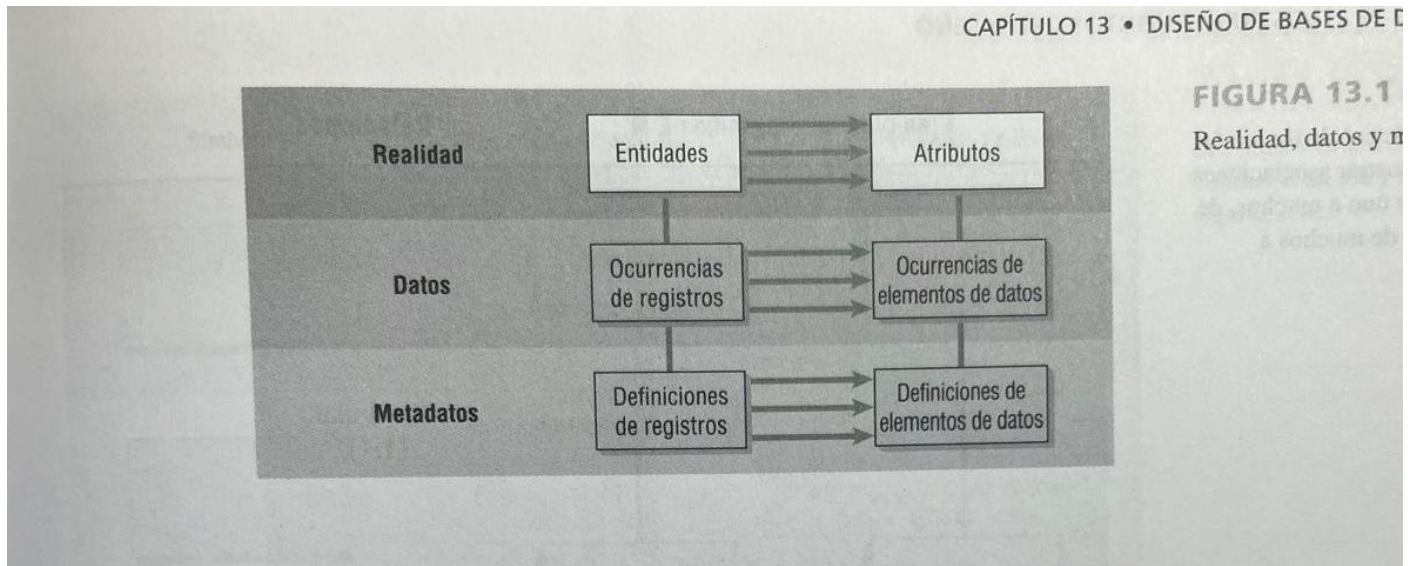


## Modelos de datos

Un modelo de datos es una colección de figuras conceptuales para describir los datos, sus relaciones, su semántica y restricciones de consistencia.



**FIGURA 13.1**  
Realidad, datos y metadatos

### Realidad, datos y metadatos

Nos referiremos al mundo real como la realidad. Los datos que se recopilen sobre personas, lugares o eventos en la realidad se almacenarán en un momento dado en un archivo o en una base de datos. Para comprender la forma y estructura de los datos, se requiere información sobre los datos en sí. La información que describe a los datos se denomina metadatos.

En la figura 13.1 se muestra la relación entre realidad, datos y metadatos. Dentro del ámbito de la realidad están las entidades y los atributos; dentro del ámbito de los datos reales están las ocurrencias de registros y las ocurrencias de elementos de datos; y dentro del ámbito de los metadatos están las definiciones de registros y las definiciones de elementos de datos. En las siguientes subsecciones hablaremos sobre el significado de estos términos.

**ENTIDADES** Cualquier objeto o evento sobre el que alguien decida recolectar datos es una entidad. También puede ser una persona, lugar o cosa (por ejemplo, un vendedor, una ciudad o un producto). Una entidad puede ser también un evento o unidad de tiempo, como la descompostura de una máquina, una venta, un mes o un año. Además de las entidades que vimos en el capítulo 2 hay una entidad menor adicional, conocida como subtipo de entidad; se representa con un rectángulo más pequeño dentro del rectángulo de entidad.

### Modelos: MR-DER-DOO-MSE

- **Modelo relacional.** El modelo relacional usa una colección de tablas para representar tanto los datos como sus relaciones. Cada tabla tiene varias columnas, y cada columna tiene un nombre único. El modelo relacional es un ejemplo de un modelo basado en registros. Los modelos basados en registros se denominan así porque la base de datos se estructura en registros de formato fijo de varios tipos. Cada tabla contiene registros de un tipo dado. Cada tipo de registro define un número fijo de campos, o atributos. Las columnas de la tabla se corresponden con los atributos.

del tipo de registro. El modelo de datos relacional es el modelo de datos más ampliamente usado, y una gran mayoría de sistemas de bases de datos actuales se basan en el modelo relacional. Los Capítulos 2 al 7 tratan el modelo relacional en detalle.

- **El modelo entidad-relación.** El modelo de datos entidad-relación (E-R) se basa en una percepción del mundo real que consiste en una colección de objetos básicos, denominados *entidades*, y de las *relaciones* entre ellos. Una entidad es una “cosa” u “objeto” del mundo real que es distinguible de otros objetos. El modelo entidad-relación se usa mucho en el diseño de bases de datos y en el Capítulo 6 se examina detalladamente.
- **Modelo de datos orientado a objetos.** El modelo de datos orientado a objetos es otro modelo de datos que está recibiendo una atención creciente. El modelo orientado a objetos se puede considerar como una extensión del modelo E-R con los conceptos de la encapsulación, los métodos (funciones) y la identidad de los objetos. En el Capítulo 9 se examina este modelo de datos.
- **Modelo de datos semiestructurados.** El modelo de datos semiestructurados permite la especificación de datos donde los elementos de datos individuales del mismo tipo pueden tener diferentes conjuntos de atributos. Esto lo diferencia de los modelos de datos mencionados anteriormente, en los que cada elemento de datos de un tipo particular debe tener el mismo conjunto de atributos. El **lenguaje de marcas extensible** (XML, eXtensible Markup Language) se emplea mucho para representar datos semiestructurados. Se estudia en el Capítulo 10.

El **modelo de datos de red** y el **modelo de datos jerárquico** precedieron cronológicamente al relacional. Estos modelos estuvieron íntimamente ligados a la implementación subyacente y complicaban la tarea del modelado de datos. En consecuencia, se usan muy poco hoy en día, excepto en el código de bases de datos antiguas que sigue estando en servicio en algunos lugares. Se describen brevemente en los Apéndices A y B para los lectores interesados.

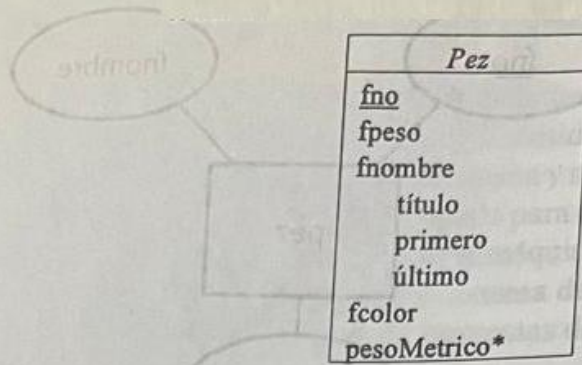
s de bases de datos y métodos de acceso

## Diagramas de entidad-relación elemental

Una aplicación relacional utiliza dos notaciones distintas para describir tablas, relaciones y restricciones. La primera notación sirve a las necesidades de los diseñadores de bases de datos cuando desarrollan las entidades de aplicación y sus relaciones. Los diagramas de entidad-relación (ER) son la notación más conocida para este propósito, y son el tema de esta sección. La segunda notación es un guión que la máquina puede leer y que presenta el diseño final (tablas, relaciones, restricciones) al sistema de administración de bases de datos (DBMS). En la siguiente sección acerca de esquemas de bases de datos se analiza esta segunda notación.

► *Los diagramas ER constituyen una notación para documentar un diseño tentativo de bases de datos. Los analistas los utilizan para facilitar el proceso de diseño. Un diagrama ER, junto con otra documentación, capta las características importantes de la aplicación, pero su traducción legible para la máquina, que es el esquema de bases de datos, en realidad lleva la información a los DBMS.*

Modelos de bases de datos y métodos de acceso



Los diagramas ER representan entidades como rectángulos, ya sea con los atributos en óvalas circundantes o con los atributos que aparecen en lista en un rectángulo. La notación especializada marca atributos de llave, valores múltiples, derivados y compuestos.

### Relaciones binarias en diagramas ER

En un diagrama de entidad-relación (ER), aparece una relación como un diamante que contiene el nombre de la relación. Unas líneas enlazan el diamante de relación a las entidades participantes. Normalmente, las líneas emanan de los vértices del diamante, que parece que limita la notación a relaciones de grado 4 o menos; pero ninguna regla expresa que las líneas deben originarse en un vértice y, por tanto, las relaciones de grado 5 o mayores son posibles. La mayor parte de las relaciones son binarias; ocasionalmente aparece una relación ternaria. Debido a que las relaciones de grado más alto son raras, los cuatro vértices del diamante suelen ser más que suficientes.

En la figura 2.17 se ilustra la relación Representa entre Especie y Pez. Recordemos que se puede tabular cualquier clase de relaciones binarias (una a muchas, muchas a muchas, o una a una) como un conjunto de pares. La relación Representa es una a muchas, y cada instancia es un par (especie, pez). Desde luego que un pez determinado aparece a lo sumo en un par, indicando una agrupación con una especie única. Una sola línea que conecta una entidad con una relación denota *participación parcial*: ciertos objetos de la clase de entidad podrían no aparecer en ningún par. En la figura 2.17, la entidad Especie participa parcialmente en la relación Representa. Esto significa que puede haber una especie que no se agrupa con ningún pez. El acuario puede haber registrado una especie en la base de datos pero no tener todavía ningún pez en esa especie. En contraste, la entidad Pez utiliza una doble línea para conectar con el diamante de relación. Esa conexión indica *participación total*, lo que significa que todo objeto de la entidad debe aparecer en por lo menos un par. Un pez debe pertenecer a una especie, o no puede existir en la base de datos. Este ejemplo muestra la forma en que la notación puede captar ciertos matices de una aplicación. En este contexto, un diagrama ER constituye una *herramienta semántica para modelar*. Un diseñador la utiliza para captar la semántica (significado) de una aplicación.

Un diagrama ER muestra una relación como un diamante, con líneas que lo conectan a sus entidades participantes. Una línea individual denota *participación parcial*; una doble línea significa *participación total*.

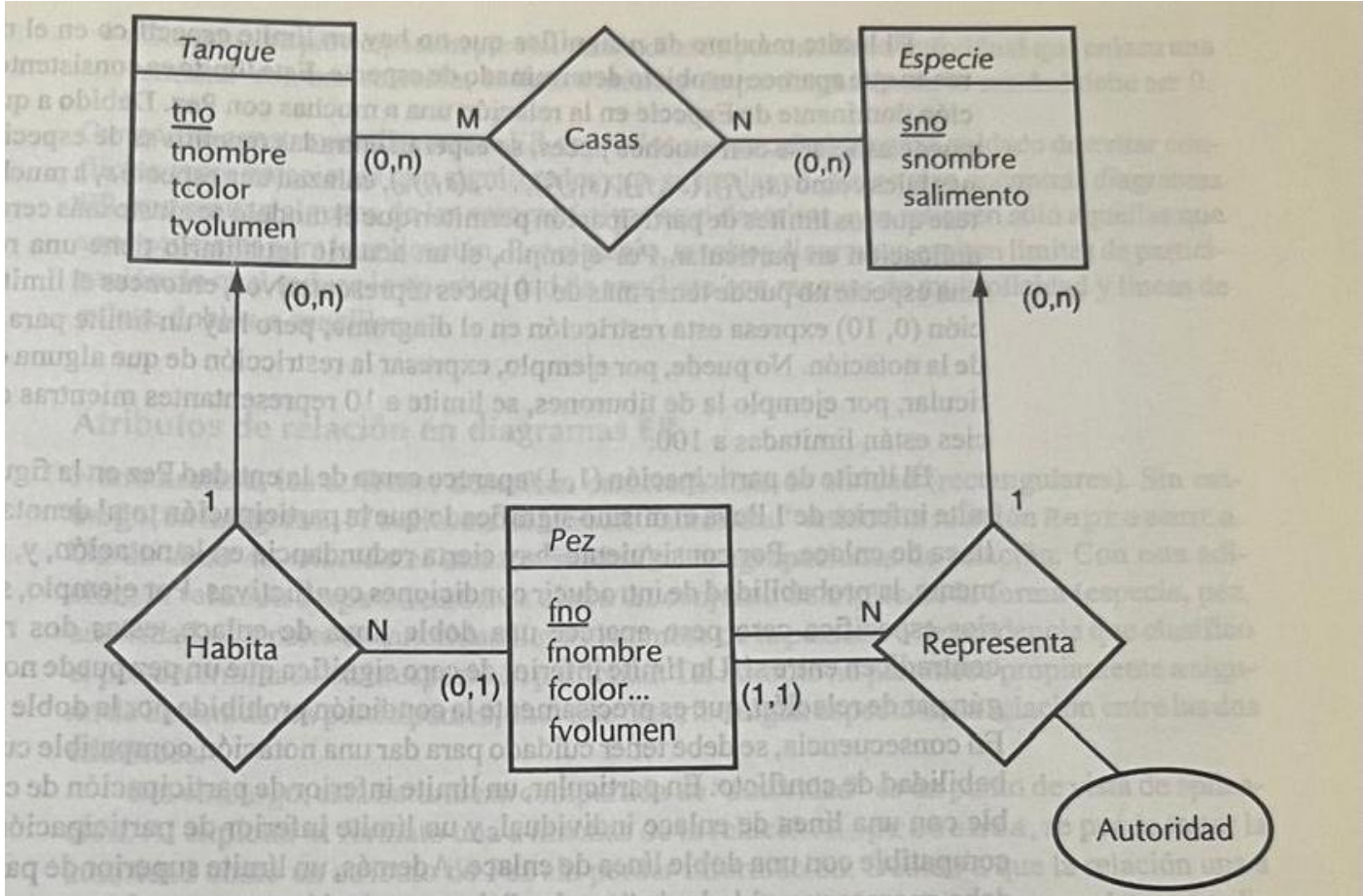


Figura 2.17. Representación de relaciones mediante un diagrama ER.

Aparece una punta de flecha en el extremo de la línea que enlaza el diamante de relación con la relación dominante de una relación una a muchas, como se ilustra en la figura 2.17. La ausencia de una punta de flecha marca la otra entidad como la parte no dominante. Del mismo modo, se pueden anotar relaciones una a muchas con un “1” o una “N” donde la línea sale del diamante de relación para unirse a la entidad dominante o no dominante, respectivamente. Un “1” en ambos lados indica una relación una a una, y las letras M y N representan una relación muchas a muchas. Las puntas de flecha en los extremos de ambas líneas conectan entidades en una relación una a una, y cuando no haya puntas de flecha en absoluto se muestra una relación muchas a muchas. Las designaciones 1:1, 1:N y M:N se denominan razones de multiplicidad.

Se sitúan límites de participación cerca de una línea de conexión donde enlace una entidad. Un límite de participación contiene dos números (x, y), donde x da el número mínimo de pares en que un objeto de entidad determinado puede participar y donde y da el número máximo. La letra n significa “muchos”, cuando el número exacto no es importante. Es decir, n significa 0, 1 o más. En la figura 2.17, el límite de participación (0, n) se presenta cerca de la entidad Especie. El límite mínimo de cero significa que una especie dada no necesita participar en ningún par de relación. Su agrupación de especie-peze puede contener sólo la especie y ningún pez. Un límite mínimo de cero lleva el mismo significado que la participación parcial denotada por la línea de conexión individual.

## Modelos de bases de datos y métodos de acceso

El límite máximo de  $n$  significa que no hay un límite específico en el número de pares en que aparece un objeto determinado de especie. Este límite es consistente con la condición dominante de Especie en la relación una a muchas con Pez. Debido a que una especie puede asociarse con muchos peces, se esperan entradas repetitivas de especies. Repeticiones tales como  $(s_1, f_1), (s_1, f_2), (s_1, f_3), \dots, (s_1, f_n)$ , enlazan una especie  $s_1$  a muchos peces. Nótese que los límites de participación permiten que el modelo se ajuste más cercanamente a la aplicación en particular. Por ejemplo, si un acuario igualitario tiene una regla en la que una especie no puede tener más de 10 peces representativos, entonces el límite de participación  $(0, 10)$  expresa esta restricción en el diagrama; pero hay un límite para la flexibilidad de la notación. No puede, por ejemplo, expresar la restricción de que alguna especie en particular, por ejemplo la de tiburones, se limite a 10 representantes mientras que otras especies están limitadas a 100.

El límite de participación  $(1, 1)$  aparece cerca de la entidad Pez en la figura 2.17. Un límite inferior de 1 lleva el mismo significado que la participación total denotada por la doble línea de enlace. Por consiguiente, hay cierta redundancia en la notación, y, desafortunadamente, la probabilidad de introducir condiciones conflictivas. Por ejemplo, si el límite inferior especifica cero pero aparece una doble línea de enlace, estas dos restricciones se contradicen entre sí. Un límite inferior de cero significa que un pez puede no aparecer en algún par de relación, que es precisamente la condición prohibida por la doble línea de enlace. En consecuencia, se debe tener cuidado para dar una notación compatible cuando haya probabilidad de conflicto. En particular, un límite inferior de participación de cero es compatible con una línea de enlace individual, y un límite inferior de participación de 1 o más es compatible con una doble línea de enlace. Además, un límite superior de participación de 1 debe aparecer en el lado de "muchas" de una relación una a muchas.

La relación Habitantes entre Tanque y Pez también aparece en la figura 2.17. Esta relación es una a muchas, pero la participación de Pez es parcial, lo que lleva a un límite inferior de participación de cero. El límite superior de participación de Pez debe ser 1 porque la entidad es el lado no dominante de una relación una a muchas. Por último, la relación Casas de la figura 2.17 ilustra una típica relación muchas a muchas. Estos ejemplos demuestran la necesidad de notación compatible. Las siguientes reglas deben ayudar a construir diagramas de entidad-relación (ER) consistentes.

- Al representar una relación binaria una a X, donde X puede ser una o muchas, el límite máximo de participación para la entidad en el lado X debe ser 1.
- Al representar una relación binaria una a X, donde X puede ser una o muchas, la conexión a la entidad en el lado "uno" debe terminar en una punta de flecha.
- Si en una relación binaria aparece un límite máximo de participación de 1 en una de las entidades, entonces la otra entidad debe recibir una punta de flecha en el extremo de su línea de enlace.
- Cuando una participación total está indicada por una doble línea que enlaza una entidad con una relación, el límite inferior de participación de la entidad debe ser 1 o más.

- Cuando una participación parcial está indicada por una línea individual que enlaza una entidad con una relación, el límite inferior de participación de la entidad debe ser 0.

Cuando se construyan diagramas ER con todos sus arreglos, téngase cuidado de evitar conflictos entre notaciones con significados que se traslapen. No es raro encontrar diagramas ER que omitan algunas de las características aquí descritas, que retienen sólo aquellas que son esenciales para la aplicación. Por ejemplo, muchos diagramas omiten límites de participación, lo cual reduce la oportunidad de conflicto con razones de multiplicidad y líneas de enlace dobles o sencillas.

### Atributos de relación en diagramas ER

Normalmente, los atributos describen características de entidad (rectangulares). Sin embargo, en la figura 2.17 se ilustra un atributo “autoridad” unido a la relación Representa. Un *atributo de relación* es una extensión de las agrupaciones de relación. Con esta adición, la relación Representa es ahora un conjunto de triples de la forma (especie, pez, autoridad). La parte de autoridad lleva el nombre de la persona o dependencia que clasificó el pez determinado en la especie en particular. Tal atributo no pertenece propiamente a alguna de las entidades participantes, sino que describe algún aspecto de la relación entre las dos entidades.

Sin embargo, esta naturaleza compartida de “autoridad” es un punto de vista de aplicación. Al explotar el formato una a muchas de la relación Representa, se puede tratar la autoridad como un atributo de Pez sin perder información. Debido a que la relación una a muchas siempre determina de manera única la especie relacionada con un pez determinado, se puede reconstruir la agrupación (especie, pez, autoridad) de la tupla de peces, usando su *sno* incrustado y su atributo de autoridad.

Sin considerar si se cambia el atributo autoridad en Pez en el diagrama de entidad-relación (ER), cuando se vuelca el diseño final en tablas relacionales, tendrá que ponerse en la tabla de peces. Sin embargo, esta conveniencia introduce una tendencia en la representación, sugiriendo que autoridad está relacionado más estrechamente con pez que con especie. Si la aplicación del acuario ve la autoridad como una característica descriptiva de un par (especie, pez), no debe ponerse dentro de la entidad pez en el diagrama ER, aun cuando finalmente se cambiará ahí cuando se construya el esquema relacional. En términos de la aplicación del acuario, se puede argumentar que autoridad debe ser una característica conjunta de un par (especie, pez). Por ejemplo, se podría cometer un error de clasificación porque no se tuvo cuidado al examinar el pez o porque no había claridad en cuanto a definir las características de las especies. En otras palabras, tanto pez como especie son igualmente sospechosos. Además, los otros atributos de peces, excepto por la llave artificial, directamente describen un pez. Los atributos nombre, color y peso sirven para distinguir un pez en particular de otro. En contraste, el atributo autoridad es un juicio externo. Por tanto, cuando se elabore el diseño de la base de datos con diagramas ER, se debe unir el atributo autoridad a la relación Representa porque esa ubicación representa de manera más estrecha la intención de la aplicación. Cuando se convierta el resultado en una base de datos relacional, se tiene que tratar con el inconveniente del modelo relacional que no permite atributos de relación. En ese punto, se debe cambiar el atributo autoridad a la tabla de Pez.

## los de bases de datos y métodos de acceso

En realidad, ya antes se encontraron atributos de relación cuando estudiamos la técnica estándar de descomposición para factorizar una relación muchas a muchas en dos relaciones una a muchas. Este proceso crea una entidad de intersección, que se convierte en el hijo común de las nuevas relaciones una a muchas, pero es frecuente que se descubra que la entidad de intersección corresponda a alguna entidad de aplicación por su propio derecho, y se pueden añadir más atributos para completar la descripción de la entidad recién descubierta. Estos atributos adicionales corresponden a atributos de relación de la relación original muchas a muchas.

Por ejemplo, un diagrama ER de la aplicación bote-marinero aparece en la parte superior de la figura 2.18. Se visualizan ejemplos de *Tripulación* como pares (bote, marinero) y se pueden especificar otros atributos para detallar la interacción entre un bote y un marinero, por ejemplo *fechaSalida...*, *condicionesMar...*, etc. Debe entenderse que *fechaSalida* se aplica a un par (bote, marinero) en el sentido de que estas fechas dicen cuándo es que el bote determinado salió con el marinero específico a bordo. Por tanto, *fechaSalida* no es un atributo descriptivo de Bote ni de Marinero. Lo mismo se aplica al atributo de *condicionesMar*. Cuando se descompone *Tripulación* como se muestra en la parte inferior de la figura, se transfieren los atributos de relación a la entidad de intersección.

### Relaciones de grado más alto en diagramas ER

En la figura 2.19 se ilustra una relación ternaria con el ejemplo de barco-carga-puerto. Aun cuando no son apropiadas las razones de multiplicidad para las relaciones no binarias, los límites de participación siguen siendo importantes. Estos valores especifican el número mínimo y máximo de agrupaciones de relación en que una tupla determinada de entidad puede participar. En este ejemplo, todos los límites de participación son  $(0, n)$ , reflejando, por ejemplo, que un barco en particular puede aparecer en muchos triples barco-carga-puerto. La relación ternaria BCP no es equivalente a un conjunto de relaciones por pares entre las entidades, sino que es equivalente a tres relaciones binarias una a muchas entre las entidades originales y una entidad de intersección (misión). Esta descomposición aparece en la parte inferior de la figura 2.19. Nótese que la participación de misión debe ser total en las tres relaciones derivadas. Una misión existe específicamente para representar una agrupación de relación BCM de la forma (barco, carga, puerto). En consecuencia, cada objeto de misión debe participar en un par de relación por medio de cada uno de los pares BM, CM, PM. Por ejemplo, la misión  $m_1$  aparece en un par SM  $(s_i, m_1)$  porque  $m_1$  es algún triple  $(s_i, c_j, p_k)$  de la relación BCP.

### Relaciones recursivas en diagramas ER

La relación recursiva prototípica una a muchas entre supervisores y trabajadores aparece en la figura 2.20. Aquí, los brazos de una relación tienen ciertos nombres de papeles, aun cuando normalmente esta notación no contribuye al contenido de información del diagrama. Por ejemplo, en la figura 2.17, se pueden rotular los brazos de la relación *Habita* con los papeles "tanque" y "pez", pero, debido a que en la relación aparecen dos entidades diferentes, estos papeles son simplemente los nombres de las entidades participantes. Podrían usarse los

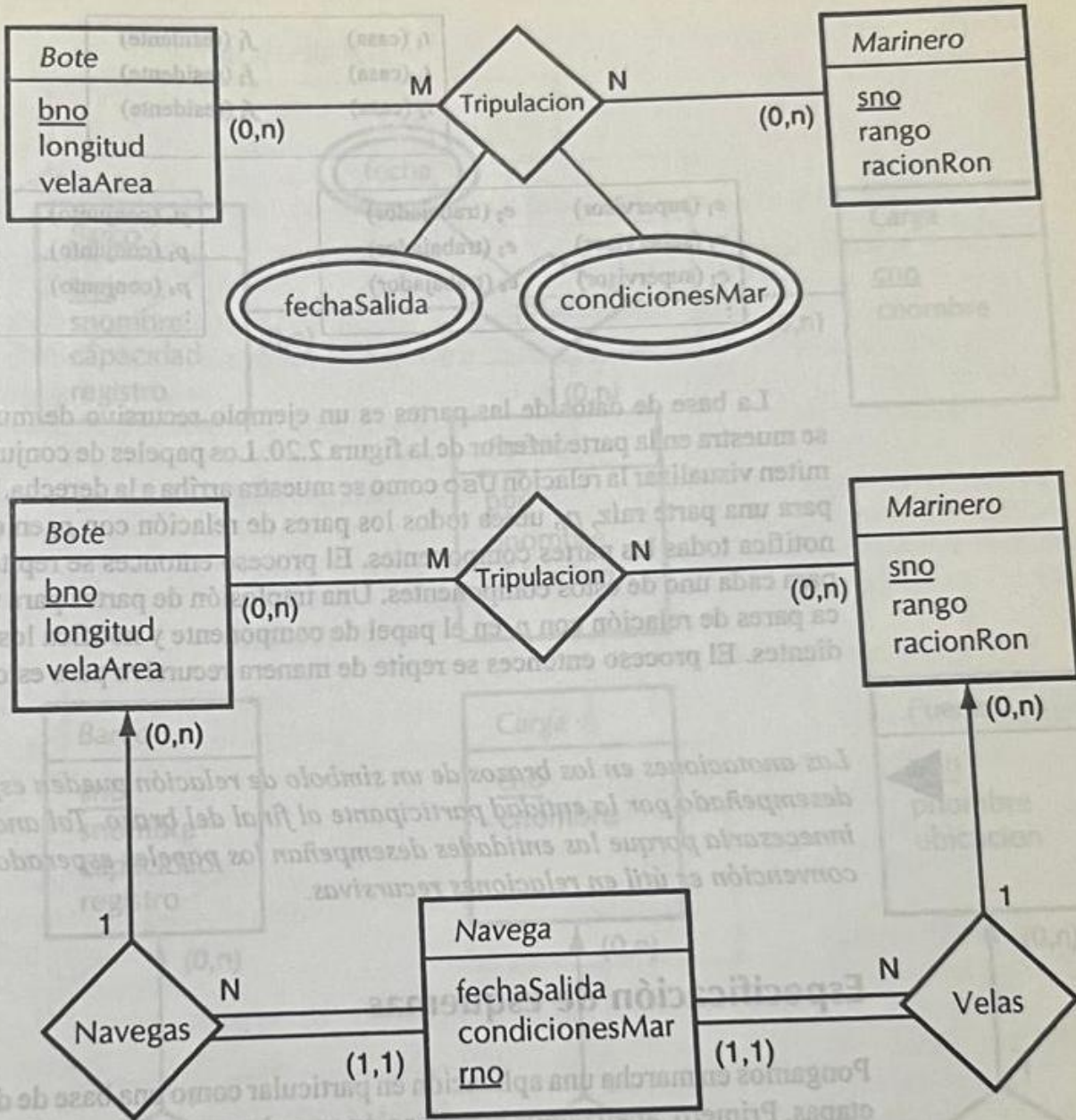
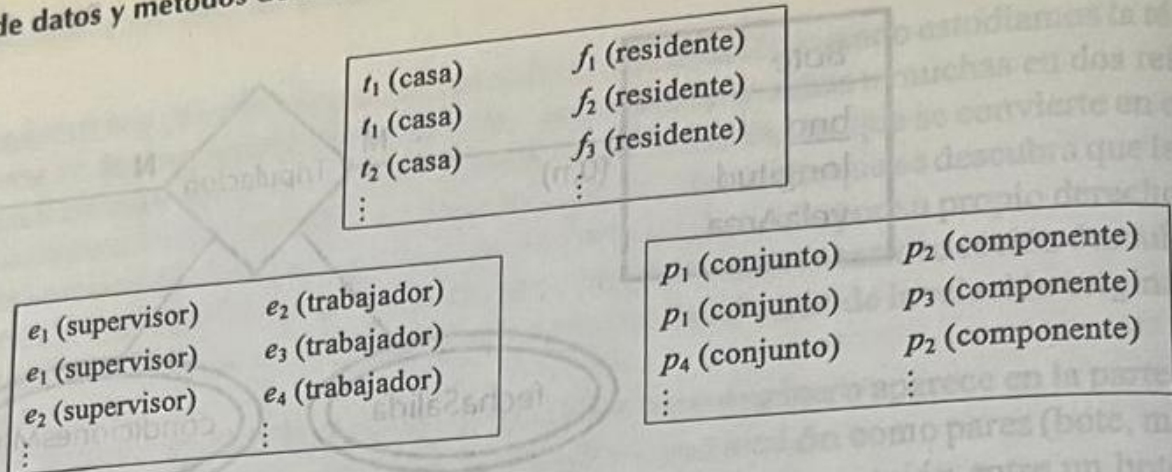


Figura 2.18. Atributos de relación que llevan al descubrimiento de una nueva entidad.

papeles “casa” y “residente” en lugar de “tanque” y “pez”, lo que lleva a una tabulación de la relación Habita como se muestra en el lado izquierdo. Claramente se ve que hay muy pocas razones al asignar de manera explícita papeles a los brazos Tanque y Pez de la relación, porque los objetos tanque y pez sirven en los papeles esperados.

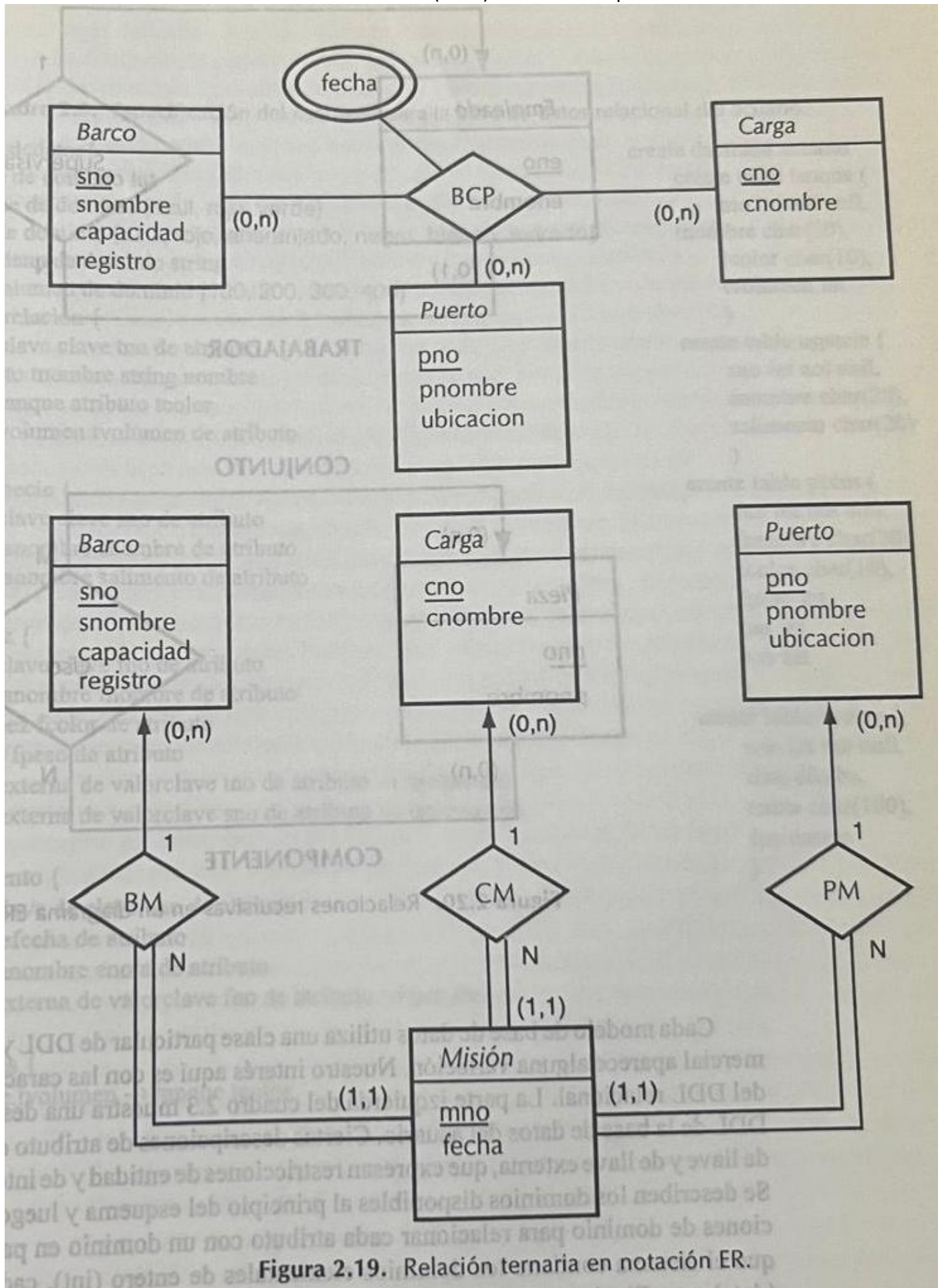
La situación es menos clara cuando una entidad entra en una relación más de una vez. La relación Supervisa de la figura 2.20 hace buen uso de las asignaciones de papel para determinar qué empleado es el supervisor en una agrupación de relación. Se puede visualizar la relación Supervisa como la tabla del centro que viene a continuación.

e bases de datos y métodos de acceso



La base de datos de las partes es un ejemplo recursivo de muchas a muchas, como se muestra en la parte inferior de la figura 2.20. Los papeles de conjunto y componente permiten visualizar la relación  $U_{so}$  como se muestra arriba a la derecha. La explosión de partes para una parte raíz,  $p_i$ , ubica todos los pares de relación con  $p_i$  en el papel de conjunto y notifica todas las partes componentes. El proceso entonces se repite de manera recursiva para cada uno de estos componentes. Una implosión de partes para una parte raíz,  $p_i$ , busca pares de relación con  $p_i$  en el papel de componente y notifica los conjuntos correspondientes. El proceso entonces se repite de manera recursiva para estos conjuntos.

*Las anotaciones en los brazos de un símbolo de relación pueden especificar el papel desempeñado por la entidad participante al final del brazo. Tal anotación suele ser innecesaria porque las entidades desempeñan los papeles esperados; pero la convención es útil en relaciones recursivas.*



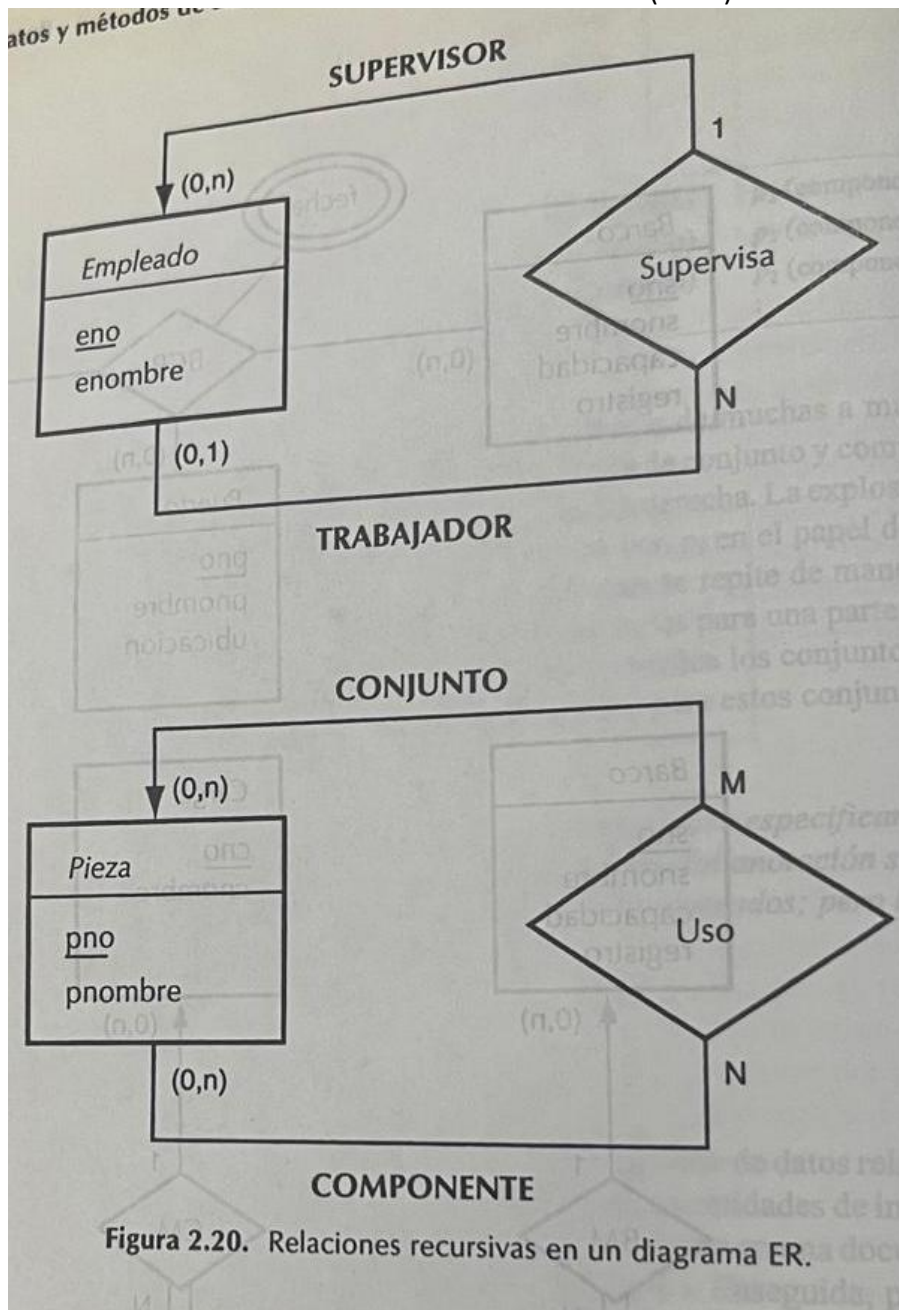


Figura 2.20. Relaciones recursivas en un diagrama ER.

MODELO DER – RELACIONES

**RELACIONES** Las relaciones son asociaciones entre entidades (algunas veces se les conoce como asociaciones de datos). La figura 13.2 es un diagrama entidad-relación (E-R) que muestra varios tipos de relaciones.

El primer tipo de relación es una relación de uno a uno (se designa como 1:1). El diagrama muestra que sólo hay un PAQUETE DE PRODUCTO para cada PRODUCTO. La segunda relación de uno a uno muestra que cada EMPLEADO tiene una OFICINA única. Cabe mencionar que todas esas entidades se pueden describir con más detalle (un precio de producto no sería una entidad; tampoco lo sería una extensión telefónica).

Otro tipo de relación es la asociación de uno a muchos (1:M) o de muchos a uno. Como se muestra en la figura, a un MÉDICO en una organización al cuidado de la salud se le asignan muchos PACIENTES, pero un PACIENTE se asigna a un solo MÉDICO. Otro ejemplo muestra que un EMPLEADO es miembro de sólo un DEPARTAMENTO, pero cada DEPARTAMENTO tiene muchos EMPLEADOS.

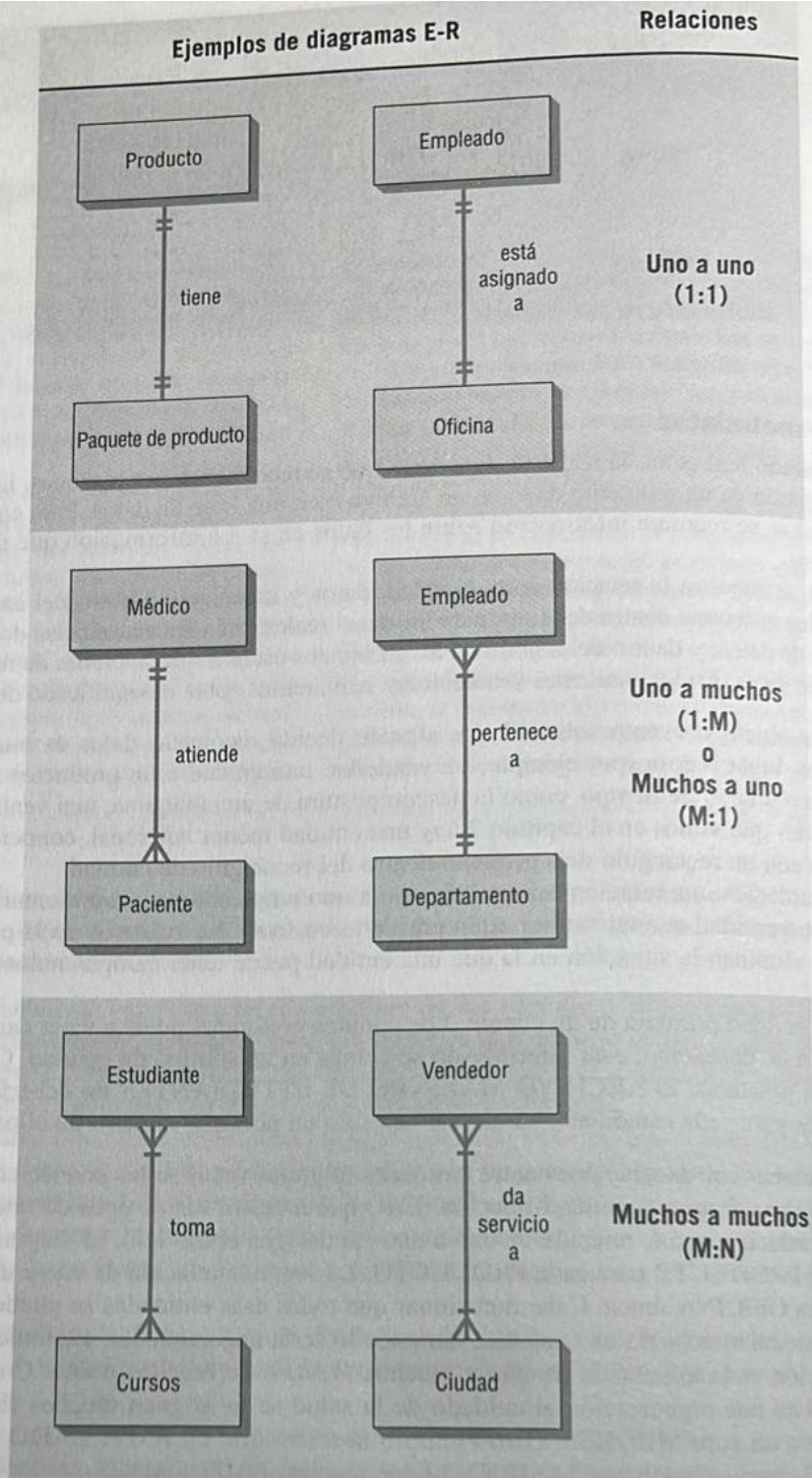
Por último, una relación de muchos a muchos (se designa como M:N) describe la posibilidad de que las entidades puedan tener muchas asociaciones en cualquier dirección. Por ejemplo, un ESTUDIANTE puede tener muchos CURSO(s) y al mismo tiempo un CURSO puede tener muchos ESTUDIANTE(s) inscritos en él. El segundo ejemplo muestra que un VENDEDOR puede llamar a muchas CIUDAD(es) y una CIUDAD puede ser un área de ventas para muchos VENDEDOR(es).

La figura 13.3 muestra los símbolos estándar para la notación de pata de cuervo, la explicación oficial de los símbolos y lo que en realidad significan. Observe que el símbolo para una entidad es un rectángulo. Una entidad se define como una clase de persona, lugar o cosa. Un rectángulo con un diamante en su interior representa a una entidad asociativa, la cual se utiliza para unir dos entidades. Un rectángulo con un óvalo en su interior representa a una entidad atributiva, la cual se utiliza para repetir grupos.

Las demás notaciones necesarias para dibujar diagramas E-R son las conexiones, de las cuales hay cinco tipos distintos. En la porción inferior de la figura se explica el significado de la notación. Cuando una línea recta

**FIGURA 13.2**

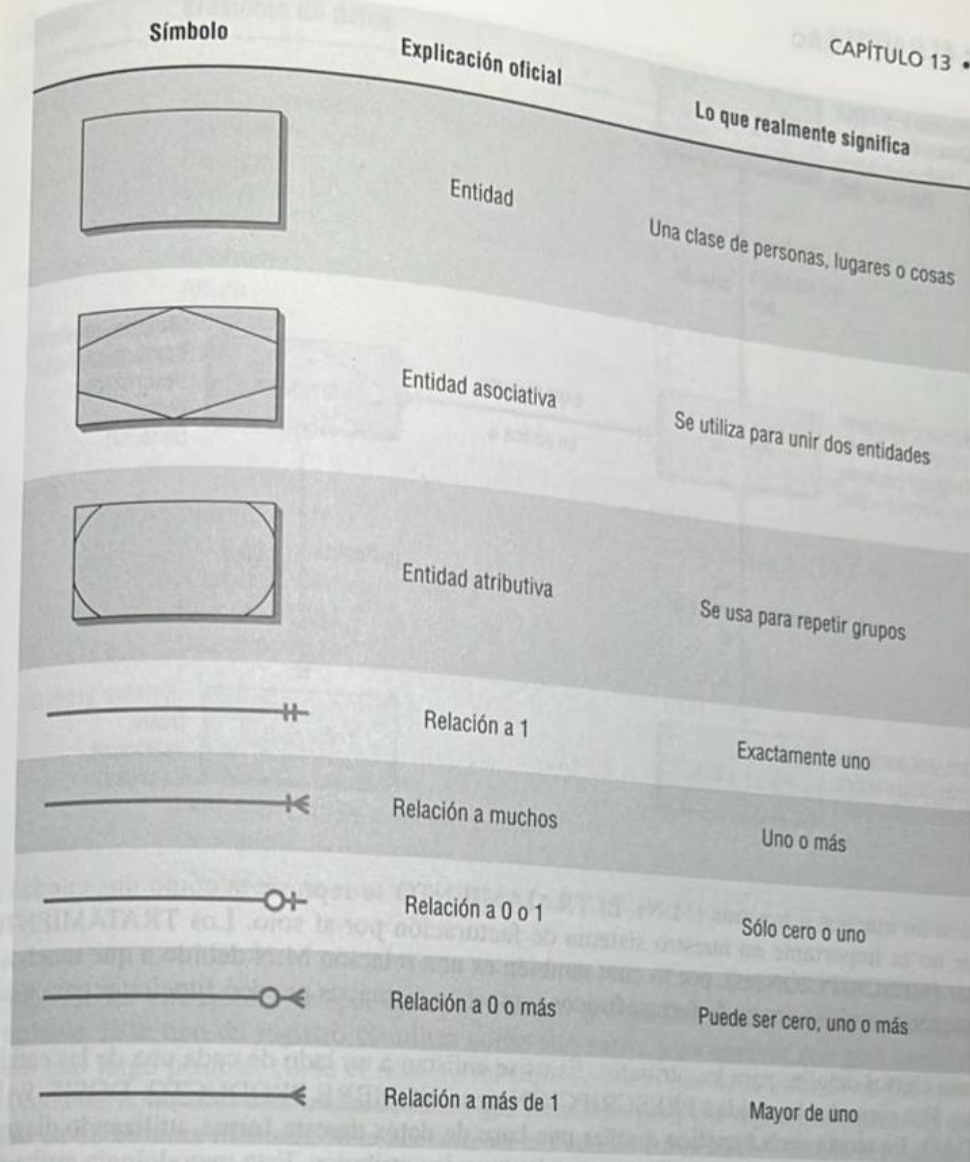
Los diagramas de entidad-relación (E-R) pueden mostrar asociaciones de uno a uno, de uno a muchos, de muchos a uno o de muchos a muchos.



conecta dos entidades simples y los extremos de la línea están marcados con dos rayas pequeñas (||), existe una relación de uno a uno. Después observará una pata de cuervo con una raya pequeña (|); cuando esta notación vincula a las entidades, indica una relación de uno a uno o de uno a muchos (a uno o más).

Las entidades vinculadas con una línea recta más una raya corta (|) y un cero (que parece más un círculo, O) describen una relación de uno a cero o de uno a uno (sólo cero o uno). Un cuarto tipo de vínculo para relacionar entidades se dibuja con una línea recta marcada en un extremo con un cero (O) seguido de una pata de cuervo. Este tipo muestra una relación de cero a cero, de cero a uno o de cero a muchos. Por último, una línea recta que vincula entidades con una pata de cuervo en un extremo describe una relación a más de uno.

Una entidad puede tener una relación que se conecta a sí misma. Este tipo de relación se denomina relación de autounión; la implicación es que debe haber una forma de vincular un registro de un archivo a otro registro



**FIGURA 13**  
Los símbolos d  
entidad-relación  
significados.

del mismo archivo. Podemos encontrar un ejemplo de relación de autounión en las simulaciones de HyperCase que se incluyen en estos capítulos. Una tarea puede tener una tarea anterior (es decir, una que se deba completar antes de poder empezar la tarea actual). En esta situación, un registro (la tarea actual) apunta a otro registro (la tarea anterior) en el mismo archivo.

Las relaciones en palabras se pueden escribir a lo largo de la parte superior o lateral de cada línea conectora. En la práctica se ve la relación en una dirección, aunque podemos escribir relaciones en ambos lados de la línea, cada una de las cuales representa el punto de vista de una de las dos entidades (vea el capítulo 2 para obtener más detalles sobre cómo dibujar diagramas E-R).

**UN EJEMPLO DE ENTIDAD-RELACIÓN** En la figura 13.4 podemos ver un ejemplo de un diagrama entidad-relación que contiene muchas entidades, muchos tipos distintos de relaciones y numerosos atributos. El diagrama E-R tiene que ver con un sistema de facturación, en especial con la parte del sistema relacionada con la prescripción (por cuestión de simpleza vamos a suponer que las visitas al consultorio se manejan de una manera distinta y están fuera del alcance de este sistema).

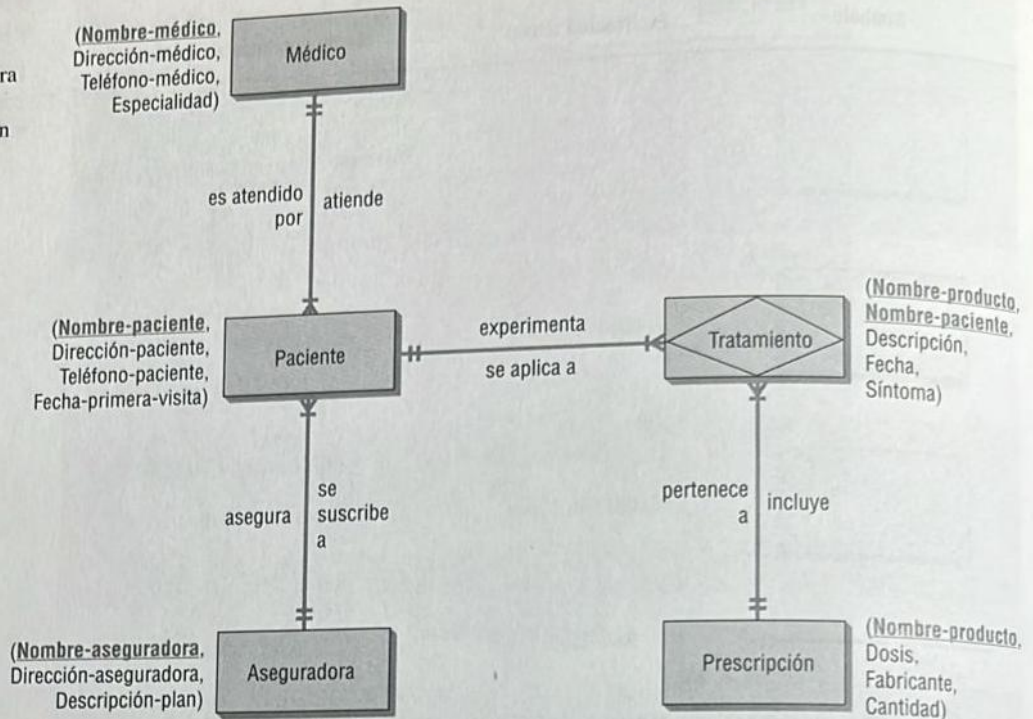
Las entidades son PRESCRIPCIÓN, MÉDICO, PACIENTE y COMPAÑÍA DE SEGUROS. La entidad TRATAMIENTO no es importante para el sistema de facturación, pero forma parte del diagrama E-R debido a que se utiliza como puente para PRESCRIPCIÓN y PACIENTE. Por lo tanto la dibujamos como entidad asociativa en la figura.

Aquí, un MÉDICO atiende a muchos PACIENTE(s) (1:M), cada uno de los cuales está suscrito a una COMPAÑÍA DE SEGUROS individual. Desde luego que el PACIENTE es sólo uno de los muchos pacientes que susciben a esa COMPAÑÍA DE SEGUROS (M:1) específica.

Para completar los registros del MÉDICO, éste necesita mantener información sobre los tratamientos que ha llevado un PACIENTE. Muchos PACIENTE(s) experimentan muchos TRATAMIENTO(s), lo cual la co

**FIGURA 13.4**

El diagrama entidad-relación para el tratamiento de pacientes. Los atributos se pueden enlistar en un lado de las entidades. En cada caso, la clave está subrayada.



vierte en una relación de muchos a muchos (M:N). El TRATAMIENTO se representa como una entidad asociativa debido a que no es importante en nuestro sistema de facturación por sí solo. Los TRATAMIENTO(s) pueden incluir tomar PRESCRIPCIÓN(es), por lo cual también es una relación M:N debido a que muchos tratamientos pueden requerir combinaciones de fármacos y muchos fármacos pueden funcionar para muchos tratamientos.

Después se llenan ciertos detalles para los atributos. Éstos se enlistan a un lado de cada una de las entidades y se subraya la clave. Por ejemplo, la entidad PRESCRIPCIÓN tiene NOMBRE PRODUCTO, DOSIS, FABRICANTE y CANTIDAD. En teoría sería benéfico diseñar una base de datos de esta forma, utilizando diagramas entidad-relación y después completando los detalles concernientes a los atributos. Esta metodología arriba-abajo es conveniente, pero algunas veces es difícil de lograr.

**ATRIBUTOS** Un atributo es cierta característica de una entidad. Puede haber muchos atributos para cada entidad. Por ejemplo, un paciente (entidad) puede tener muchos atributos tales como apellido paterno, primer nombre, dirección, ciudad, estado, etcétera; la fecha de la última visita del paciente, así como los detalles de las prescripciones también son atributos. Cuando construimos el diccionario de datos en el capítulo 8, el elemento más pequeño que describimos se denominó elemento de información. Al hablar sobre archivos y bases de datos, estos elementos de datos se denominan comúnmente elementos de datos. De hecho, los elementos de datos son las unidades más pequeñas en un archivo o base de datos. El término *elemento de datos* se puede intercambiar con la palabra *atributo*.

Los elementos de datos pueden tener valores. Estos valores pueden ser de longitud fija o variable; pueden ser caracteres alfabéticos, numéricos, especiales o alfanuméricos. En la figura 13.5 encontrará ejemplos de elementos de datos y sus valores.

Algunas veces a un elemento de datos también se le conoce como campo. Sin embargo, un campo representa algo físico, no lógico. Por lo tanto, muchos elementos de datos se pueden empaquetar en un campo; el campo se puede leer y convertir en varios elementos de datos. Un ejemplo común de esto es guardar la fecha en un solo campo como MM/DD/AAAA. Para ordenar el archivo por fecha, se extraen tres elementos de datos separados del campo y se ordenan primero por AAAA, después por MM y finalmente por DD.

**REGISTROS** Un registro es una colección de elementos de datos que tienen algo en común con la entidad descrita. La figura 13.6 es una ilustración de un registro con muchos elementos de datos relacionados. El registro que se muestra es para un pedido realizado con una empresa de pedidos por correo. PEDIDO-#, APELLIDO PATERNO, INICIAL, DIRECCIÓN CALLE, CIUDAD, ESTADO y TARJETA CRÉDITO son todos atributos. La mayoría de los registros son de longitud fija, por lo que no hay necesidad de determinar la longitud del registro cada vez que se utiliza.

| Entidad  | Elemento de datos                | Valor                           |
|----------|----------------------------------|---------------------------------|
| Vendedor | Número vendedor                  | 87254                           |
|          | Nombre vendedor                  | Kaytell                         |
|          | Nombre empresa                   | Music Unlimited                 |
|          | Dirección                        | 45 Arpeum Circle                |
|          | Ventas                           | \$20,765                        |
| Paquete  | Anchura                          | 2                               |
|          | Altura                           | 16                              |
|          | Longitud                         | 16                              |
|          | Peso                             | 3                               |
|          | Dirección postal                 | 765 Dulcinea Drive              |
|          | Dirección retorno                | P.O. Box 341, Spring Valley, MN |
| Pedido   | Producto(s)                      | B521                            |
|          | Descripción(es)                  | Disco compacto "Mi Bella Dama"  |
|          | Cantidad ordenada                | 1                               |
|          | Apellido paterno del cliente     | Kiley                           |
|          | Primera inicial                  | R.                              |
|          | Dirección calle                  | 765 Dulcinea Drive              |
|          | Ciudad                           | La Mancha                       |
|          | Estado                           | CA                              |
|          | Código postal                    | 93407                           |
|          | Número tarjeta crédito           | 65-8798-87                      |
|          | Fecha en que se colocó el pedido | 01/03/2010                      |
|          | Cantidad                         | \$6.99                          |
|          | Estado                           | Pedido pendiente                |

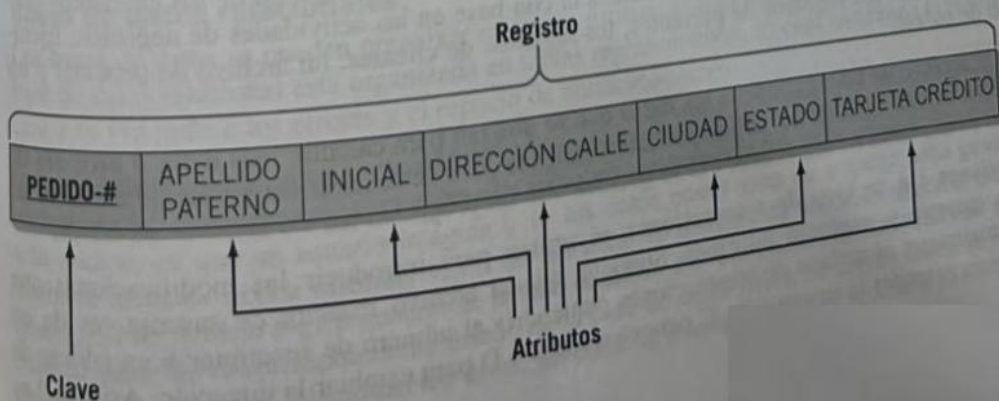
**FIGURA 13.5**  
Los valores típicos asignados a los elementos de datos pueden ser números, caracteres alfabéticos, caracteres especiales y combinaciones de los tres.

Bajo ciertas circunstancias (por ejemplo, cuando el espacio es una prioridad) se utilizan registros de longitud variable. Este tipo de registro se utiliza como alternativa para reservar una gran cantidad de espacio para el registro más largo posible, como el número máximo de veces que haya visitado un paciente a un médico. Cada visita contendría muchos elementos de datos que formarían parte del registro completo del paciente (o carpeta de archivo en un sistema manual). Más adelante en el capítulo hablaremos sobre la normalización de una relación. La normalización es un proceso que elimina a los grupos repetitivos que se encuentran en los registros de longitud variable.

**CLAVES** Una clave es uno de los elementos de datos en un registro que se utiliza para identificarlo. Cuando una clave identifica a un registro en forma única, se le llama clave primaria. Por ejemplo, PEDIDO-# puede ser una clave primaria, ya que sólo se asigna un número a cada pedido del cliente. De esta forma, la clave primaria identifica a la entidad del mundo real (pedido del cliente).

Hay que tener cuidado especial a la hora de diseñar la clave primaria. A menudo es un número secuencial o un número secuencial con un número autoverificable (llamado dígito de verificación) al final. A veces hay cierto significado integrado en la clave primaria, pero definir una clave primaria con base en un atributo se considera un riesgo: si cambia el atributo, la clave primaria también cambiará y se creará una dependencia entre la clave primaria y los datos.

**FIGURA 13.6**  
Un registro tiene una clave primaria y puede tener atributos.



PARTE IV • LOS FUNDAMENTOS DEL DISEÑO

Un ejemplo de clave primaria basada en datos es el uso de una abreviatura de estado para el nombre de un estado o el código de equipaje de una aerolínea para el nombre de un aeropuerto. A un atributo o colección de atributos que pueden servir como clave primaria se le denomina clave candidata. Una clave primaria también debe ser mínima y no debe contener atributos adicionales que sean necesarios para identificar a un registro.

A una clave se le denomina clave secundaria si no puede identificar a un registro en forma única. Las claves secundarias pueden ser únicas o pueden identificar a varios registros en una base de datos. Las claves secundarias se pueden utilizar para seleccionar un grupo de registros que pertenezcan a un conjunto (por ejemplo, pedidos del estado de Virginia).

Cuando no es posible identificar a un registro en forma única mediante el uso de los elementos de datos encontrados en un registro, podemos construir una clave al elegir dos o más elementos de datos y combinarlos. A esta clave se le denomina clave concatenada o compuesta. Cuando se utiliza un elemento de datos como clave en un registro, la descripción se subraya. Por lo tanto, en el registro de PEDIDO (PEDIDO-# APELLIDO PATERNO, INICIAL, DIRECCIÓN CALLE, CIUDAD, ESTADO, TARJETA CRÉDITO) la clave es PEDIDO-#.

Si un atributo es una clave en otro archivo, se debe subrayar con una línea punteada. Algunas bases de datos permiten al desarrollador usar un identificador de objetos (OID), una clave única para cada registro en la base de datos, no sólo en una tabla. Dado un identificador de objetos, se obtendrá un registro sin importar la tabla en la que exista. Éste se puede incluir con un pedido o una confirmación de pago, junto con un mensaje como: "Éste es su número de confirmación".

**METADATOS** Los metadatos son datos sobre los datos del archivo o base de datos. Los metadatos describen el nombre proporcionado y la longitud asignada a cada elemento de datos. Los metadatos también describen la longitud y la composición de cada uno de los registros.

La figura 13.7 es un ejemplo de metadatos para una base de datos para cierto software genérico. La longitud de cada elemento de datos se indica de acuerdo con una convención, donde 7.2 significa que se reservan siete espacios para el número, dos de los cuales están a la derecha del punto decimal. La letra N significa "numérico" y la A significa "alfanumérico". La D representa a la "fecha" y se establece de manera automática en la forma MM/DD/AAAA. Algunos programas como Microsoft Access usan inglés o español simple para los metadatos, por lo que se utilizan palabras como *text (texto)*, *currency (moneda)* y *number (número)*. Microsoft Access provee un valor predeterminado de 50 caracteres como longitud de campo para los nombres, lo cual está bien si trabajamos con sistemas pequeños, sin embargo, si va a trabajar con una base de datos extensa para un banco o una compañía de servicios públicos, por ejemplo, no es conveniente que dedique todo ese espacio a ese campo, pues de lo contrario la base de datos crecería mucho y se llenaría de espacio desperdiciado. Aquí es cuando podemos usar metadatos para planear con anticipación y diseñar una base de datos más eficiente.

## Archivos

Un archivo contiene grupos de registros que se utilizan para proveer información para operaciones, planeación, administración y toma de decisiones. Primero hablaremos sobre los tipos de archivos utilizados y después describiremos las diversas formas en que se pueden organizar los archivos convencionales.

**TIPOS DE ARCHIVOS** Podemos usar los archivos para guardar datos durante un periodo indefinido o almacenarlos provisionalmente con un propósito específico. Los archivos maestros y los archivos de tablas se utilizan para almacenar datos por un periodo prolongado. Los archivos temporales por lo general se denominan archivos de transacciones, archivos de trabajo o archivos de informes.

**Archivos maestros** Los archivos maestros contienen registros para un grupo de entidades. Estos atributos se pueden actualizar con frecuencia, pero los registros en sí son relativamente permanentes. Estos archivos tienden a tener registros extensos que contienen toda la información sobre una entidad de datos. Por lo general cada registro contiene una clave primaria y varias claves secundarias.

Aunque el analista tiene la libertad de ordenar los elementos de datos en un archivo maestro en cualquier orden, una disposición habitual es colocar el campo de la clave primaria primero, seguido de los elementos descriptivos y al último los elementos que cambian con frecuencia con base en las actividades de negocios. Ejemplos de un archivo maestro son los registros de pacientes, los registros de clientes, un archivo de personal y un archivo de inventario de piezas.

**Archivos de tablas** Un archivo de tablas contiene los datos que se utilizan para calcular más datos o medidas de desempeño. Un ejemplo es una tabla de las tarifas postales empleada para determinar los costos de envío de un paquete; otro ejemplo es una tabla de impuestos. Por lo general sólo un programa lee los archivos de tablas.

**Archivos de transacciones** Un archivo de transacciones se utiliza para introducir las modificaciones que actualizan el archivo maestro y producen informes. Suponga que el archivo maestro de suscriptores de un periódico necesita actualizarse; el archivo de transacciones contendría el número de suscriptor y un código de transacción, como E para extender la suscripción, C para cancelarla o D para cambiar la dirección. Así sólo hay

| Elemento de datos              | Valor |     |
|--------------------------------|-------|-----|
| Número de vendedor             | N     |     |
| Nombre del vendedor            | A     | 5   |
| Nombre de la empresa           | A     | 20  |
| Dirección                      | A     | 26  |
| Ventas                         | N     | 36  |
| Anchura                        | N     | 9.2 |
| Altura                         | N     | 2   |
| Longitud                       | N     | 2   |
| Peso                           | N     | 2   |
| Dirección de envío             | A     | 2   |
| Dirección del remitente        | A     | 36  |
| Producto(s)                    | A     | 36  |
| Descripción(es)                | A     | 4   |
| Cantidad solicitada            | A     | 30  |
| Apellido del cliente           | N     | 2   |
| Primera inicial                | A     | 24  |
| Dirección (calle)              | A     | 1   |
| Ciudad                         | A     | 28  |
| Estado                         | A     | 12  |
| Código postal                  | A     | 2   |
| Número de tarjeta de crédito   | N     | 9   |
| Fecha en que se hizo el pedido | N     | 10  |
| Cantidad                       | D     | 8   |
| Estado                         | \$    | 7.2 |
|                                | A     | 22  |

Campos

N Numérico  
 A Alfanumérico o texto  
 D Fecha MM/DD/AAAA  
 \$ Moneda  
 M Memo

**FIGURA 13.7**  
 Los metadatos incluyen una descripción de la apariencia que debe tener el valor de cada elemento de datos.

7.2 significa que el campo contiene 7 dígitos, dos de los cuales están a la derecha del punto decimal.

Se pueden especificar formatos especiales para los campos.  
 MM/DD/AAAA

que introducir la información relevante a las necesidades de actualización; es decir, la longitud de la renovación si es E y la dirección si es D. No se requeriría información adicional si se cancelara la suscripción. El resto de la información ya existe en el archivo maestro. Los archivos de transacciones pueden contener varios tipos de registros, como los tres que se utilizan para actualizar el archivo maestro de suscripciones al periódico, con un código en el archivo de transacciones para indicar el tipo de transacción.

**Archivos de informes** Cuando es necesario imprimir un informe y no hay una impresora disponible (por ejemplo, cuando está ocupada imprimiendo otros trabajos), se utiliza un archivo de informes. Al proceso de enviar la salida a un archivo en vez de enviarla a una impresora se le conoce como puesta en cola, o "spooling". Después, cuando el dispositivo está listo se puede imprimir el documento. Los archivos de informes son muy útiles, ya que los usuarios pueden llevar los archivos a otros sistemas computarizados y enviarlos como salida a dispositivos especializados.