en cada celda el valor

de la interacción existente entre su nodo fila y su nodo columna. Esta forma de describir las relaciones directas entre individuos permite calcular también sus relaciones indirectas. Si las relaciones han sido descritas por tablas, la operación será una fusión de tablas que compartan al menos un atributo. Si las relaciones han sido registradas como matrices cuadradas, la operación será una multiplicación de matrices.

Una operación común a todos los SIG, la primera y primordial, es la búsqueda, flexible y eficaz, de datos en el sistema. Cualquier persona puede estar interesada en localizar objetos por su nombre, sus propiedades o sus relaciones, con la ayuda de tablas y mapas índice. Ésta es la porción del sistema que atiende a las demandas del gran público, que no conoce, ni tiene por qué conocer, los intrínquilis de la base de datos.

Hasta aquí, las herramientas que los SIG ofrecen para archivar y gestionar grandes masa de información, y para seleccionar los datos que demanden múltiples usuarios. La respuesta del SIG debe ser, por defecto, una combinación de gráficos y texto, pues difícilmente se puede atender una pregunta relacionada con el territorio sin entregar mapas, fotos aéreas, croquis, diagramas, etc.

Más allá de la presentación gráfica y textual de información de base, los SIG están equipados con numerosas funciones que permiten la construcción algebraica de soluciones. La superposición de mapas (*overlay*) y las funciones de proximidad son las más frecuentemente utilizadas.

Las funciones de superposición de mapas pueden tener un resultado puramente visual, justificable por sí mismo. En la pantalla de un monitor gráfico se visualizarán distintos estratos de información, con la pretensión de que determinadas combinaciones de colores, o de tonos de gris, llamen la atención de la persona que está dialogando con el sistema, facilitándole el descarte de zonas extensas que no tengan relación alguna con el problema a resolver.

Más interesantes todavía son las funciones de superposición que tienen un resultado no meramente visual. En la actualidad ya no se pueden construir modelos espaciales verosímiles sin recurrir en algún momento a este tipo de procedimientos. Las posibilidades son innumerables. Por citar un ejemplo, piénsese en una capa de información que contenga el trazado de varias carreteras. Mediante transformaciones sucesivas, conceptualmente simples, se puede generar un mapa de los lugares que se encuen-

tran a lo sumo a una determinada distancia de una carretera. Posteriormente, ese mapa puede utilizarse, como una máscara, para analizar otros mapas, otras propiedades del espacio, en las proximidades de las vías rápidas de comunicación.

La interpolación espacial es una de las funciones de proximidad más notables. Se invoca a ella para obtener una estimación del valor en un punto del espacio de una determinada variable. Para ello, los procedimientos de interpolación analíticos investigan qué valores conocidos se sitúan en las proximidades y calculan un valor en función de ellos y de sus distancias al punto en cuestión.

Tareas

Aunque aparentemente, sólo aparentemente, la flexibilidad de las funciones de un SIG hace posible que sea mantenido por pocas personas, capaces de desarrollar tareas muy diversas, la realidad es que un SIG funciona si está atendido por un número adecuado de profesionales especializados en determinadas parcelas el espectro, tan dilatado en este caso, que abarca esta tecnología.

El organigrama de tareas de un SIG es una estructura jerárquica, acorde con su objetivo primordial: sintetizar información empírica para su conocimiento generalizado, seguido en muchos casos de una acción coherente con él.

En la base del organigrama (ver figura 5) se localizan las tareas de *recogida de información*. Normalmente, un SIG no incluye entre sus objetivos los trabajos de campo, que están encomendados a otro tipo de instituciones (institutos cartográficos, estadísticos, geologicomineros, agrarios, forestales, etc.). Consecuentemente, las personas que trabajan en el nivel de introducción de datos en el sistema suelen ser grabadores en soporte magnético de información presentada en forma de imágenes, gráficos, croquis, mapas y tablas.

Cada vez es más frecuente que las agencias que se encargan de la recogida de datos en el campo entreguen sus productos en un soporte compatible con el ordenador. Ello es así porque ya existen muchos aparatos de grabado digital de la información en lugares de recogida. También, porque estas agencias tienden a absorber las tareas, y las personas que las

llevan a cabo, de grabación de la información, bien recurriendo al personal del propio instituto o, cada vez más frecuentemente, a contratar empresas pequeñas que trabajan a destajo. La tendencia, por tanto, es que los SIG sean nutridos, cada vez en mayor medida, desde el exterior.

El siguiente nivel sería el de la *verificación de la información* que se va a incorporar al sistema. Se trata de filtrar los errores que toda masa de información acarrea. No existen filtros normalizados, porque cada SIG, en función de sus objetivos, tiene distintos umbrales de tolerancia. Los filtros se basan en el principio de continuidad de la información, subrayando, para su posterior eliminación a juicio de un operador (a veces automáticamente), los valores excepcionales. También se pueden filtrar datos para eliminar la información redundante (duplicada, triplicada...) y la escasamente significativa. En los procesos de verificación los operadores cuentan con muchas y muy sofisticadas ayudas visuales (gráficos por computadora). Como se aprecia en la figura 5, entre el estrato de recogida de información y el de verificación de los datos de partida hay una interacción circular, ya que al flujo de entrada de información se opone el de rechazo de la información de baja calidad, que genera una demanda de nueva información sobre el mismo tema.

Los restantes niveles del SIG serían dedicados a la estructuración de la información verificada, al análisis de la misma, recurriendo a los más diversos métodos, y a la creación de modelos. En todos los casos se recurre continuamente a las prestaciones de los módulos de presentación: alfanumérico, gráfico y cartográfico.

Brevemente diremos que *estructuración de la información* verificada es la tarea de agrupar entidades y variables en bloques significativos, y diseñar índices eficaces que puedan ser explotados por los procedimientos de búsqueda, para localizar a gran velocidad cualquier átomo de información requerido.

Cualquier manipulación de la información que sobrepase su almacenamiento y recuperación es un procedimiento de *análisis*. Las tareas de análisis tienden a conocer la estructura de los fenómenos y sus relaciones: tienen un carácter desarticulador. Lo contrario hay que decir respecto a la construcción de *modelos*, que no necesariamente han de ser dinámicos. Un modelo es como un rompecabezas que recuerda, explica, simula... la realidad, siempre mucho menos complejo que ella, siempre de dimensio-

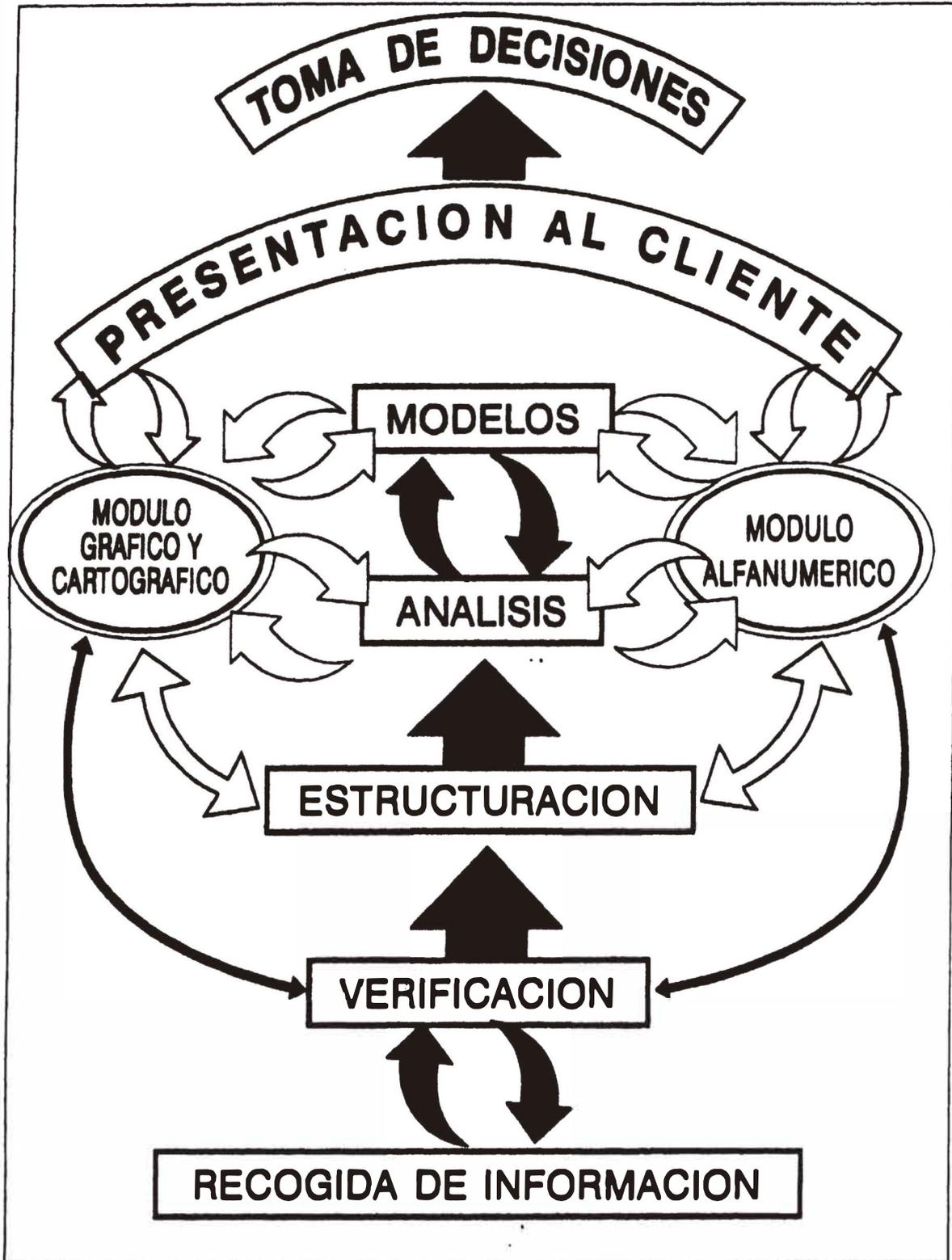


FIGURA 5.—El organigrama de tareas de un SIG

nes más reducidas (no en vano en inglés modelo significa primordialmente maqueta: modelo a escala).

Aún cuando en el organigrama parece que no hay vuelta atrás una vez verificados los datos (solamente en el vértice se produce un bucle que enlaza las tareas de análisis y construcción de modelos), si se presta atención, se descubre que cualquier trayectoria es posible pasando por los *módulos de presentación del sistema*. En distintos momentos y, por supuesto, en el de presentación al cliente se puede tener que demandar todo el trayecto hasta volver a la recogida de nueva información.

Merece la pena hacer una última consideración, antes de terminar este apartado. Unos párrafos arriba señalábamos que el acceso a un SIG no reduce la necesidad de una experiencia profesional contrastada, todo lo contrario. Pero esto no es obvio para muchos usuarios de esta tecnología. Ello ha causado situaciones poco deseables, como la de *los cartógrafos en una hora*. Tampoco es infrecuente el caso de la persona que aplica *un método tras otro* (según el *menú* que le presenta el SIG), a un conjunto de datos recogidos por *quién sabe*, para resolver un problema que no puede definir en pocas palabras. Nuestra experiencia nos dicta que el uso de SIG es muy útil para el que sabe de antemano qué quiere hacer. Si, además, eso que se quiere hacer es interesante, el SIG facilitará enormemente la elaboración de interesantes proposiciones.

Agradecimientos

Esta publicación es el resultado de mi participación en el Curso de Cartografía Automática y SIG, celebrado en Laredo en septiembre de 1992. Por ello quiero expresar mi agradecimiento a Pedro Reques (su organizador), a los otros profesores y a los alumnos del curso. Por otro lado, quiero hacer constar la ayuda de Jesús Monge (ilustraciones) y Rafael Rubio (mecnografía) en la preparación del texto original.

BIBLIOGRAFÍA

- BOSQUE, J.: *Sistemas de Información Geográfica*. Madrid, 1992, Rialp, 451 pp.
BOUILLE, F.: «Structuring Cartographic Data and Spatial Processes with the Hypergraph-based Data Structure», en DUTTON, G. ed., *Harvard Papers on Geographic Information*

Systems, Harvard University, 1978, Lab. for Computer Graphics and Spatial Analysis, vol. 5, 16 pp.

- CEBRIÁN, J. A.: «Una categorización de las principales funciones de establecimiento y explotación de sistemas de información geográfica», *Anales de Geografía de la Universidad Complutense*, 1987, n. 7, pp. 29-41.
- «Gramática de Comunicación con Sistemas de Información Geográfica», *Estudios Geográficos*, 1990, n. 199-200, pp. 381-396.
- *Información Geográfica y Sistemas de Información Geográfica (SIGs)*, Santander, 1992, Servicio de Publicaciones, Universidad de Cantabria, 85 pp.
- CEBRIÁN, J. A. y MARK, D. M.: «Sistemas de Información Geográfica. Funciones y estructuras de datos», *Estudios Geográficos*, 1986, n. 184, pp. 277-299.
- CHEN, P. P.: «The Entity-Relationship Model: Towards a Unified View of Data», *ACM Transactions on Database Systems*, 1976, vol. 1, n. 1, pp. 9-36.
- EGENHOFER, M. J. y FRANK, A. U.: «Object-Oriented Database Requirements for GIS», *IGIS Symposium: The Research Agenda, Proceedings*, Arlington, Virginia, 1987, pp. II-189 a II-211.
- SMITH, J. M. y SMITH, D. C. P.: «Database Abstractions: Aggregation and Generalization», *ACM Transactions on Database Systems*, 1977, vol. 2, n. 2, pp. 105-133.
- TSICHRITZIS, D. C. y LOCHOVSKY, F. H.: *Data Models*, Englewood Cliffs, N. J., 1982, Prentice-Hall Inc., 380 pp.

RESUMEN.—*Información Geográfica y Sistemas de Información Geográfica*, 2. En este artículo pasamos revista a algunos de los aspectos cruciales de la tecnología de los Sistemas de Información Geográfica (SIG).

Hemos estudiado los diferentes modelos de datos espaciales y las relaciones entre los SIG y los sistemas generales de manipulación de información, subrayando los aspectos conceptuales más importantes.

PALABRAS CLAVE.—Información geográfica. Sistemas de información geográfica. Tratamiento de datos.

ABSTRACT.—*Geographic Information and Geographic Information Systems*, 2. A review of some crucial aspects of Geographic Information Systems (GIS) technology is given.

The different spatial data models, the relationships between GIS and general database management systems, as well as several GIS procedures are discussed. The approach is logical, i.e. conceptual, for the most part.

KEY WORDS.—Geographic Information. Geographic Information Systems. Spatial data.

RÉSUMÉ.—*Systèmes d'Information Géographique*, 2. Dans cet article nous revisons importants aspects de la technologie des Systèmes d'Information Géographique (SIG).

Nous étudions les différents modèles des données spatiales, les relations entre les SIG et les systèmes généraux de manipulation d'information, ainsi que plusieurs procédures de SIG. Les aspects théoriques sont particulièrement soulignés.

MOTS CLÉ.—Information géographique. Systèmes d'Information Géographique. Données spatiales.