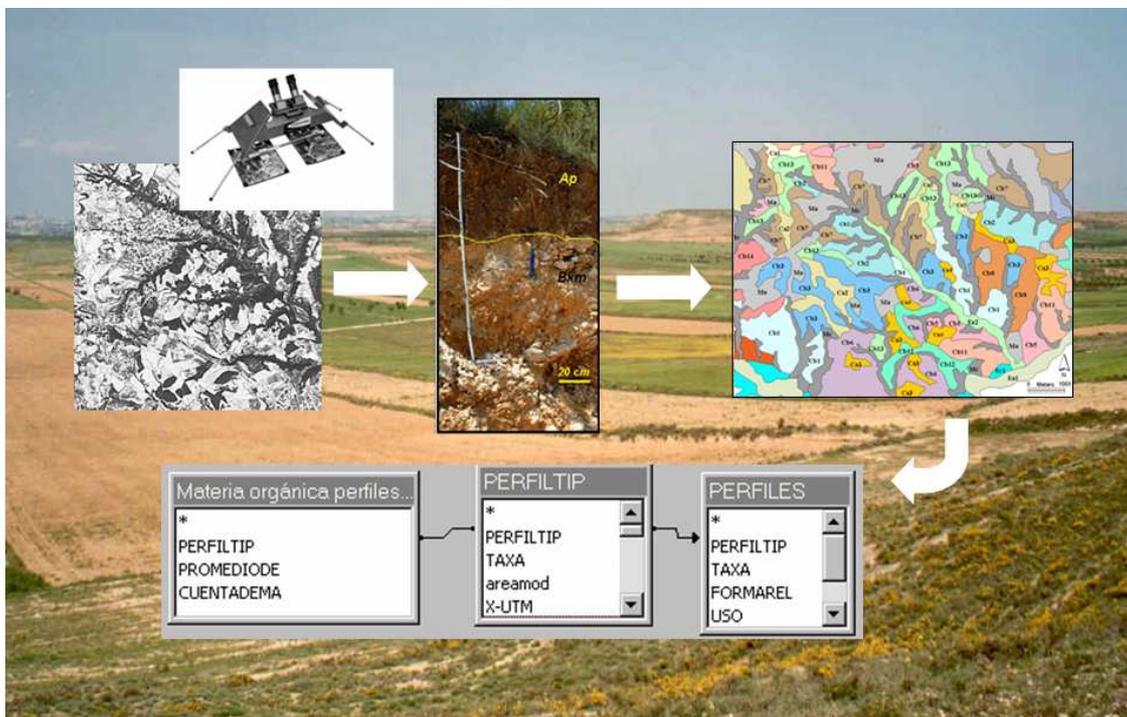


BASES DE DATOS GEO-RELACIONALES

Aplicación al diseño, implementación y consulta de un Sistema de Información de Suelos



José Antonio Martínez Casasnovas



Universitat de Lleida
Departament de Medi Ambient
i Ciències del Sòl

Título: Bases de datos geo-relacionales. Aplicación al diseño, implementación y consulta de un Sistema de Información de Suelos

Serie: Quaderns DMACS núm. 35

Autor: José Antonio Martínez Casasnovas

Editor: Universitat de Lleida, Departament de Medi Ambient i Ciències del Sòl

Primera Edición: 100 ejemplares

Imprime: Copisteria Universitaria Armengol

Depósito Legal: L-1263-2010



Esta obra está licenciada bajo la Licencia Creative Commons Atribución-NoComercial-SinDerivadas 3.0 España. Esta licencia no permite modificaciones de la obra ni usos comerciales. Para ver una copia de esta licencia, visita <http://creativecommons.org/licenses/by-nc-nd/3.0/es/>.

ÍNDICE

| | | |
|---------|---|----|
| 1.- | INTRODUCCIÓN | 4 |
| 2.- | MODELIZACIÓN DE DATOS | 4 |
| 2.1.- | Modelización externa..... | 5 |
| 2.2.- | Modelización conceptual | 8 |
| 2.2.1.- | Modelo jerárquico | 9 |
| 2.2.2.- | Modelo en red | 11 |
| 2.2.3.- | Modelo Relacional | 11 |
| 2.2.4.- | Modelo orientado al objeto | 12 |
| 2.3.- | Modelización interna | 13 |
| 3.- | DISEÑO DE BASES DE DATOS RELACIONALES | 13 |
| 3.1.- | Tablas | 14 |
| 3.1.1.- | Normalización de tablas..... | 15 |
| 3.1.2.- | Identificadores..... | 20 |
| 3.2.- | Relaciones | 20 |
| 3.2.1.- | Características generales..... | 21 |
| 3.2.2.- | Grado y obligatoriedad de una relación..... | 21 |
| 3.3.- | El modelo entidad-relación de la base de datos | 25 |
| 4.- | ÁLGEBRA RELACIONAL | 27 |
| 4.1.- | Concepto de álgebra relacional..... | 27 |
| 4.2.- | SQL..... | 27 |
| 5.- | EJEMPLO DE APLICACIÓN DEL DISEÑO DE BASES DE DATOS RELACIONALES: BASE DE DATOS DE SUELOS PARA ESTUDIOS DE EROSIÓN Y EVALUACIÓN DE TIERRAS | 30 |
| 5.1.- | Introducción | 30 |
| 5.2.- | Modelización conceptual de suelos y paisaje | 30 |
| 5.2.1.- | Componentes del mapa de suelos y relaciones..... | 31 |
| 5.2.2.- | Modelo conceptual de la base de datos de suelos | 33 |
| 5.3.- | Entidades y atributos (tablas esqueleto)..... | 36 |
| 6.- | PRÁCTICA. DISEÑO E IMPLEMENTACIÓN DE BASES DE DATOS RELACIONALES. APLICACIÓN A UNA BASE DE DATOS DE SUELOS. | 38 |
| 6.1.- | Objetivos de la práctica:..... | 38 |
| 6.2.- | Parte 1: diseño de la base de datos..... | 38 |
| 6.3.- | Parte 2: implementación y consulta de la base de datos | 42 |
| 6.3.1.- | Realización de consultas complejas mediante ACCESS. Consultas a partir de una tabla. 42 | |
| 6.3.2.- | Realización de consultas complejas mediante ACCESS. Consultas a partir de más de una tabla..... | 46 |
| 6.3.3.- | Otras consultas..... | 47 |
| 7.- | REFERENCIAS BIBLIOGRÁFICAS..... | 48 |

1.- INTRODUCCIÓN

La componente temática de la información geográfica se almacena con una estructura de datos diferente a la de los geométricos. No obstante ambos tipos de información están permanentemente conectados. Físicamente, los atributos se almacenan en la memoria del ordenador de forma similar a las coordenadas, como un conjunto de caracteres o números, aunque son organizados de una forma lógica.

Una base de datos es un modelo simplificado de la realidad, consecuencia de un ejercicio de abstracción que el diseñador hace a partir del propósito para el que se vaya a utilizar. Este propósito indica cuales son los aspectos relevantes que debe contener el modelo: objetos, fenómenos, atributos, relaciones entre objetos, etc. Así, el conocimiento del problema será determinante para extraer y reflejar la información relevante. Esta información, para ser almacenada y posteriormente utilizada, tendrá que organizarse debidamente. Para ello existen varios métodos: estructura jerárquica, en red, relacional y orientada al objeto. El Software para el manejo y la manipulación de estos datos será diferente según el tipo de organización. Estos programas son los denominados Gestores de Bases de Datos (DBMS).

El proceso de abstracción o traducción de la realidad a una base de datos se denomina modelización de datos, y consiste de tres niveles o fases fundamentales: modelización externa, conceptual e interna o física. El presente capítulo explica estos tres niveles del diseño de bases de datos, haciendo especial énfasis en los modelos conceptuales. El modelo relacional, que es el más usado en los SIG, se explica más detalladamente más adelante.

2.- MODELIZACIÓN DE DATOS

Una base de datos es una colección de datos interrelacionados y bien estructurados, que contienen la menor redundancia posible y que sirven el propósito de uno o más usuarios. El proceso de diseño de una base de datos es un ejercicio mental de abstracción y simplificación de la realidad a estructuras manejables por el ordenador.

El proceso de modelización incluye tres niveles (Date, 1990), que se corresponden con niveles de abstracción que son explicados mediante estructuras cercanas a como un usuario las entiende e interpreta, o más cercanas a como el ordenador las almacena y procesa. Los tres niveles son: modelo externo, modelo conceptual y modelo interno o físico. La Figura 1 esquematiza el paso de la realidad a una base de datos a través de las tres fases de la modelización. (Este ejemplo muestra un caso sencillo de modelo conceptual relacional).

2.1.- Modelización externa

El primer paso en el diseño de una base de datos es la concepción por la mente humana de las situaciones y fenómenos en el mundo real que se quieren representar. Es decir, se identifican un conjunto de conceptos generales, nombres, descripciones y relaciones (Herring, 1991), de forma que se crea una abstracción de la realidad, con las propiedades relevantes a la finalidad del estudio o proyecto. De esta forma los datos son reducidos a una cantidad finita y manejable.

Por medio de diagramas denominados de Entidad-Relación se representan los objetos, sus atributos y las relaciones entre ellos, de una forma muy cercana a como la mente humana percibe la realidad.

En términos de base de datos una entidad u objeto es simplemente algo (cosa, evento, fenómeno, concepto) que tiene sentido dentro del contexto de una aplicación, que es reconocido con existencia independiente, y que puede ser identificado de forma única. Ejemplos: persona, curso, delineación de suelos, edificio, etc. Dentro del concepto objeto se pueden distinguir los objetos espaciales como aquellos que pueden ser localizados en el espacio.

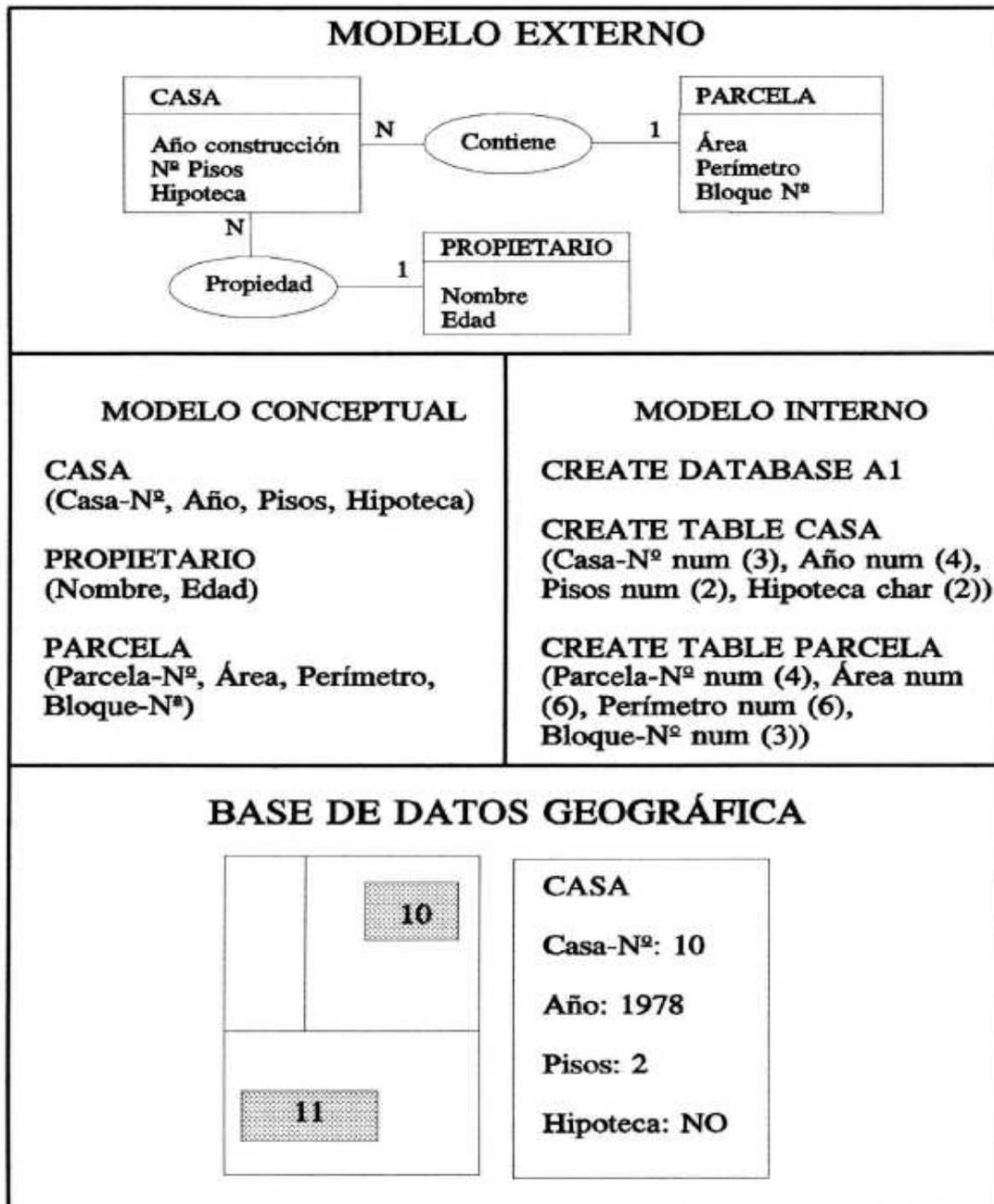
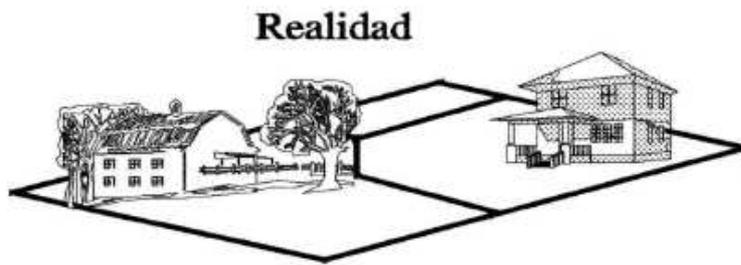


Figura 1.- Niveles de modelización en el diseño de bases de datos.

En ocasiones el término objeto es definido de forma ambigua. Bien puede referirse a una cosa o a un conjunto de cosas con similares características. En este caso se precisará objeto como clase de objeto, tipo de objeto o simplemente clase. Una clase de objeto agrupa a objetos individuales que tienen similares propiedades (atributos), relaciones comunes con objetos de otro tipo y similar comportamiento.

Un atributo es una propiedad de un objeto. Ejemplos: Persona (nombre, edad, altura, domicilio, etc.); Suelo (clase, profundidad, textura, contenido en materia orgánica, etc.).

Una relación es una asociación entre uno o más objetos o entidades. Ejemplos: Una parcela puede ser poseída por una o más personas, un suelo está compuesto por uno o más horizontes, etc. Las relaciones pueden ser de diferente tipo según el número de objetos de cada conjunto implicados en la asociación: uno-a-uno (1:1), uno-a-muchos (1:M) y muchos-a-muchos (M:N).

Una misma base de datos puede servir a diferentes usuarios o aplicaciones, que pueden requerir de diferentes datos. El nivel de modelización externo concierne a como es vista la realidad por parte de cada uno de los usuarios o aplicaciones individuales. Este nivel puede estar compuesto por varios modelos, tantos como usuarios o aplicaciones tenga la base de datos. Posteriormente las visiones externas se traducirán en una única base de datos.

En la Figura 1 se mostraba un ejemplo de diagrama de entidad-relación, donde una clase de objetos está representada por un recuadro con el nombre de la clase en la parte superior y los atributos en la parte inferior. Las relaciones están representadas por elipses, aunque también es usual hacerlo en forma de rombos. La Figura 2 muestra otro ejemplo de visión externa de una base de datos. En este caso el modelo recoge la información relevante para una aplicación que consiste en obtener mapas de erosión a partir de información hidrográfica del territorio: densidad de drenaje, área directamente erosionada por los elementos de la red de drenaje en relación a sus cuencas, grado de actividad de los elementos de la red, profundidad, etc. (Martínez Casanovas, 1994). En el modelo externo se recogen las clases de objetos (espaciales y no espaciales), sus atributos y las relaciones entre objetos, de las que se especifica su nombre y su grado. En caso de que sólo se considerase esta aplicación, la Figura 2 representaría la única visión externa de la base de datos. Si hubiese más aplicaciones, que

requiriesen la consideración de otros objetos o atributos, este modelo sólo representaría la visión parcial para la obtención del mapa de erosión.

En esta forma la información es fácilmente comprensible por el usuario aunque todavía no tiene la estructura que puede ser manejada por el ordenador. Previamente el modelo que integre todas las visiones externas debe traducirse a una forma lógica de organización de los datos en uno de los modelos de base de datos existentes.

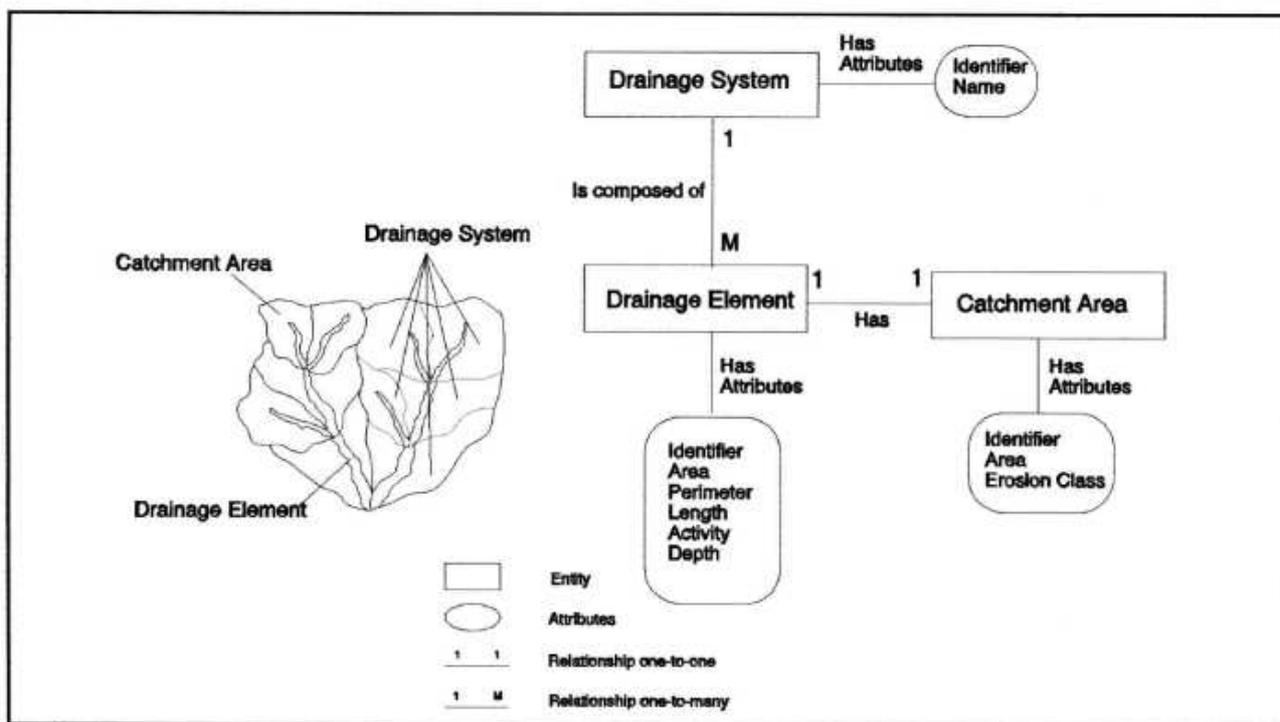


Figura 2.- Visión externa de la base de datos para la obtención de mapas de erosión en función recharacterísticas hidrográficas del terreno (Martinez Casasnovas, 1994)

2.2.- Modelización conceptual

La modelización conceptual (también denominada lógica) es el proceso de organizar los datos del modelo externo de modo que éstos puedan ser manejados por un programa gestor de base de datos. Éste es un modelo que supone una visión de los datos intermedia a como el usuario los concibe (modelo externo) y a como el ordenador los implementa y procesa (modelo interno). Este modelo todavía es independiente del Software usado, aunque el modo de organización indica el tipo de programa gestor de base de datos a utilizar. Lo más

importante del modelo conceptual es que expresa como se establecen las relaciones entre objetos. También se muestran los atributos, indicando los identificadores de los objetos.

Existen diferentes modos lógicos de estructurar los datos de forma que los objetos, sus atributos y relaciones queden registrados. Estos modelos son: modelo jerárquico, en red, relacional y orientado al objeto.

2.2.1.- Modelo jerárquico

En este tipo de estructura la información es organizada en forma de árbol. Las relaciones entre las diferentes entidades vienen definidas por la organización de la jerarquía. El único tipo posible de relaciones entre entidades es del tipo uno-a-muchos o padre-hijo. Es decir, la entidad padre puede tener muchos hijos, pero una entidad hijo sólo puede tener un padre.

En el ejemplo adjunto se representa la definición de polígonos y arcos de un mapa según este modelo (Figura 3). Cada entidad sería registrada con todos sus atributos cada vez que ocurre. Así por ejemplo, en el nivel inferior los nodos deben registrarse cada vez que un arco ocurre en la base de datos.

Otra característica de este modelo es que no permite establecer relaciones entre elementos del mismo nivel. En el ejemplo anterior no se pueden relacionar directamente arcos entre arcos. Para saber los arcos que se conectan entre sí se debe buscar en la base de datos las conexiones en el nivel inferior.

Físicamente las relaciones entre diferentes entidades se hacen mediante punteadores que se especifican en un programa. Es decir, que la base de datos jerárquica es programada una vez que se conocen las relaciones existentes entre las diferentes entidades.

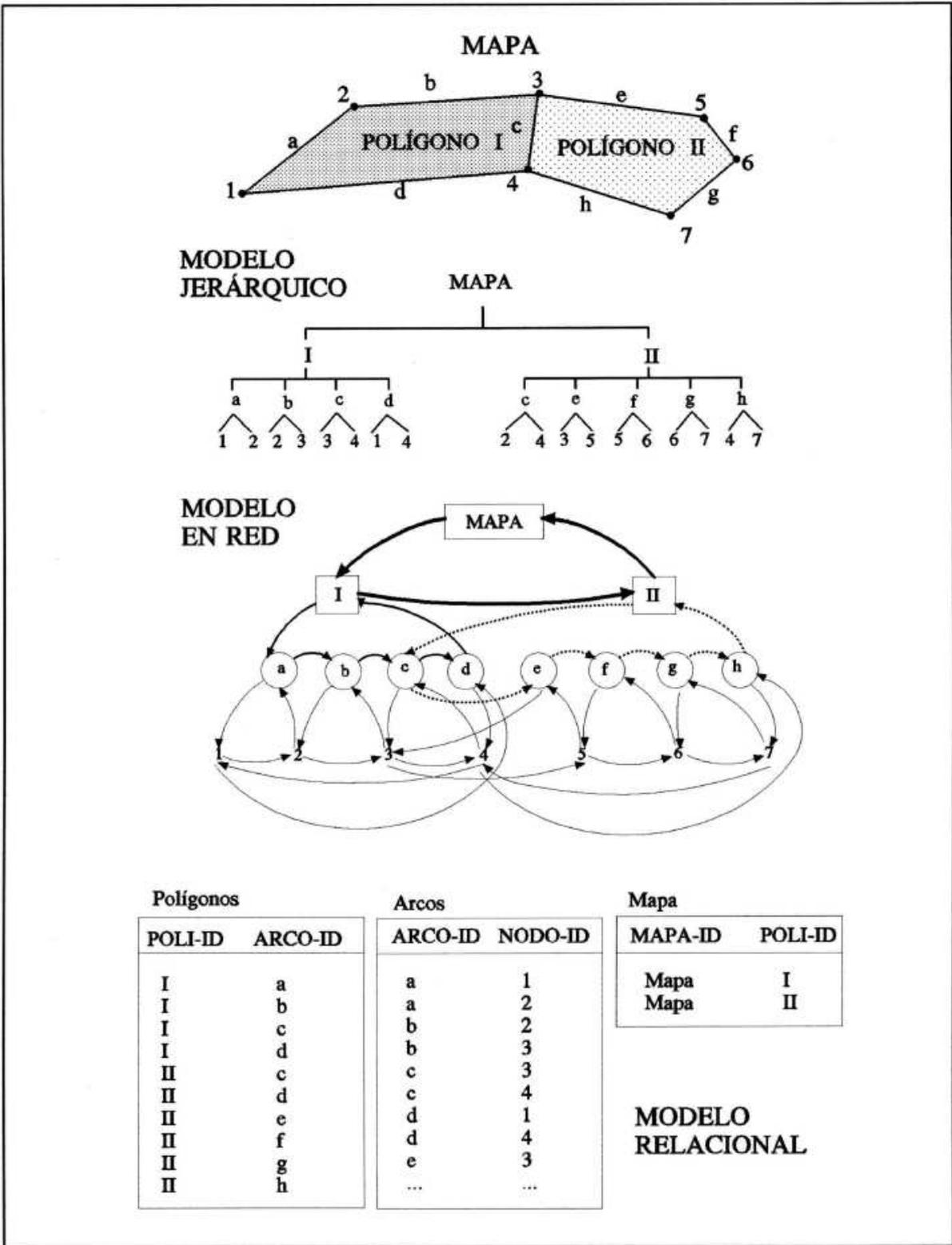


Figura 3.- Modelos de organización lógica de datos para el diseño de bases de datos: modelo jerárquico, en red y relacional.

Como conclusión, se puede decir que este modelo de datos resulta bastante inflexible, ya que todas las relaciones deben hacerse explícitas mediante programación y por tanto deben ser conocidas de antemano. Esto implica que para aplicaciones como análisis de información geográfica esta estructura no es útil, ya que las preguntas que normalmente se desean son de carácter exploratorio y por tanto no se conocen a priori para ser programadas.

2.2.2.- Modelo en red

Este modelo resuelve algunos problemas de inflexibilidad planteados por el modelo jerárquico, ya que una entidad puede tener múltiples padres así como múltiples hijos. Al igual que en la estructura jerárquica, las relaciones entre individuos de diferentes entidades son registrados con punteadores. En el caso del ejemplo anterior, se observa la complejidad de la estructura de punteadores o conexiones necesaria para definir las relaciones, que también se establecen mediante programación (Figura 3).

Ventajas de este tipo de estructura son la rapidez de respuesta a consultas programadas y la menor redundancia de datos, ya que los atributos de cada entidad son registrados una sola vez. Las desventajas son la complejidad de la estructura y la necesidad de programar con antelación las relaciones entre entidades. Además si se necesita actualizar la base de datos o modificarla eso supone reconstruir parte de la estructura de punteadores. Este tipo de modelo tampoco viene siendo muy utilizado en aplicaciones en SIG dada la complejidad e inflexibilidad de cara a consultas de carácter exploratorio.

2.2.3.- Modelo Relacional

Al igual que en los casos anteriores, los datos sobre un individuo de una clase de objeto son almacenados en un registro. Éste contiene los valores de cada atributo que describe a la clase. El modelo de datos relacional no emplea punteadores para indicar las relaciones entre individuos de diferentes entidades. Todos los registros correspondientes a los miembros de una entidad se agrupan en una tabla. Las distintas tablas, correspondientes a diferentes entidades, se asocian compartiendo un atributo común (Figura 3)

La estructura relacional se caracteriza por:

- Simplicidad, ya que todos los datos son almacenados en tablas, siendo muy fácil añadir atributos o registros.
- Permite la asociación de tablas por medio de atributos comunes, de forma que toda la información contenida en diferentes tablas está relacionada.
- El método para representar modelos externos (modelos Entidad-Relación) son muy similares a como se representan este tipo de estructuras.
- Permite la representación de relaciones muchos-a-muchos (M:N), para lo cual se necesita definir una nueva tabla que contenga la relación, de modo que se crean dos relaciones del tipo uno-a-muchos (1:M).

El modelo relacional es el más utilizado para estructurar la información temática en SIG dada su flexibilidad y sencillez de manipulación de datos.

2.2.4.- Modelo orientado al objeto

Los modelos de datos jerárquico y en red estaban basados en la programación pura de las relaciones entre diferentes clases de objetos, mientras que en el relacional la estructura de datos es independiente de cualquier tipo de programación. Las clases de objetos se representan mediante tablas que agrupan las ocurrencias en registros. Las relaciones se representan mediante atributos comunes en las tablas o bien mediante nuevas tablas.

El modelo de datos orientado al objeto, que no se debe confundir con el modo objeto de representación de la información, es un método de organización de datos que complementa la estructura relacional con la incorporación de programas embebidos en la propia estructura de datos. Estos programas sirven para implementar el concepto de comportamiento de los objetos, que son las operaciones o métodos que una clase de objetos puede realizar o ser sometido a.

En la actualidad este modelo está conceptualmente desarrollado (Rumbough, 1991), y parece ser el modelo que sucederá en el futuro próximo a las estructuras relacionales puras en muchos campos de la información. También se está desarrollando Software para crear y manipular bases de datos orientadas al objeto. Uno de los enfoques en Software está basado en gestores de bases de datos relacionales (POSTGRES, sucesor de INGRES RDBMS). Éste considera el almacenamiento de la estructura jerárquica de la clasificación de los objetos en tablas relacionales. El llamamiento a programas en lenguaje de base de datos para realizar las operaciones también se almacena en atributos especiales dentro de las tablas.

Respecto a la implementación del modelo orientado al objeto en SIG, el mejor ejemplo es el Software TIGRIS (Intergraph), que utiliza programación orientada al objeto. No obstante, el modelo más utilizado es el relacional, ya que el orientado al objeto no está plenamente desarrollado (Healey, 1991).

La sección 3 está dedicada al modelo relacional o georelacional, como se viene denominando dentro del entorno SIG.

2.3.- Modelización interna

La modelización interna o física es la última parte del proceso de creación de la base de datos, que describe como los datos se hallan físicamente almacenados en la memoria del ordenador (Figura 8.1). Este modelo es dependiente del Software empleado. Cada programa tendrá sus particularidades, aunque dos sistemas gestores del mismo tipo implementan los mismos conceptos. En este modelo es donde se crea una tabla, para cada atributo se especifica el número de bytes reservados para almacenar los valores, el tipo de dato (numérico, carácter, fecha, lógico, etc., y el número de decimales (para números reales.

3.- DISEÑO DE BASES DE DATOS RELACIONALES

En la sección 2.2 se introducían los diferentes modelos lógicos de estructuras de datos temáticos, de los cuales el modelo relacional es el más usado en los SIG. Éste está

caracterizado por la simplicidad de su estructura, y por la mayor facilidad de manejo de los datos mediante lenguajes similares al hablado (desde el punto de vista del usuario), como por ejemplo el SQL o "Structured Query Language". El modelo relacional se compone de: la estructura de datos y la parte operativa encargada de realizar las operaciones de manipulación (basada en álgebra de conjuntos). En esta sección se exponen las normas de diseño de una estructura de datos relacional, mientras que la parte operativa (álgebra relacional y SQL) es explicada en la sección 4.

3.1.- Tablas

Una base de datos relacional es una colección de tablas. Tanto las entidades (clases de objetos) como las relaciones se representan del mismo modo, como tablas. El nombre de la tabla se corresponde con el nombre de la clase de objeto o de la relación. La tabla está compuesta de filas y columnas. Las columnas representan los atributos que describen la clase, mientras que las filas son ocurrencias de la clase (objetos individuales). Todos los individuos se caracterizan por tener la misma lista de atributos. La intersección fila/columna contiene el valor del atributo para un determinado objeto de la clase.

Las tablas relacionales deben cumplir una serie de normas:

- Cada ocurrencia de una tabla debe ser diferente en al menos un atributo.
- Cada intersección atributo/ocurrencia debe contener un solo valor. No se permiten valores múltiples ni espacios vacíos.
- El orden de las ocurrencias o atributos en la tabla no es importante.
- Cada individuo (ocurrencia) debe estar identificado de forma única. La identificación se hace mediante un atributo que se denomina identificador o atributo clave.

Una tabla que cumple con estas normas se dice que está normalizada. La Figura 4 muestra los componentes y características de una tabla según el modelo relacional.

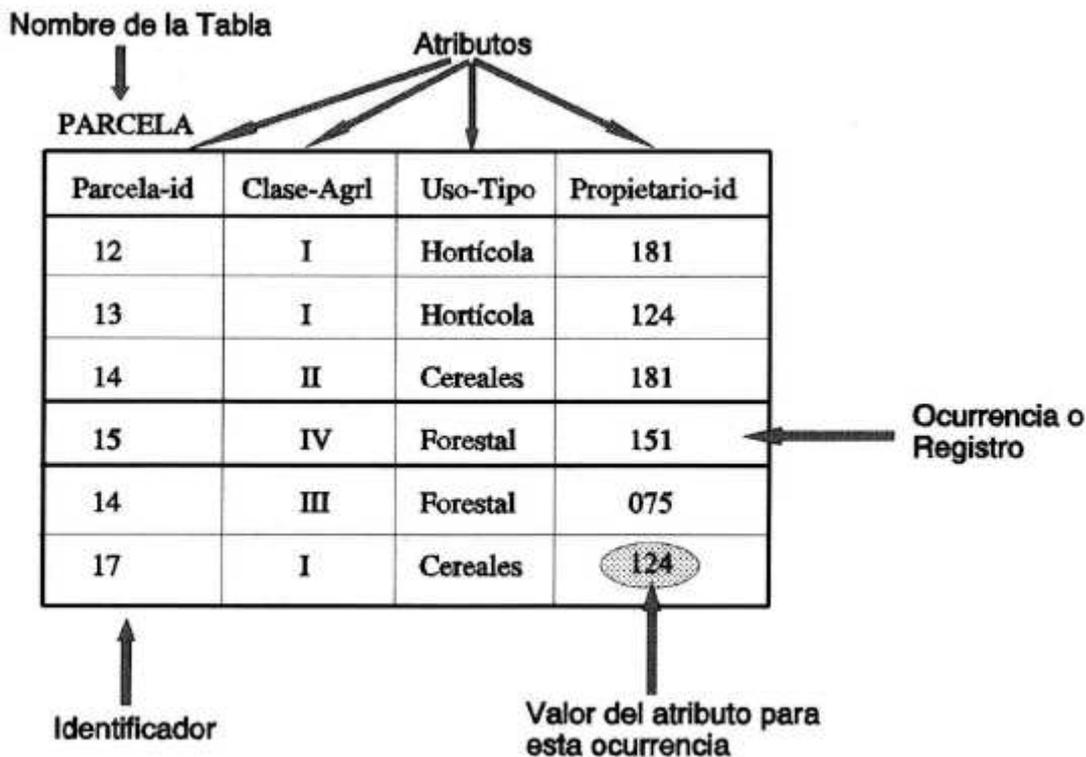


Figura 4.- Componentes de una tabla según el modelo relacional.

3.1.1.- Normalización de tablas

La normalización es el proceso de aplicación de las normas que debe cumplir un conjunto de datos para convertirse en una tabla relacional. Este proceso se describe en forma de estados. Así se habla de 1ª forma normal, 2ª forma normal y 3ª forma normal en referencia a las diferentes fases de la normalización. En cada fase son eliminadas determinadas características que no son deseables en una tabla relacional, como por ejemplo: valores o grupos repetidos, datos redundantes o dependencia única del atributo clave, y dependencia transitiva entre atributos.

1ª Forma Normal: Grupos repetidos.

Una norma de las tablas relacionales es que cada intersección atributo/ocurrencia debe contener un único valor, no pudiendo ser ni múltiple ni vacío. Supongamos una tabla (Tabla 1) que contiene datos sobre parcelas, relevantes para una aplicación catastral o para un mapa de usos del suelo. En esta tabla se observa que una parcela puede pertenecer a más de una

persona (propietario). Así lo indican, por ejemplo, los registros correspondientes a las parcelas 12 y 14. Este hecho incumple una de las normas para que esta tabla forme parte de una base de datos relacional: grupos repetidos o casillas con más de un valor.

Tabla 1.- Parcela

| Parcela-id | Clase-Agrl | Uso-tipo | Propietario-id |
|------------|------------|-----------|----------------|
| 12 | I | HORTÍCOLA | 181 124 |
| 13 | I | HORTÍCOLA | 124 |
| 14 | II | CEREALES | 181 075 |
| 15 | IV | FORESTAL | 151 |
| 16 | III | FORESTAL | 075 |
| 17 | I | CEREALES | 124 |

La eliminación de los grupos repetidos tiene dos posibles soluciones:

a) El modo más sencillo es la creación de nuevos registros en la misma tabla con los mismos valores de los atributos, excepto para el caso de los valores duplicados (Tabla 2). Sin embargo, aunque la tabla resultante sería correcta, de esta forma se crea duplicación innecesaria de datos.

Tabla 2.- Parcela

| Parcela-id | Clase-Agrl | Uso-tipo | Propietario-id |
|------------|------------|-----------|----------------|
| 12 | I | HORTÍCOLA | 181 |
| 12 | I | HORTÍCOLA | 124 |
| 13 | I | HORTÍCOLA | 124 |
| 14 | II | CEREALES | 181 |
| 14 | II | CEREALES | 075 |
| 15 | IV | FORESTAL | 151 |
| 16 | III | FORESTAL | 075 |
| 17 | I | CEREALES | 124 |

b) Otro modo de eliminar los grupos repetidos, evitando la duplicación de datos, es creando una nueva tabla (Tabla 4) para contener la relación entre "Parcela" y "Propietario". Los atributos de esta tabla serán el identificador de la tabla original y los valores individuales del atributo que contenía los valores repetidos. El resto de los atributos se mantiene en la primera tabla (Tabla 3), excepto el atributo con valores repetidos que desaparece. La nueva tabla está relacionada con la original a través del identificador "Parcela-id", que es común en ambas tablas.

Tabla 3.- Parcela

| Parcela-id | Clase-Agrl | Uso-Tipo |
|------------|------------|-----------|
| 12 | I | HORTÍCOLA |
| 13 | I | HORTÍCOLA |
| 14 | II | CEREALES |
| 15 | IV | FORESTAL |
| 16 | III | FORESTAL |
| 17 | I | CEREALES |

Tabla 4.- Parcela/Propietario

| Parcela-id | Propietario-id |
|------------|----------------|
| 12 | 181 |
| 12 | 124 |
| 13 | 124 |
| 14 | 181 |
| 14 | 075 |
| 15 | 151 |
| 16 | 075 |
| 17 | 124 |

2ª Forma Normal: Datos Redundantes.

Redundancia significa duplicación innecesaria de datos en una tabla. Un valor es redundante si puede ser eliminado sin pérdida de información. En el ejemplo siguiente (Tabla 5) se observa que si en el primer registro de la tabla se elimina el valor del atributo "Nombre" todavía se puede conocer el nombre del propietario de la parcela correspondiente, ya que éste

se encuentra duplicado cada vez que aparece en la tabla el mismo identificador del propietario. Esta tabla se dice que tiene valores duplicados que además son redundantes. Esto ocurre siempre que hay atributos que no son dependientes del identificador de la tabla, sino de otro atributo, como en este caso que "Nombre", "Dirección" o "Edad" son dependientes de "Propietario-id".

Tabla 5.- Parcela

| Parcela-id | Clase-Agrl | Uso-Tipo | Propietario-id | Nombre | Dirección | Edad |
|------------|------------|-----------|----------------|------------|------------|------|
| 12 | I | HORTÍCOLA | 181 | J. Sánchez | Ronda 14 | 48 |
| 13 | I | HORTÍCOLA | 124 | R. Álvarez | Gracia 67 | 56 |
| 14 | II | CEREALES | 181 | J. Sánchez | Ronda 14 | 48 |
| 15 | IV | FORESTAL | 151 | J. Lobato | Tamarit 6 | 23 |
| 16 | III | FORESTAL | 075 | F. Marcial | Lope V. 25 | 38 |
| 17 | I | CEREALES | 124 | R. Álvarez | Gracia 67 | 56 |

La eliminación de la redundancia se hace dividiendo la tabla en dos tablas relacionadas que comparten un atributo común (Tabla 6 y

Tabla 7), aquel del cual dependen los atributos que son redundantes, ("Propietario-id" en el caso del ejemplo). Evitando los datos redundantes se consigue que todos los atributos de la tabla tengan dependencia única de su atributo clave o identificador.

Tabla 6.- Parcela

| Parcela-id | Clase-Agrl | Uso-Tipo | Propietario-id |
|------------|------------|-----------|----------------|
| 12 | I | HORTÍCOLA | 181 |
| 13 | I | HORTÍCOLA | 124 |
| 14 | II | CEREALES | 181 |
| 15 | IV | FORESTAL | 151 |
| 16 | III | FORESTAL | 075 |
| 17 | I | CEREALES | 124 |

Tabla 7.- Propietario

| Propietario-id | Nombre | Dirección | Edad |
|----------------|------------|------------|------|
| 181 | J. Sánchez | Ronda 14 | 48 |
| 124 | R. Álvarez | Gracia 67 | 56 |
| 151 | J. Lobato | Tamarit 6 | 23 |
| 075 | F. Marcial | Lope V. 25 | 38 |

Es importante saber distinguir entre datos redundantes y datos duplicados. La duplicación de datos es la repetición de valores de un atributo en diferentes registros. Sin embargo, se distingue de los datos redundantes en que si aquellos son eliminados se produce pérdida de información. La siguiente tabla (Tabla 8) es un ejemplo de datos duplicados. El primer y segundo registro contienen los mismos valores para los atributos "Clase-Agrl" y "Uso-Tipo", sin embargo si el valor de la clase de uso para la parcela 12 es eliminado es imposible conocer, a partir de la tabla, cual es el uso en esta parcela. En este caso la duplicación de datos no es redundante, y la tabla está bien construida.

Tabla 8.- Parcela

| Parcela-id | Clase-Agrl | Uso-Tipo |
|------------|------------|-----------|
| 12 | I | HORTÍCOLA |
| 13 | I | HORTÍCOLA |
| 14 | II | CEREALES |
| 15 | IV | FORESTAL |
| 16 | III | FORESTAL |
| 17 | I | CEREALES |

3ª Forma Normal: No Dependencia Transitiva.

La no dependencia transitiva significa la eliminación de atributos que sean función o puedan ser derivados a partir de atributos ya existentes en la tabla. Un caso muy frecuente es el de la edad y la fecha de nacimiento o de comienzo de una actividad. Conocida la fecha de nacimiento o comienzo puede ser calculada la edad, por lo que este atributo no es necesario

incluirlo en la tabla. Otro ejemplo podría ser una tabla que contenga medidas sobre el área de las cuencas, la longitud de los elementos de la red de drenaje en dichas cuencas, la densidad de drenaje, etc. Ya que la densidad de drenaje se calcula como la longitud de los elementos de la red dividido por el área de la cuenca no será necesario incluir este atributo en la tabla. Siempre que sea preciso se podrá calcular a partir de la longitud y el área.

3.1.2.- Identificadores

Una norma que deben cumplir las tablas es que cada individuo de una clase de objeto debe estar únicamente identificado. Un identificador es un atributo o atributos (caso de los identificadores compuestos) que es suficiente para distinguir un individuo de forma única. En la Tabla 6 el atributo "Parcela-id" es el atributo identificador. Una parcela no queda identificada por ser de clase agrológica I o tener un uso hortícola. Con estas características podemos encontrar infinidad de parcelas. Sin embargo, dado que el número de "Parcela-id" se ha asignado de forma única a cada individuo, constituye un código que sirve para identificar cada parcela dentro de esta clase de objetos.

En Tabla 7 el identificador puede ser tanto el nombre del propietario como su número de identificación. En este caso el atributo "Nombre" constituye un identificador candidato. Un identificador compuesto se daría en el caso de que el nombre estuviese dividido en diferentes columnas para los apellidos y el nombre propio. En este caso el identificador estaría formado por tres atributos (dos apellidos y nombre), formando un identificador compuesto.

Los identificadores son también usados para establecer las relaciones entre tablas. Un identificador de una tabla (A) puede ser anexado como atributo en otra tabla (B) para crear una relación o conexión entre ellas. A este atributo en la segunda tabla se le denomina identificador externo. En la Tabla 6 el atributo "Propietario-id" es un identificador externo.

3.2.- **Relaciones**

3.2.1.- Características generales

Una relación es una asociación entre elementos de dos conjuntos, o en el caso de tablas relacionales entre ocurrencias de dos tablas. Las relaciones entre entidades se representan en forma de Diagramas de Entidad-Relación. Estos diagramas muestran las entidades, el tipo de relación y el grado, que indica con cuantos elementos de una entidad se relaciona (B) o se puede relacionar un elemento de la otra (A).

Las relaciones vienen reguladas por una serie de reglas o condiciones también denominadas por algunos autores "Reglas de la Empresa". Por ejemplo entre parcelas y personas pueden definirse muchos tipos de relaciones de propiedad: una parcela puede ser propiedad de una sola persona, una parcela puede ser propiedad de más de una persona, una persona puede poseer sólo una parcela, o más de una parcela, etc. La definición del tipo de relación dependerá del usuario y de la aplicación, es decir, de la visión particular que al usuario le interesa de la realidad para su aplicación. La Figura 5 muestra un ejemplo de relación entre las entidades "Parcela" y "Persona" asociadas mediante una relación de propiedad en la que se ha limitado a que una parcela sólo esté poseída por una persona y viceversa.

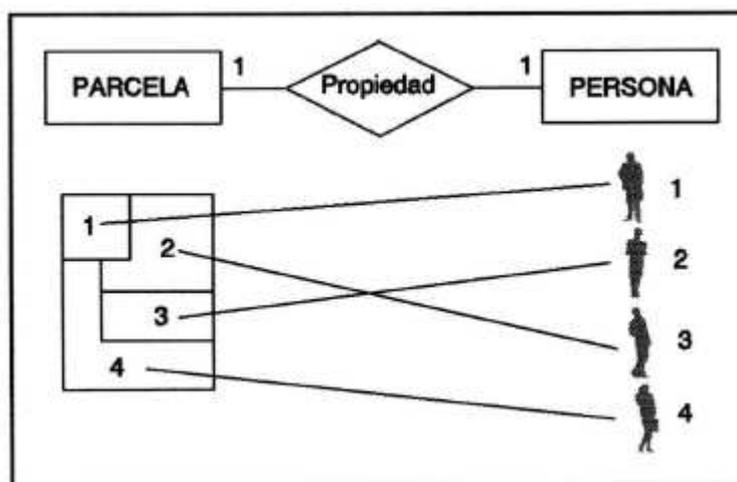


Figura 5.- Ejemplo de relación.

Asociación entre miembros de dos entidades.

3.2.2.- Grado y obligatoriedad de una relación

Una propiedad importante de la relación es el grado. Existen tres tipos de relaciones según el grado, que estará determinado por las reglas de la relación: uno-a-uno (1:1), uno-a-muchos (1:M) y muchos-a-muchos (M:N). Además, hay dos formas de como un tipo de objetos puede participar en una relación: a) de forma obligatoria, es decir, todas las ocurrencias de la tabla A se relacionan con algún miembro de la tabla B; o b) de forma no obligatoria, es decir, no

todos los miembros de la tabla A se tienen que relacionar obligatoriamente con miembros de la tabla B.

Dependiendo del grado y obligatoriedad de una relación entre dos tablas, ésta se llevará a cabo de diferente forma, o bien anexando el identificador de una tabla a la otra o bien creando una nueva tabla relacional.

3.2.2.1 Relaciones 1:1

Hay tres tipos de relaciones (1:1): obligatorias, obligatoria-no obligatoria y no obligatorias. (Figura 6).

En la relación obligatoria (por ambos lados) cada individuo de la entidad A se asocia a uno de B y viceversa. La relación entre dos tablas se llevará a cabo anexando indistintamente el identificador de una de las tablas en la otra. En el ejemplo de la Figura 6 el atributo "Propietario-id" se anexa a la tabla "Parcela", aunque también podría ser al revés. También es posible formar una sola tabla.

Otro caso de relaciones 1:1 es cuando la obligatoriedad sólo se mantiene por uno de los lados de la relación. Es decir, todo miembro de la tabla A se asocia con uno de la tabla B, pero no todos los miembros de la tabla B están asociados con uno de la tabla A. La relación entre tablas debe establecerse anexando obligatoriamente el identificador de la tabla B en la tabla A, ya que al contrario se crearían espacios vacíos, violando las normas de las tablas del modelo relacional (Figura 6).

Por último está el caso de la no obligatoriedad por ambas partes de la relación. Un miembro de la tabla A puede asociarse con uno de la tabla B, y viceversa un miembro de B puede asociarse a uno de A aunque de forma no obligatoria en ambos casos. En este caso la relación entre tablas no puede hacerse mediante anexión de uno de los identificadores a la otra tabla ya que en cualquier caso se crearían espacios vacíos. La solución es la creación de una tabla que contenga la relación. El identificador de esta tabla será el compuesto por los identificadores anexados de las tablas originales. Además, la nueva tabla puede tener

atributos propios de la relación (Figura 6). Este mismo hecho ocurre en otros dos casos: relaciones (1:M) obligatoria-no obligatoria y relaciones (M:N).

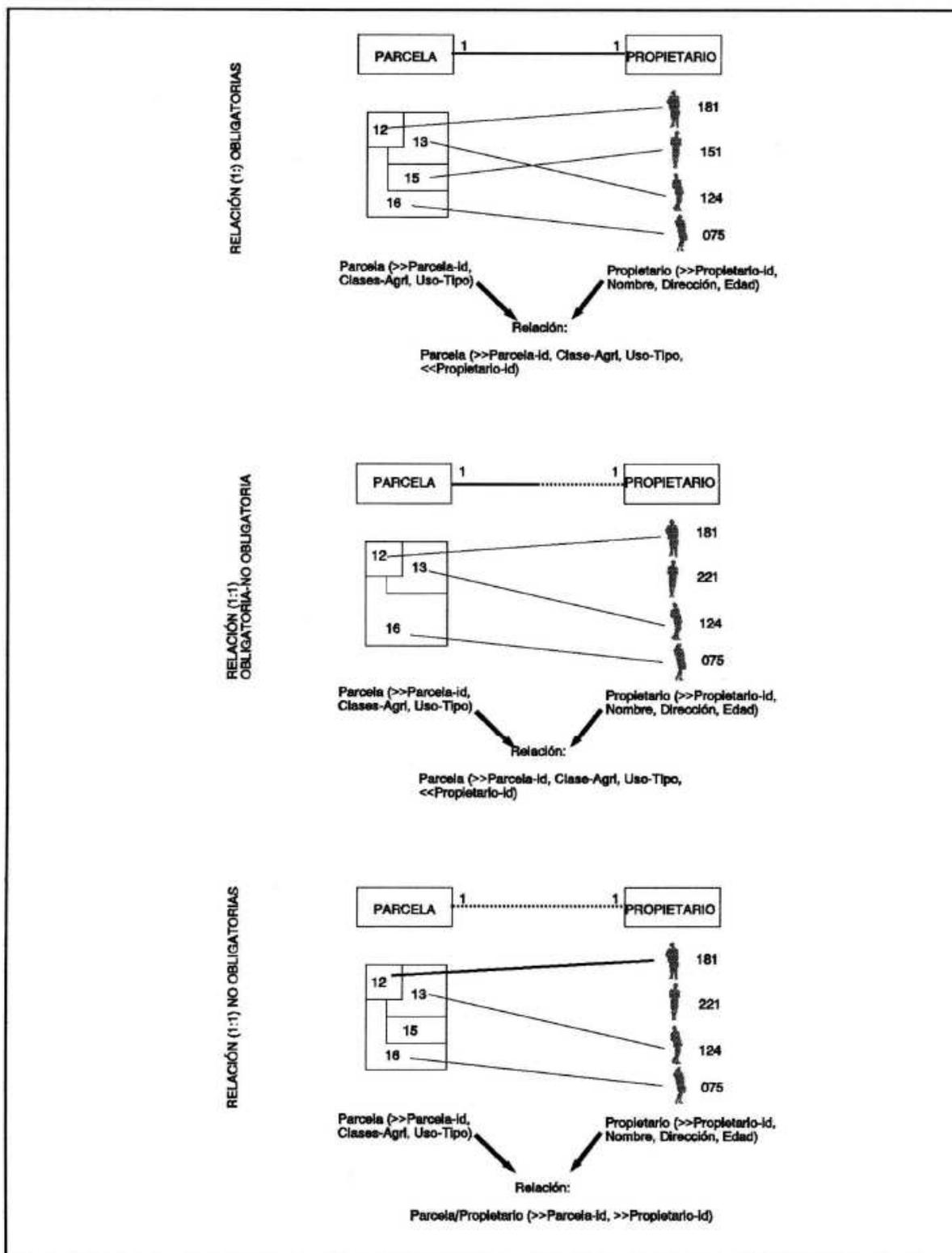


Figura 6.- Relaciones (1:1), todos los casos. (La línea de puntos indica la relación no obligatoria, >> es el símbolo para el identificador clave y << para el identificador externo).

3.2.2.1 Relaciones 1:M

Dentro de estas relaciones hay que distinguir dos casos generales: a) relaciones (1:M) obligatorias y (1:M) no obligatoria-obligatoria; y b) relaciones (1:M) obligatoria-no obligatoria y (1:M) no obligatorias.

En el primer caso (a) la relación se establece anexando el identificador de la entidad donde el grado de la relación es 1 en la entidad donde el grado es M. De otro modo se crearían valores repetidos. El segundo caso (b) se resuelve creando una nueva tabla relacional como se ha comentado para el caso (1:1) no obligatorias. De otra forma, al anexar uno de los identificadores a la otra tabla se puede crear o bien valores repetidos o bien valores vacíos que incumplen las normas del modelo relacional.

3.2.2.1 Relaciones M:N

Este tipo de relación exige la creación de una nueva tabla relacional entre las dos entidades que se asocian (Figura 7). Un miembro de la tabla A puede asociarse a más de un elemento de la tabla B (una parcela puede ser propiedad de diversas personas), y al contrario un miembro de la tabla B puede asociarse a más de uno de la tabla A (una persona puede poseer más de una parcela). Toda relación (M:N), al normalizarse, queda descompuesta en dos relaciones entre tres tablas que son del tipo (1:M).

A modo de resumen, a la hora de establecer asociaciones entre tablas según el modelo relacional se debe tener en cuenta:

- No anexar un identificador que pueda crear valores vacíos ni grupos repetidos.
- No representar una relación por una tabla adicional excepto en los casos necesarios.

Con esta serie de normas se puede diseñar la estructura de cualquier base de datos relacional. A partir de aquí la base de datos está dispuesta para ser consultada. Para ello se utiliza un conjunto de operaciones denominadas de álgebra relacional. Estas operaciones se

implementan informáticamente por medio de un lenguaje de manejo y manipulación de datos (Ej. SQL).

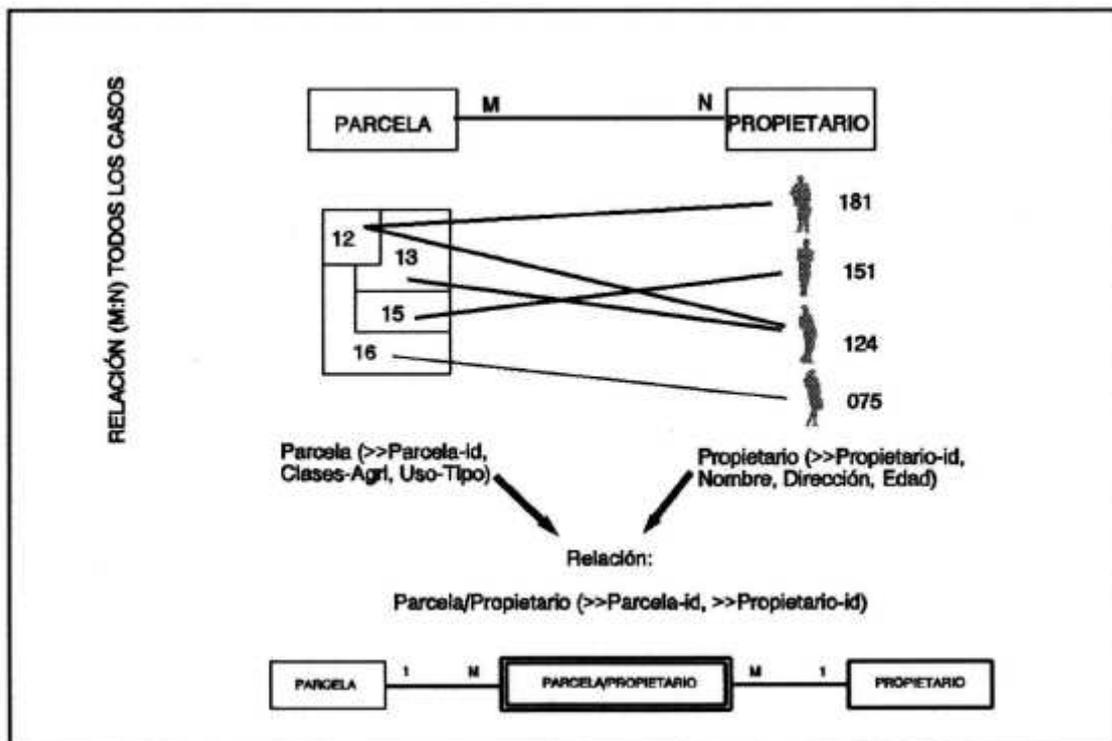


Figura 7.- Relaciones (M:N), todos los casos.

3.3.- El modelo entidad-relación de la base de datos

En la sección 2.1 se introdujo el término Diagramas de Entidad-Relación. Estos diagramas son una técnica, introducida por Chen (1976), que se usa para representar conceptos tales como objetos, sus atributos y relaciones de una forma cercana a como la mente humana los percibe del mundo real.

Esta técnica también es usada para representar el modelo conceptual de una base de datos relacional. En éste se representan todas las entidades (tablas), con sus atributos (indicando los identificadores claves y los externos), y las relaciones. Así, el usuario tiene una visión de conjunto de la base de datos a implementar en el ordenador, que resalta los aspectos temáticos. La Figura 8 muestra un ejemplo del modelo conceptual de una base de datos relacional. En concreto se trata del modelo correspondiente a la visión externa representada

en la Figura 2. Para completar la información, en el modelo conceptual se puede especificar cuales son los objetos con componente espacial. No todas las entidades de una base de datos geográfica tendrán componente espacial directamente asociado. Esto ocurre ya que el proceso de normalización habrá producido el reparto de los datos en diferentes tablas relacionadas. Así, aunque haya datos sin componente espacial, al estar relacionadas con tablas que sí están asociadas a este componente, siempre se podrá seleccionar objetos espaciales por características o información contenida en el primer tipo de tablas. Este es el ejemplo de la tabla "Drainsys" en la Figura 8. En este caso, un sistema de drenaje es un concepto que se define como la unión de varios elementos de la red conectados y sus cuencas de drenaje. La geometría de una sistema de drenaje (Ej. red de drenaje) podrá derivarse a partir de la geometría de las partes elementales, que está almacenada a nivel de los elementos individuales. La pertenencia de un elemento a un determinado sistema se especifica por medio de un atributo ("Drainsys-id"), que actúa como identificador externo en la entidad con componente espacial "Drainarea".

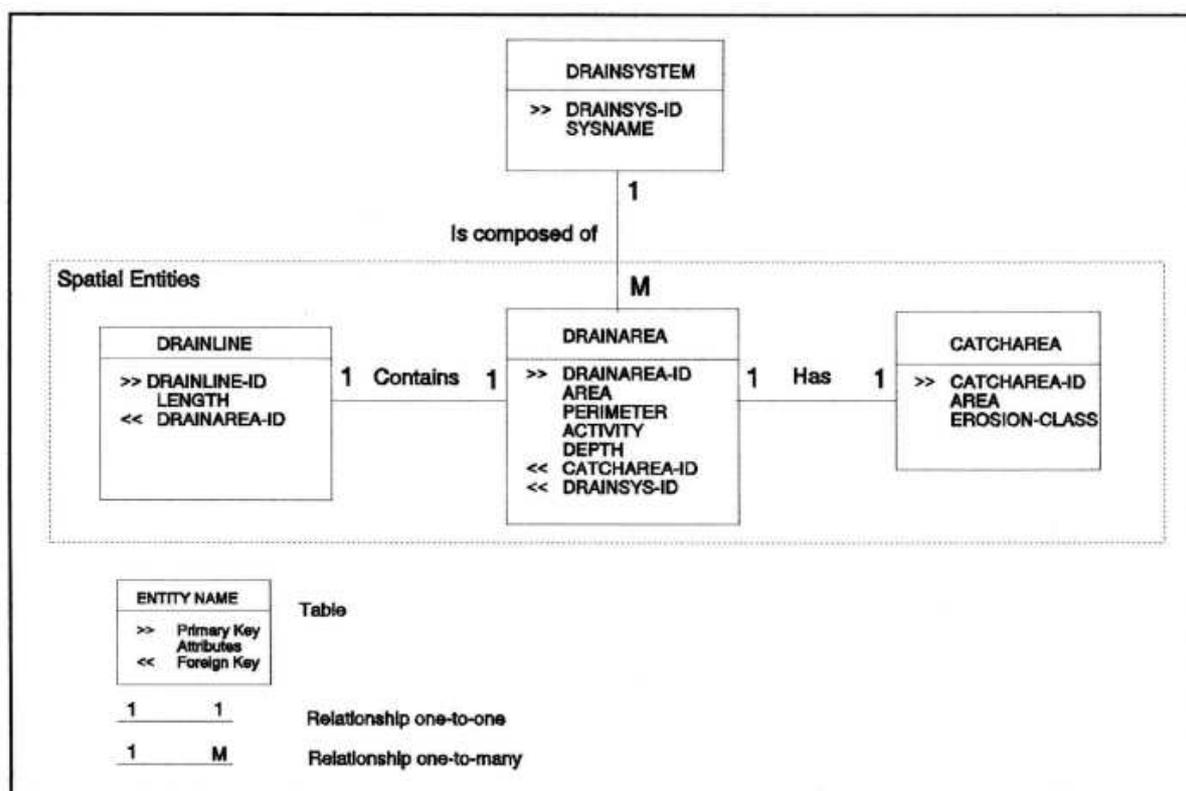


Figura 8.- Ejemplo de modelo conceptual de base de datos relacional (Martínez Casanovas, 1994).

4.- ÁLGEBRA RELACIONAL

4.1.- Concepto de álgebra relacional

La última parte del modelo relacional está compuesta por la parte operativa, es decir, las operaciones de manipulación de los datos contenidos en las tablas. El conjunto de operaciones de manipulación de datos en tablas relacionales se denomina álgebra relacional. Está basado en una serie de operaciones que consideran una o dos tablas como datos de partida y producen una nueva tabla como resultado.

Los tres tipos de operaciones principales de manipulación de datos contenidos en tablas son:

a.- Selección de ocurrencias de acuerdo a una condición (restricción). Las condiciones se describen mediante operadores de álgebra de conjuntos (Figura 9).

Condiciones: =, <, >, <=, =>, A and B, A not B, A or B, A xor B.

b.- Selección de atributos especificados (proyección).

c.- Conexión de tablas ("join") por medio de un atributo común.

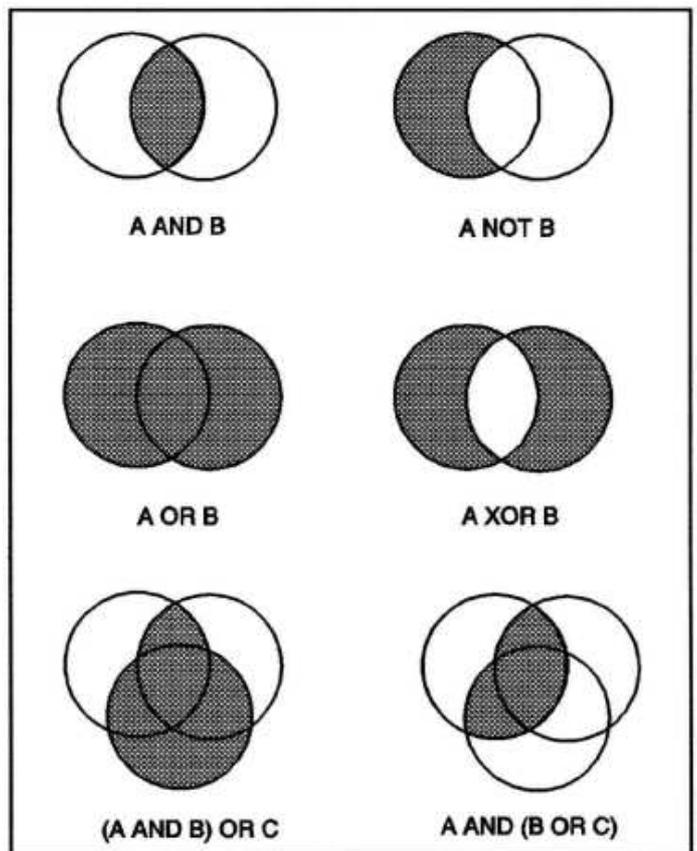


Figura 9.- Operadores de álgebra booleana.

4.2.- SQL

La realización de operaciones en las tablas mediante el álgebra relacional necesita de un lenguaje específico para servir de interfase entre el usuario y el sistema gestor de la base de datos. El lenguaje más extendido para la manipulación de datos en bases relacionales es el "Structured Query Language" o SQL. Éste es un lenguaje de 4ª generación desarrollado por

IBM a finales de los 70 y que ha llegado a ser un estándar en bases de datos relacionales. SQL es parecido en la estructura de sus frases al inglés. Es usado para controlar la base de datos (creación, manipulación de datos contenidos en tablas, etc.).

Hay muchas ventajas del uso de SQL sobre la programación con el fin de manipular datos. Usando SQL el operador simplemente especifica los resultados que quiere obtener y no tiene que pensar en definir un procedimiento para obtenerlos. El usuario especifica las operaciones en términos de "qué" es lo que se quiere, no en términos de "cómo" hay que obtenerlo. Es decir, SQL, junto con las bases de datos relacionales, son no procedurales, al contrario que los lenguajes de programación.

Las operaciones se realizan por medio de sentencias estándar que son ejecutables. Éstas emplean comandos como CREATE TABLE (para crear tablas), INSERT (para añadir registros a una tabla), UPDATE (para actualizar valores), DELETE (para borrar valores), SELECT (para selección de registros o atributos en una o varias tablas), etc.

Ejemplo de definición de una tabla y entrada de datos:

```
CREATE TABLE Parcela
(Parcela-id num (3),
Clase-Agrl char (3),
Uso-Tipo char (15),
Propietario-id num (8));
```

```
INSERT into PARCELA values (12, I, HORTÍCOLA, 181);
```

La sentencias para consultas o selección de datos a partir de tablas es del tipo:

```
SELECT (atributos)
FROM (tablas)
WHERE (restricciones);
```

Ejemplos:

a.- Selección de todos los registros de la tabla Parcela.

```
SELECT * FROM Parcela;
```

b.- Selección de registros con restricciones. Ejemplo de selección de parcelas con uso cereal y clase agrológica I o II.

```
SELECT * FROM parcela
WHERE Uso-Tipo = "CEREAL" and
(Classe-Agrl = "I" or Clase-Agrl = "II");
```

c.- Selección de atributos. Ejemplo de selección de los diferentes tipos de usos de las parcelas. Hay que tener en cuenta que SQL no elimina los valores repetidos a menos que se indique lo contrario mediante la cláusula SELECT DISTINCT.

```
SELECT DISTINCT Uso-Tipo
FROM Parcela;
```

d.- Selección a partir de más de una tabla. Este tipo de operación implica la conexión de tablas relacionadas. Ejemplo de selección del número de parcela y del tipo de uso de las parcelas pertenecientes al propietario R. Álvarez (se supone una relación M:1 entre parcelas y propietarios). En este caso el nombre del propietario no figura en la tabla parcela. Éste está en la tabla "Propietario", relacionada con "Parcela" a través del atributo "Propietario-id". Para resolver esta consulta habrá que conectar estas dos tablas (Tablas 8.6 y 8.7).

```
SELECT Parcela-id, Uso-Tipo
FROM Parcela, Propietario
WHERE Parcela.Propietario-id = Propietario.Propietario-id and
Nombre = "R. Álvarez";
```

Estos son algunos ejemplos de carácter introductorio. Más información sobre las operaciones de manipulación de bases de datos relacionales mediante SQL, como diseño de bases de datos, relleno de tablas, selección de datos, agrupamiento y cálculo de estadísticas, conexión entre tablas, etc., puede obtenerse consultando las siguientes publicaciones:

- Emerson, S.L. et al., 1989. *The Practical SQL handbook*. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts.
- Wellesley Software, 1992. *Learning advanced SQL*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

5.- EJEMPLO DE APLICACIÓN DEL DISEÑO DE BASES DE DATOS RELACIONALES: BASE DE DATOS DE SUELOS PARA ESTUDIOS DE EROSIÓN Y EVALUACIÓN DE TIERRAS

5.1.- Introducción

Como caso de aplicación del diseño de bases de datos relacionales, en este capítulo se presenta un ejemplo de diseño de una base de datos de suelos con información relevante para estudios de erosión y evaluación de tierras a escalas semidetalladas (Martínez Casasnovas, 1993).

El suelo y el relieve son dos factores principales que influyen la erosión hídrica de los suelos. Información precisa sobre los suelos, sus características y distribución espacial en el paisaje, es conveniente para determinar las posibles relaciones con los procesos erosivos, formas de erosión existentes y riesgos de erosión.

Los mapas de suelos son modelos donde estos aparecen clasificados en unidades taxonómicas (clases) de acuerdo a sus propiedades externas e internas, y las delineaciones muestran su distribución en el espacio. Dado que los suelos son reconocidos como objetos tridimensionales (SSS, 1951), la geomorfología es usada como clave para identificar los límites de los suelos en el terreno, para explicar la génesis de suelos (Wambeke & Forbes, 1985, Zink & Valenzuela, 1990), y para estructurar bases de datos de suelos.

La aparición de las bases de datos digitales y los Sistemas de Información Geográfica imponen una serie de condiciones para la adquisición de datos, su organización y su manejo, siendo la relación suelos-paisaje un concepto fundamental para diseñar modelos conceptuales de bases de datos.

5.2.- Modelización conceptual de suelos y paisaje

La realización de cualquier prospección de suelos debe tener un objetivo. Éste sirve para identificar las características de los suelos y propiedades que son relevantes para la

aplicación. Para estudios de erosión propiedades de especial importancia son las relacionadas con:

- Estructura del suelo y estabilidad estructural.
- Infiltración y permeabilidad.

En la modelización de suelos la clasificación es un método para organizar y agrupar los atributos en clases, mientras que las delineaciones en los mapas muestran como se distribuyen geográficamente dichas clases. A pesar del hecho de que las delineaciones de suelos contengan más de una clase taxonómica, este es uno de los mejores métodos para representar los suelos, sus características y su distribución espacial.

5.2.1.- Componentes del mapa de suelos y relaciones

En los mapas de suelos, los límites se representan por medio de líneas. La delineación de estas líneas se basa en relaciones conceptuales entre el paisaje y los suelos, ya que comparten factores formadores comunes (Zink, 1988). Por ello los mapas de suelos suelen tener una entrada fisiográfica. En primer lugar se delimitan las unidades de paisaje teniendo en cuenta un sistema de clasificación. Posteriormente, se describen los suelos existentes sobre dichas unidades de paisaje. Posteriormente, en base a la prospección de suelos, se distinguen nuevos límites dentro de las unidades de relieve, los cuales constituirán las delineaciones de suelo definitivas.

El modelo de suelos-paisaje es fundamental para comprender como es posible realizar una cartografía de suelos a partir de una serie de observaciones puntuales que suponen < 1% de la superficie del suelo. Las bases para la formación de este modelo son: los conocimientos previos del área de estudio, la fotointerpretación y el estudio de suelos en áreas modelo (Martínez Casasnovas, 1992).

Para la clasificación de las unidades de paisaje se puede utilizar el sistema de taxonomía geomorfológica propuesto por Zink (1988). Éste es un sistema jerárquico que considera seis niveles categóricos:

| | |
|-------------|------------------------|
| Orden: | Geoestructura |
| Suborden: | Ambiente morfogenético |
| Grupo: | Paisaje |
| Subgrupo: | Relieve |
| Familia: | Substrato |
| Subfamilia: | Forma del terreno |

La entrada fisiográfica para cartografías semidetalladas de suelos estaría compuesta por los cuatro últimos niveles de la clasificación geomorfológica. También el nivel de agregación de la información de suelos debe estar de acuerdo a la escala del mapa. A nivel semidetallado las unidades cartográficas de suelos son, en su mayor parte, asociaciones o complejos de unidades taxonómicas. La unidad de paisaje básica será la unidad de relieve, la cual viene definida por una combinación de formas elementales del terreno.

Dado que el sistema de clasificación geomorfológica es un sistema jerárquico, las relaciones entre diferentes niveles serán del tipo uno-a-muchos (1:M). Por otra parte, una unidad de relieve podrá contener delineaciones con diferentes clases de suelos. El conjunto de las delineaciones de suelos identificadas por un mismo símbolo constituye la unidad cartográfica de suelos.

Las clases taxonómicas se definen sin tener en cuenta la distribución geográfica de los suelos. Estas clases agrupan a suelos con similares características. Cada unidad taxonómica puede ser descrita por perfiles representativos o perfiles tipo, los cuales consideran el rango de características de diferentes suelos que pertenecen a la misma clase. Las unidades taxonómicas pueden ser subdivididas en Fases de acuerdo a características potencialmente significantes para el uso o manejo del suelo que no son consideradas en la definición de las clases.

Una unidad cartográfica puede incluir una o más clases de suelos representadas por sus perfiles tipo. Estos componentes pueden diferir en tamaño, forma y patrón de distribución. Dependiendo de la composición de la unidad se pueden distinguir diferentes tipos de

unidades cartográficas: consociación, asociación y complejos. En ellas se consideran también las inclusiones de clases de suelos que son diferentes a los que dan nombre a la unidad.

Una unidad cartográfica de suelos viene determinada por cuatro componentes básicos:

- El nombre de las clases de suelos que la forman y las fases.
- La proporción de cada clase de suelos en términos de porcentaje de área dentro de la unidad.
- El patrón de distribución de las clases de suelos dentro de la unidad.
- La unidad de paisaje dentro de la que se distribuyen las clases de suelos.

Cada unidad taxonómica está relacionada con su pedón modal (representado por un perfil tipo), el cual concentra el máximo rango de variabilidad de cada perfil descrito perteneciente a la misma clase (características morfométricas, físicas, químicas, mineralógicas y biológicas).

Estas reglas y relaciones, que sirven para modelizar los suelos de la realidad a mapas, también forman la base para estructurar la información sobre suelos en bases de datos digitales (Zink & Valenzuela, 1990).

5.2.2.- Modelo conceptual de la base de datos de suelos

Partiendo de las reglas y relaciones mencionadas anteriormente, se ha diseñado el siguiente modelo para estructurar una base de datos de suelos con fines de estudios de erosión y evaluación de tierras (por extensión). El modelo conceptual se muestra en la Figura 10, y está basado en las siguientes reglas:

a.- Cada área en el mapa de suelos se denomina delineación. La delineación es la unidad espacial básica, y contiene información geométrica y temática. Se relaciona con la unidad cartográfica de suelos mediante una asociación (1:M). Una unidad cartográfica es el conjunto de delineaciones con el mismo símbolo.

- Una unidad cartográfica de suelos tiene al menos una delineación.

- Una delineación pertenece a una única unidad cartográfica de suelos.

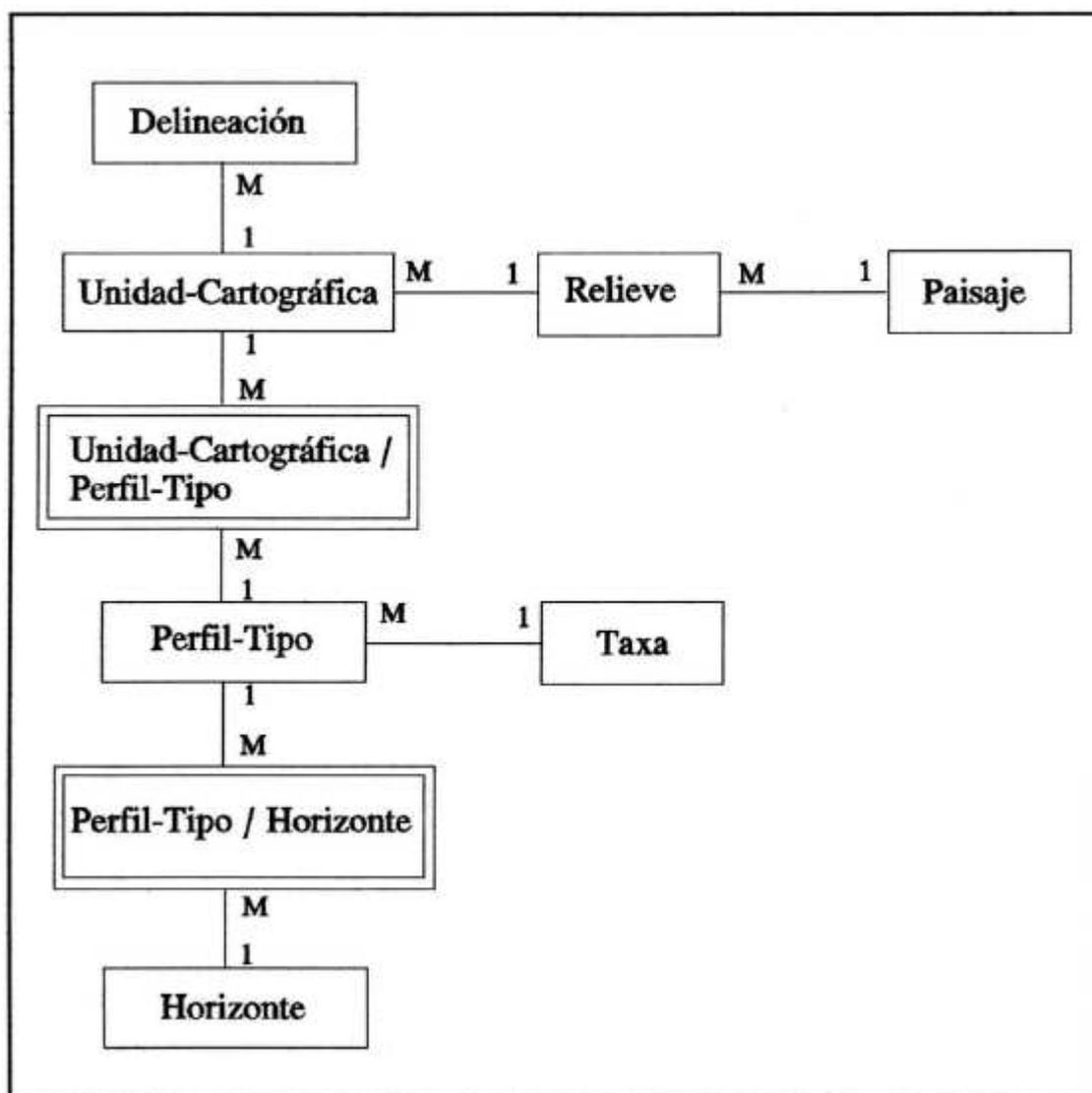


Figura 10.- Diagrama de Entidad relación que representa el modelo conceptual de la base de datos de suelos (Martinez Casanovas, 1993)

b.- La unidad cartográfica de suelos es la entidad básica que contiene información sobre los componentes del suelo y paisaje.

- Una unidad cartográfica está contenida en una unidad de relieve.

- Una unidad de relieve puede contener una o más unidades cartográficas de suelos.

- Una unidad de paisaje tiene al menos una unidad de relieve.

- Una unidad de relieve debe pertenecer a una unidad de paisaje.

- Una unidad cartográfica de suelos puede contener más de una clase de suelos, representada por su perfil tipo.

- Una clase de suelos puede estar presente en más de una unidad cartográfica.

Como relaciones del tipo muchos-a-muchos (M:N) no están permitidas según el modelo relacional, se debe crear una tabla relacional entre la entidad representando a las unidades cartográficas y a los perfiles tipo. De este modo se aparecen dos relaciones del tipo (1:M).

c.- Una unidad taxonómica de suelos, como por ejemplo una Familia (SSS, 1975, 1992) agrupa diferentes perfiles de suelos que tienen similares características.

- Una unidad taxonómica puede tener más de un perfil modal.

- Un perfil modal debe pertenecer a una única unidad taxonómica.

- La cartografía de suelos considerada en este ejemplo se realiza a escala semidetallada, siendo el nivel taxonómico considerado es el de familia. Sin embargo una familia puede tener uno o más perfiles tipo (posibles series de suelos a nivel más detallado), los cuales pueden diferir en algunos casos en propiedades importantes tales como la profundidad de suelos. Este enfoque puede limitar la utilidad de este nivel taxonómico en cartografías a escalas semidetalladas para estudios de erosión y evaluación de tierras, al menos que estas propiedades relevantes sean diferenciadas. Por ejemplo dos perfiles de la misma clase taxonómica tienen similares características como tipos y secuencia de horizontes, aunque los horizontes pueden diferir en grosor. Un modo de resolver esta situación en la estructura de datos es considerar una entidad, "Horizonte", que recoja las características comunes a los perfiles tipo de la misma clase, y distinguir las diferencias a nivel de perfil tipo. Esto conlleva una relación (M:N) entre las entidades "Perfil-Tipo" y "Horizonte". Como esto no está permitido en el modelo relacional, se debe crear una nueva tabla relacional. En esta tabla es donde se registrarán las características diferenciales.

- Un perfil tipo tiene uno o más horizontes.

- Un horizonte tipo (características comunes) puede estar presente en más de un perfil tipo.

Los diferentes valores de características como la profundidad de suelos se incorporarán a nivel de fase, y serán añadidos a la tabla relacional creada como un atributo. En caso de que la fase modifique la totalidad de la unidad cartográfica (Ej. pendiente), ésta se reflejará en la entidad de la unidad cartográfica.

5.3.- Entidades y atributos (tablas esqueleto)

Las tablas esqueleto es otra forma esquemática de representar las tablas. Consisten en el nombre de la tabla seguido de los atributos (normalmente entre paréntesis), especificando el identificador clave y los externos.

De acuerdo al modelo conceptual propuesto para la base de datos (Figura 9.1), y teniendo en cuenta las propiedades relevantes a la aplicación (estudios de erosión y evaluación de tierras), seguidamente se describen las entidades y sus atributos que contendrá la base de datos de suelos. Otras propiedades de interés para la aplicación pueden añadirse a esta estructura a conveniencia del usuario).

El diseñador de la base de datos debe adjuntar al modelo conceptual un diccionario de datos, en el que se especifique el significado de cada atributo (concepto), métodos y unidades de medida, etc.

Tablas esqueleto de la base de datos sobre suelos propuesta.

| | |
|--|---|
| <p>DELINEACIÓN >>Delineación-id Área Perímetro <<Unidad-id</p> | <p>HORIZONTE >>Horizon-id Hor-genético Hor-diagnost Color Manchas% Manchas-tamaño Manchas-color Frag-gruesos% Textura-USDA Arena-gruesa Arean-fina Limo Arcilla Estruct-grado Estruct-forma Estruct-tamaño Compactación Friabilidad Cement-grado HCl-reacción Revest% Revest-posición Acumul% Acumul-forma Acumul-comp pH CE(1:5) Mat-org% CO3Ca% Caliza-act Hdad-33kpa Hdad-1500kpa Densidad-ap</p> |
| <p>UNIDAD-CARTOGRÁFICA >>Unidad-id Relieve-id Prof-fase Pend-fase</p> | |
| <p>UNIDAD-CARTOGRÁFICA/PERFIL-TIPO >>Unidad-id >>Perfil-id Tipo-componente Porcentaje Patrón-dist</p> | |
| <p>PERFIL-TIPO >>Perfil-id Drenaje-clase Material-parent Permeabilidad Infiltración Estab-estruct Prof-restricción Horizontes <<Taxa-id</p> | |
| <p>TAXA >>Taxa-id Grangrupo Subgrupo Partículas-clase Mineralogía-clase Cal-clase Temp-clase Prof-clase FAO</p> | <p>RELIEVE >>Relieve-id <<Paisaje-id</p> |
| <p>PERFIL-TIPO/HORIZONTE >>Perfil-id >>Horizon-id Prof-min Prof-max</p> | <p>PAISAJE >>Paisaje-id Geoestructura >> <i>Identificador Clave</i> << <i>Identificador Externo</i></p> |

6.- PRÁCTICA. DISEÑO E IMPLEMENTACIÓN DE BASES DE DATOS RELACIONALES. APLICACIÓN A UNA BASE DE DATOS DE SUELOS.

6.1.- Objetivos de la práctica:

Aplicar los conocimientos teóricos adquiridos sobre el diseño de bases de datos relacionales.

Diseñar una base de datos que posteriormente se implementará en un Sistema Gestor de Base de Datos Relacional (RDBMS).

Manejo de datos y consulta de la base de datos por medio de interfaces y lenguaje SQL (Structured Query Language) y conexión con el Sistema de Información Geográfica.

6.2.- Parte 1: diseño de la base de datos

Se desea crear una base de datos que contenga la información de suelos relativa a un área de estudio. Esta base de datos se pretende utilizar como inventario de suelos y análisis del territorio.

La información relevante sobre suelos de que se dispone es:

- Mapa de suelos, con los códigos de las unidades cartográficas. (Una unidad cartográfica es el conjunto de las delineaciones del mapa con el mismo código).
- Leyenda del mapa de suelos. Para cada unidad cartográfica se indica el tipo de unidad, la clase de suelos y el porcentaje de cada clase en la unidad.
- Datos sobre perfiles representativos de las clases de suelos en el área de estudio: datos generales del perfil y de los horizontes que lo componen.

Se pide:

- a.- Diseñar una base de datos según el modelo relacional que contenga la información de suelos.

Proceso propuesto:

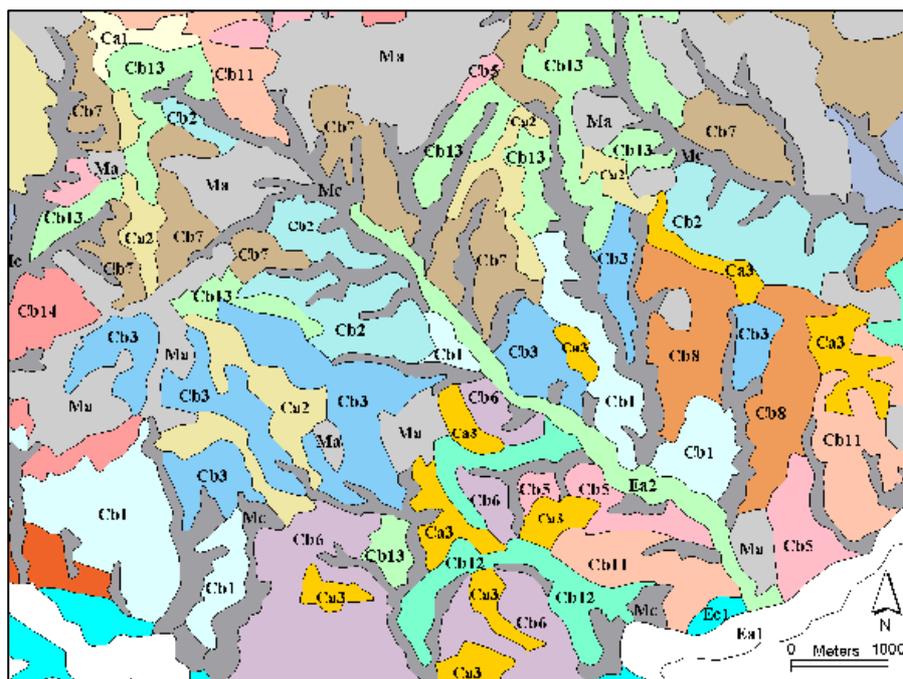
Identificar las entidades o tablas.

Determinar los atributos e identificadores.

Dibujar el diagrama de entidad-relación de la base de datos.

b.- Rellenar las tablas (dar los nombres definitivos a los atributos tal y como aparecerán en la base de datos que se implementará en el ordenador).

c.- Definición del modelo interno o físico, consistente en definir las tablas y atributos: nombre, tipo de dato (numérico o carácter), anchura del campo, etc.



Leyenda del mapa de suelos

| Código Ud. Cartográfica | Descripción |
|-------------------------|--|
| A1 | Xerorthent típico |
| A2 | Asociación Xerorthent típico (60%) y Xerofluent típico (40%) |
| B1 | Xerofluent típico |
| B2 | Complejo Xerofluent típico (65%) y Xerofluent ácuico (35%) |
| C1 | Xerochrept calcixeróllico |
| C2 | Xerochrept petrocálcico |

Perfil Tipo N₁ 101

Clase taxonómica: Xerochrept calcixeróllico
 Clase de drenaje: 5 (bien drenado)
 Clase de infiltración: 2 (buena)
 Material original: Detríticos terrígenos finos

Horizontes

| | | |
|--|----------|---------------|
| Horizonte genético | Ap | Bkn |
| Horizonte diagnóstico | Ochrico | Cálcico |
| Espesor (cm) | 40 | 110 |
| Textura (USDA) | Franca | Franco Limosa |
| Estructura | Moderada | Moderada |
| pH | 7.8 | 8.0 |
| CE 1:5 | 0.14 | 0.18 |
| CaCO ₃ % | 25.8 | 30.0 |
| Humedad -33 Kpa | 23.0 | 26.3 |
| Humedad -1500 Kpa | 7.6 | 8.5 |
| Densidad aparente (kg/m ³) | 1540 | 1440 |

Perfil Tipo N₁ 102

Clase taxonómica: Xerofluvent ácuico
 Clase de drenaje: 4 (moderadamente drenado)
 Clase de infiltración: 1 (insuficiente)
 Material original: Detríticos terrígenos finos

Horizontes

| | | |
|--|---------------|--------------------|
| Horizonte genético | Ap | Bw |
| Horizonte diagnóstico | Ochrico | - |
| Espesor (cm) | 42 | 123 |
| Textura (USDA) | Franco Limosa | Franco Arc. Limosa |
| Estructura | Débil | Moderada |
| pH | 8.3 | 8.4 |
| CE 1:5 | 0.16 | 0.15 |
| CaCO ₃ % | 30.0 | 32.0 |
| Humedad -33 Kpa | 21.2 | 23.8 |
| Humedad -1500 Kpa | 8.3 | 6.2 |
| Densidad aparente (kg/m ³) | 1570 | 1370 |

Perfil Tipo N₁ 103

Clase taxonómica: Xerofluvent típico
 Clase de drenaje: 5 (bien drenado)
 Clase de infiltración: 2 (buena)
 Material original: Detríticos terrígenos finos

Horizontes

| | | |
|--|---------|--------|
| Horizonte genético | Ap | Bw |
| Horizonte diagnóstico | Ochrico | - |
| Espesor (cm) | 44 | 106 |
| Textura (USDA) | Franca | Franca |
| Estructura | Débil | Fuerte |
| pH | 8.0 | 8.2 |
| CE 1:5 | 0.10 | 0.13 |
| CaCO ₃ % | 25.2 | 28.0 |
| Humedad -33 Kpa | 20.6 | 23.8 |
| Humedad -1500 Kpa | 7.8 | 7.9 |
| Densidad aparente (kg/m ³) | 1460 | 1620 |

Perfil Tipo N₁ 104

Clase taxonómica: Xerorthent típico
 Clase de drenaje: 5 (bien drenado)
 Clase de infiltración: 2 (buena)
 Material original: Margas

Horizontes

| | | |
|--|----------|--------------------|
| Horizonte genético | Ap | C |
| Horizonte diagnóstico | Ochrico | - |
| Espesor (cm) | 35 | 999 |
| Textura (USDA) | Franca | Franco Arc. Limosa |
| Estructura | Moderada | Material original |
| pH | 8.5 | 8.8 |
| CE 1:5 | 0.18 | 0.33 |
| CaCO ₃ % | 31.0 | 20.0 |
| Humedad -33 Kpa | 25.8 | - |
| Humedad -1500 Kpa | 12.1 | - |
| Densidad aparente (kg/m ³) | 1600 | - |

Perfil Tipo N1 105

Clase taxonómica: Xerochrept petrocálcico
Clase de drenaje: 5 (bien drenado)
Clase de infiltración: 3 (Muy buena)
Material original: Detríticos Terrígenos con gravas

Horizontes

| | | |
|---------------------------|----------|---------------------|
| Horizonte genético | Ap | Bkm |
| Horizonte diagnóstico | Ochrico | Petrocálcico |
| Espesor (cm) | 33 | 999 |
| Textura (USDA) | Franca | - |
| Estructura | Moderada | Horizonte cementado |
| pH | 7.9 | - |
| CE 1:5 | 0.14 | - |
| CaCO3 % | 25.7 | - |
| Humedad -33 Kpa | 23.2 | - |
| Humedad -1500 Kpa | 6.9 | - |
| Densidad aparente (kg/m3) | 1520 | - |

6.3.- Parte 2: implementación y consulta de la base de datos

En la Parte 1 de esta práctica se ha diseñado un modelo de Entidad-Relación de una base de datos de suelos, el cual indica las entidades que contendrá la base de datos y cómo se relacionan. Estas relaciones entre entidades se hacen por medio de identificadores que se anexan de unas tablas a otras o por medio de nuevas tablas.

En esta parte de la práctica se va a consultar la información contenida en las tablas creadas a través de un ejemplo de base de datos extraída del Mapa de Suelos del Penedès (Martínez-Casasnovas 1998). Estas consultas se van a hacer mediante ACCESS (Microsoft Office). El lenguaje de consulta a utilizar será el lenguaje SQL.

6.3.1.- Realización de consultas complejas mediante ACCESS. Consultas a partir de una tabla.

Iniciar el programa ACCESS desde el menú de inicio de Windows.

Desde el menú de Archivo, abrir la base de datos SUELO1.

Aparece una ventana con una serie de tablas. Entre ellas está la tabla SUELO.
Abrir las diferentes tablas para ver los atributos y datos que contienen

Activar la opción CONSULTAS en la ventana principal de la base de datos.

Elegir la opción NUEVO y VISTA DISEÑO.

Agregar la tabla UC y después cerrar.

a. Ejecutar la siguiente consulta para recuperar toda la información contenida en la tabla UC (unidades cartográficas).

Ir a VER, VISTA SQL

```
select * from uc;
```

(* indica que todos los atributos de la tabla son seleccionados)

Ejecutar eligiendo VISTA HOJA DE DATOS

b. Identificar todas las unidades cartográficas cuyo tipo es "CON". (Indica Consociación).

Volver a VISTA SQL y ejecutar:

```
select * from uc where tipo = "CON";
```

Estas mismas consultas pueden realizarse a través de la interface de ACCESS con SQL, que permite que las consultas sean más fáciles de realizar desde el punto de vista del usuario.

En la ventana de CONSULTAS elegir de nuevo la opción NUEVO y VISTA DISEÑO.
Agregar la tabla UC y después cerrar.

Ejecutar la siguiente consulta para recuperar toda la información contenida en la tabla UC (unidades cartográficas).

Poner el cursor en la primera casilla de CAMPO

Elegir UC.*

(* indica que todos los atributos de la tabla son seleccionados)

Ejecutar eligiendo VISTA HOJA DE DATOS

Comprobar la sentencia SQL que se ha ejecutado automáticamente mediante VISTA SQL.

Identificar todas las unidades cartográficas cuyo tipo es "CON".

Volver a VISTA DISEÑO

Poner el cursor en la primera casilla de CAMPO

Elegir CODIGO

Poner el cursor en la segunda casilla de CAMPO

Elegir TIPO

En la casilla CRITERIOS correspondiente a TIPO poner CON

Ejecutar eligiendo VISTA HOJA DE DATOS

Comprobar la sentencia SQL que se ha ejecutado automáticamente mediante VISTA SQL

Grabar la Consulta como CONSULTA1 y cerrar

Elegir una NUEVA Consulta. En vista de diseño agregar la tabla PERFIL.

Identificar los perfiles representativos de los suelos que sean bien drenados (CDRENAJE = "BIEN DRENADO") y cuyo material original sea margas (MATORIG = "MA").

Ejecutar la siguiente consulta, ya sea en modo VISTA SQL o bien en el modo INTERFASE.

```
select perfil, cdrenaje, matorig from perfil where cdrenaje = "BIEN DRENADO" and matorig = "MA";
```

Repetir la pregunta anterior pero con clase de drenaje = "BIEN DRENADO" y con material original "MA" o "DTF" (detríticos terrígenos finos).

```
select perfil, cdrenaje, matorig from perfil where cdrenaje = "BIEN DRENADO" and (matorig = "MA" or matorig = "DTF");
```

Cerrar la consulta, grabarla como CONSULTA 2, elegir una nueva consulta abriendo la tabla HORIZON.

Calcular la profundidad de suelos de cada perfil a partir de los espesores de los horizontes que componen cada perfil.

Esta consulta requiere una operación de agregación (SUMA).

select perfil, sum (espesor) from horizon group by perfil;

Ejecutar mediante VISTA DE HOJA DE DATOS

Esta misma consulta se puede realizar a través de la interfase de Access del siguiente modo en VISTA DISEÑO:

En la casilla del primer campo seleccionar el atributo PERFIL y en la segunda casilla de campo el atributo ESPESOR.

En la barra de herramientas activar el icono Σ .

En las casillas correspondientes a TOTAL aparecerá ahora la acción AGRUPAR POR. Poner el cursor en la casilla TOTAL correspondiente al atributo ESPESOR y elegir la opción SUMA.

Ejecutar la consulta mediante VISTA DE HOJA DE DATOS.

Cerrar la consulta, grabarla como CONSULTA 3, elegir una nueva consulta abriendo la tabla de nuevo la tabla HORIZON. En el primer campo seleccionar el atributo HORIZONTE.

Calcular, para cada horizonte el % máximo de humedad disponible para las plantas (kpa33 - kpa1500).

Crear una nueva consulta incorporando la tabla HORIZON

select HORIZONTE, HDAD33 -HDAD1500 from horizon;

Ejecutar mediante VISTA DE HOJA DE DATOS

Esta misma consulta se puede realizar a través de la interface de Access del siguiente modo en VISTA DISEÑO:

En la casilla del primer campo seleccionar el atributo HORIZONTE

Con el ratón en la casilla del campo 2 hacer 'click' con el botón derecho del ratón y elegir GENERAR

En la ventana de Generar Expresiones hacer doble 'click' en TABLAS y seleccionar la tabla HORIZON.

Seleccionar el atributo HDAD33 con un doble 'click'

Seleccionar el símbolo de Restar '-'

Seleccionar el atributo HDAD1500 con un doble 'click'

Aceptar

Ejecutar la consulta mediante VISTA HOJA DE DATOS

Cerrar la consulta y guardarla con el nombre CONSULTA 4.

6.3.2.- Realización de consultas complejas mediante ACCESS. Consultas a partir de más de una tabla.

Crear una nueva consulta en VISTA DE DISEÑO, agregando las tablas PERFIL y HORIZON.

Obtener el código del horizonte y el tipo de horizonte genético (HORIZGEN) de los horizontes de los perfiles con clase de drenaje = "BIEN DRENADO".

(Ejecutarla preferentemente en modo VISTA SQL)

```
select PERFIL.PERFIL, HORIZGEN from PERFIL, HORIZON where PERFIL.PERFIL = HORIZON.PERFIL and CDRENAJE = "BIEN DRENADO";
```

(Nota.- Cuando en una pregunta se incluye un atributo que existe en más de una tabla puede haber ambigüedad a la hora de cuál coger. Por ello hay que indicar a qué tabla pertenece. Ejemplo: PERFIL.PERFIL).

Crear una nueva consulta en VISTA DE DISEÑO, agregando las tablas UCPERFIL, PERFIL y TAXA.

Identificar las unidades cartográficas dónde se encuentren suelos de la clase XERORTHENT TIPICO en más de un 60 % de la unidad (visualizar también este porcentaje).

```
select UCPERFIL.CODIGO, TAXA.GGRUPO, TASA.SGRUPO, UCPERFIL.PORCENTAJE  
from UCPERFIL, PERFIL, TAXA where (UCPERFIL.PERFIL = PERFIL.PERFIL) and  
(PERFIL.TAXA = TAXA.TAXA) and TAXA.GGRUPO = "XERORTHENT" and  
TAXA.SGRUPO = "TIPICO" and UCPERFIL.PORCENTAJE > 60;
```

6.3.3.- Otras consultas.

(Anotar la pregunta SQL en la hoja de respuestas)

Identificar los horizontes cuyo contenido de CaCO₃ sea mayor que 25%.

Identificar los horizontes cuyo contenido de CaCO₃ sea mayor que 25% o el pH menor que 8.2

Calcular, para cada horizonte, la cantidad de agua disponible para las plantas (en mm), según el % de humedad y el espesor del horizonte. (Recordar que el espesor del horizonte viene dado en cm).

Calcular la humedad máxima disponible para las plantas en cada perfil, llevando a cabo la suma de la humedad máxima de cada horizonte

Identificar el código de las delineaciones de las unidades cartográficas, el tipo de las unidades cartográficas y el porcentaje dónde se da la clase de suelos "xerochrept calcixeróllico" con más del 70% (ordenar la selección por el código de la delineación de manera ascendente).

7.- REFERENCIAS BIBLIOGRÁFICAS.

Aronoff, S. 1989. *Geographic information systems: A management perspective*. WDL Publications, Ottawa, Canadá.

Burrough, P.A. 1986. *Principles of geographical information systems for land resources assessment*. Clarendon Press, Oxford, Inglaterra.

Burrough, P.A.; R.A. MacMillan & W. van Deursen, 1992. Fuzzy classification methods for determining land suitability from soil profile observations and topography. *Journal of Soil Science*, 1992, 43, pp 193-210.

Bosque, J. 1993. *Sistemas de Información Geográfica*. Editorial Rialp, S.A., Madrid.

Chen, P., 1976. The Entity-Relationship Model-Toward a unified view of data. *ACM TODSI*, Nº 1.

Date, C.J., 1990. *An introduction to database systems, Volume I*. Addison-Wesley Publish. Co., Reading, Massachusetts.

Derfler, F.J., 1992. *Guía de conectividad y redes locales*. Anaya Multimedia, S.A., Madrid.

Egenhofer, M.J. & A.U. Frank, 1989. Object-oriented modeling in GIS: inheritance and propagation. *Auto-Carto* (9). pp 588-598.

ESRI, 1991. *Map projections & Coordinate management: Concepts and procedures*. ESRI, Redlands.

Goodchild, M.F. 1988. Towards an enumeration and classification of GIS functions. In: R.T. Aangeenbrug & Y.M. Schiffman (Ed). *Int. Geo. Inf. Sys. (IGIS) Symposium: The research agenda*. AAG, Falls Church, Virginia, pp 67-77.

Healey, R.G., 1991. Database management systems. In: Maguire, D.J.; Goodchild, M. & Rhind, D. (Ed). *Geographical Information Systems. Principles and Applications*. Vol 1. Longman. Essex. pp 251-267.

Langran, G., 1992. *Time in Geographic Information Systems*. Taylor & Francis, Londres.

Laurini, R. & D. Thompson, 1990. *Fundamentals of spatial information systems*. Academic press, Londres.

NCGIA. 1990. *Core Curriculum*.

Mailing, D.H. 1991. Coordinate systems and map projections for GIS, In: Maguire, D.J.; Goodchild, M. & Rhind, D. (Ed). *Geographical Information Systems. Principles and Applications*. Vol 1. Longman. Essex. pp 135-146.

Martínez Casanovas, J.A., 1992. Mapa detallado de suelos de "Les Baixes Garrigues". *III Curso de Clasificación, Cartografía y Evaluación de Suelos*. DMCS, Lleida. (Documento de difusión restringida).

Martínez Casanovas, J.A., 1993. *Erosion modelling and conservation planning: methodological aspects and GIS. A case study in the Anoia-Penedés Vineyard Area (NE Spain)*. MSc Thesis proposal. ITC, Enschede, Holanda.

Martínez Casanovas, J.A., 1994. *Hidrographic information abstraction for erosion modelling at regional level, (a database perspective in a GIS environment)*. MSc Thesis, Wageningen Agricultural University, Wageningen, Holanda.

Molenaar, M., 1991. Status and problems of Geographical Information Systems. The necessity of a geo-information theory. *ISPRS Journal of Photogrammetry and Remote Sensing*, 46(1991), pp 85-103.

Molenaar, M., 1993. Object hierarchies and uncertainty in GIS or why is standardisation so difficult?. *GIS* (4), pp 22-28.

Molenaar, M. & D.E. Richardson, 1994. Object hierarchies for linking aggregation levels in GIS. *ISPRS Comm. IV Symposium*, May-June, Athens, USA.

Porta, J. et al., 1989. *Suelos de secano en zona semiárida*. Guía de las excursiones de la XVI Reunión de la SECS. DMCS-UPC-SECS, Lleida.

Porta, J.; M. López-Acevedo & C. Roquero, 1994. *Edafología para la agricultura y el medio ambiente*. Mundi Prensa, Madrid.

Richardson, D.E., 1993. *Automated spatial and thematic generalization using a context trasformation model*. Doctoral dissertation, Wageningen Agricultural University. R&B Publications, Ottawa, Canadá.

Rumbough, J. et a., 1991. Object-oriented modeling and design. Prentice Hall Int. Editors, New York.

Roquero, C. & Porta, J., 1992. *Agenda de Campo para el estudio del suelo*. UPM, Madrid.

SGE (Servicio Geográfico del Ejército), 1980. *Manual de topografía y lectura de planos*. SGE, Madrid.

Soil Survey Staff, 1951. *Soil Survey Manual*. Agriculture Handbook num. 18. USDA. Washington, D.C.

Soil Ssurvey Staff, 1975. *Soil Taxonomy: a basic system of soil classification for making and interpreting soil surveys*. Agric. Handbook 436. Superint of Doc, US Govt Printing Office, Washington, D.C.

Soil Survey Staff, 1992. *Keys to Soil Taxonomy, 5th revision*.

Valenzuela, C.R., 1988. *Soils geography*. ITC Journal 1988-1, 45-50 pp.

Wambeke, A. van and T. Forbes (eds.), 1985. *Guidelines for using Soil Taxonomy in the names of Soil Map Units*. SMSS Tech Monograph num. 10.

Zink, J.A. and C.R. Valenzuela, 1990. *Soil geographic database: structure and application examples*. ITC Journal 1190-3, 270-294 pp.

Zink, J.A., 1988. *Physiography and Soils*. Soil Survey Courses. Subject matter K6. ITC Lecture Notes SOL-41.