

CURSO SOBRE EL MODELADO DE DATOS GEOREFERENCIADOS

Traducido y adaptado por:

Rafael Arce Mesén

Fuente: Dr. Ivan Bédard. 1993. Université Laval,
Québec. Curso: SCG-63850

Objetivos del curso

- Determinar la importancia del modelado en el proceso de elaboración de un SIG.
- Comprender la problemática de un enfoque por la vía de los datos versus un enfoque por la vía de los tratamientos.
- Dominar el formalismo individual.
- Dominar los componentes aportados por MODUL-R, para un formalismo adaptado a la problemática de los SIG.
- Efectuar la normalización de un modelo conceptual de datos.
- Producir la documentación ligada a los modelos de datos.

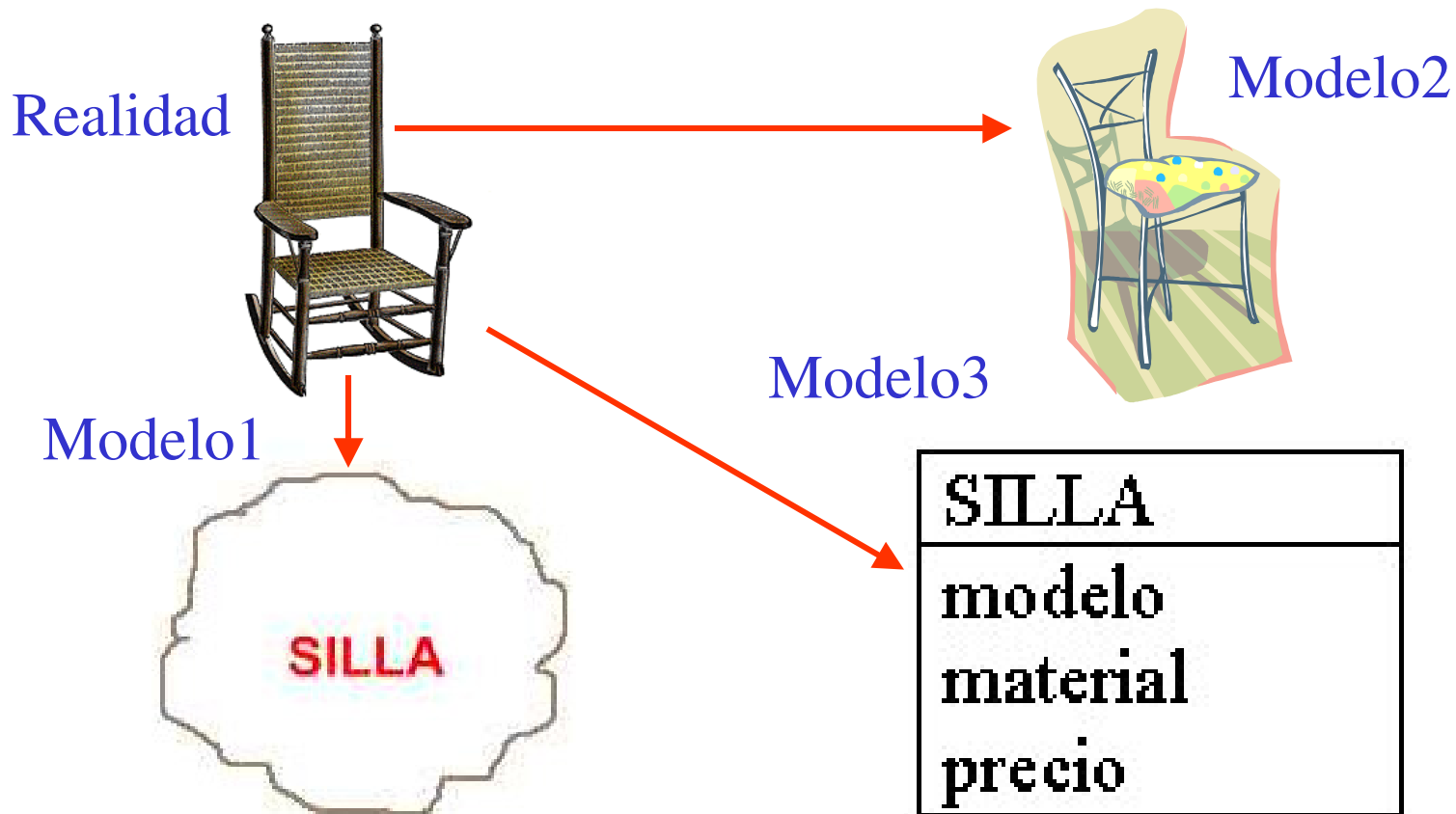
Tabla de materias

- 1. El concepto de modelado, algunos principios de base.
- 2. El modelado: datos y tratamientos; distinguir los datos de los tratamientos.
- 3. Los tipos de modelos de datos utilizados en un SIG; muchos niveles de modelos y muchos formalismos.
- 4. El formalismo individual, sus componentes y sus reglas de construcción.
- 5. El formalismo MODUL-R, un formalismo adaptado a la referencia espacial.
- 6. El formalismo entidad-relación de Chen, un sobrevuelo de este formalismo.
- 7. Normalización del MCD, las reglas de normalización.
- 8. El diccionario de datos georeferenciados, sus componentes.
- 9. El modelo lógico de datos (formalismo relacional).
- 10. El modelado de datos en el ambiente de los SIG, dónde situar los diferentes modelos en el proceso de desarrollo de un SIG.

El concepto de modelado

- Objetivos:
 - Comprender el concepto de modelado
 - Definir lo que es un modelo
 - Definir los objetivos del modelado
 - Captar las limitaciones ligadas a un modelo

De un objeto cotidiano, nosotros creamos diferentes modelos

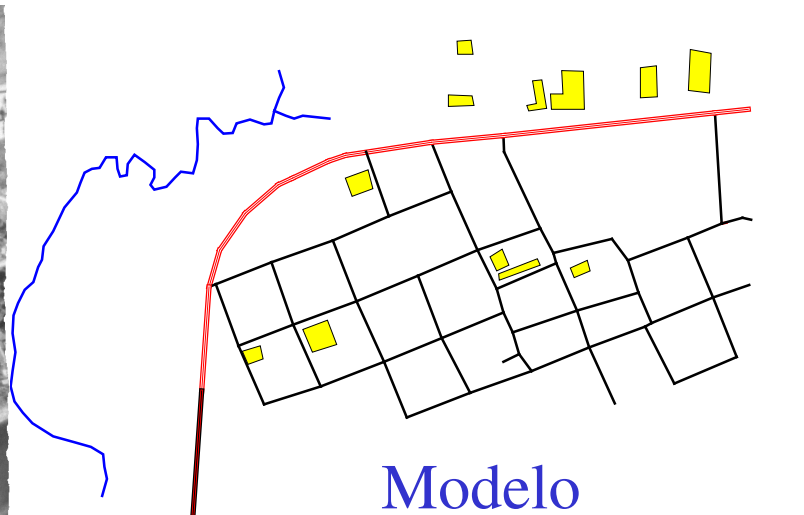


¿Qué es un modelo?

- Un modelo es una representación simplificada de la parte de la realidad que nos interesa.
- Un modelo es aceptable si el resultado de una operación efectuada sobre ese modelo es considerado como equivalente al resultado de la operación correspondiente efectuada sobre la realidad.
- Los objetivos de un modelo son:
 - 1. Comprender la realidad.
 - 2. Comunicar.
 - 3. Recordar.

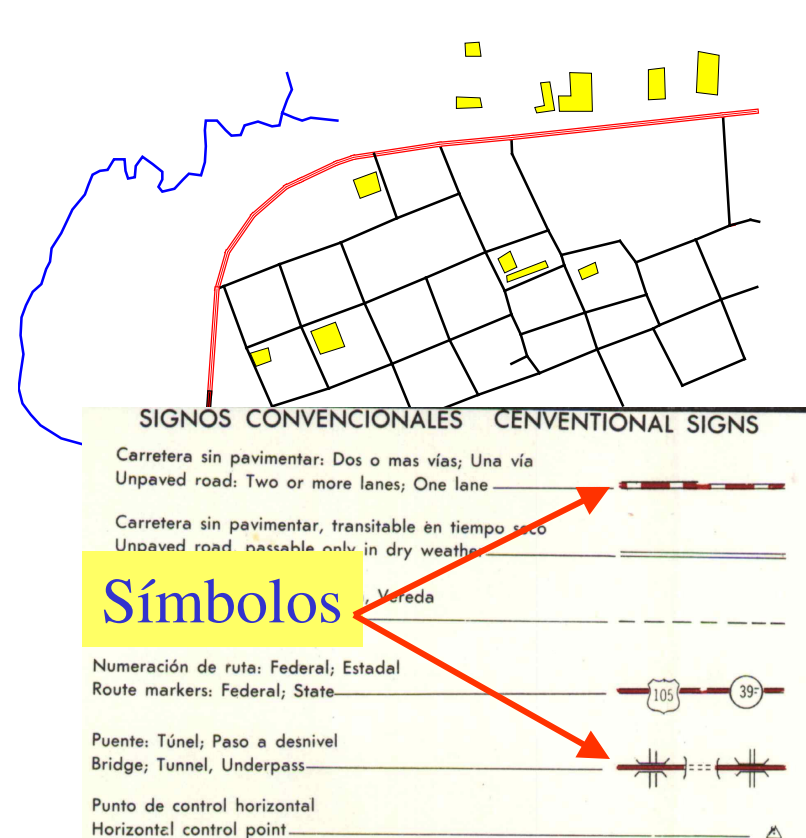
Modelo = abstracción en función de un propósito específico

- Los elementos esenciales son resaltados
- Los detalles no esenciales son eliminados



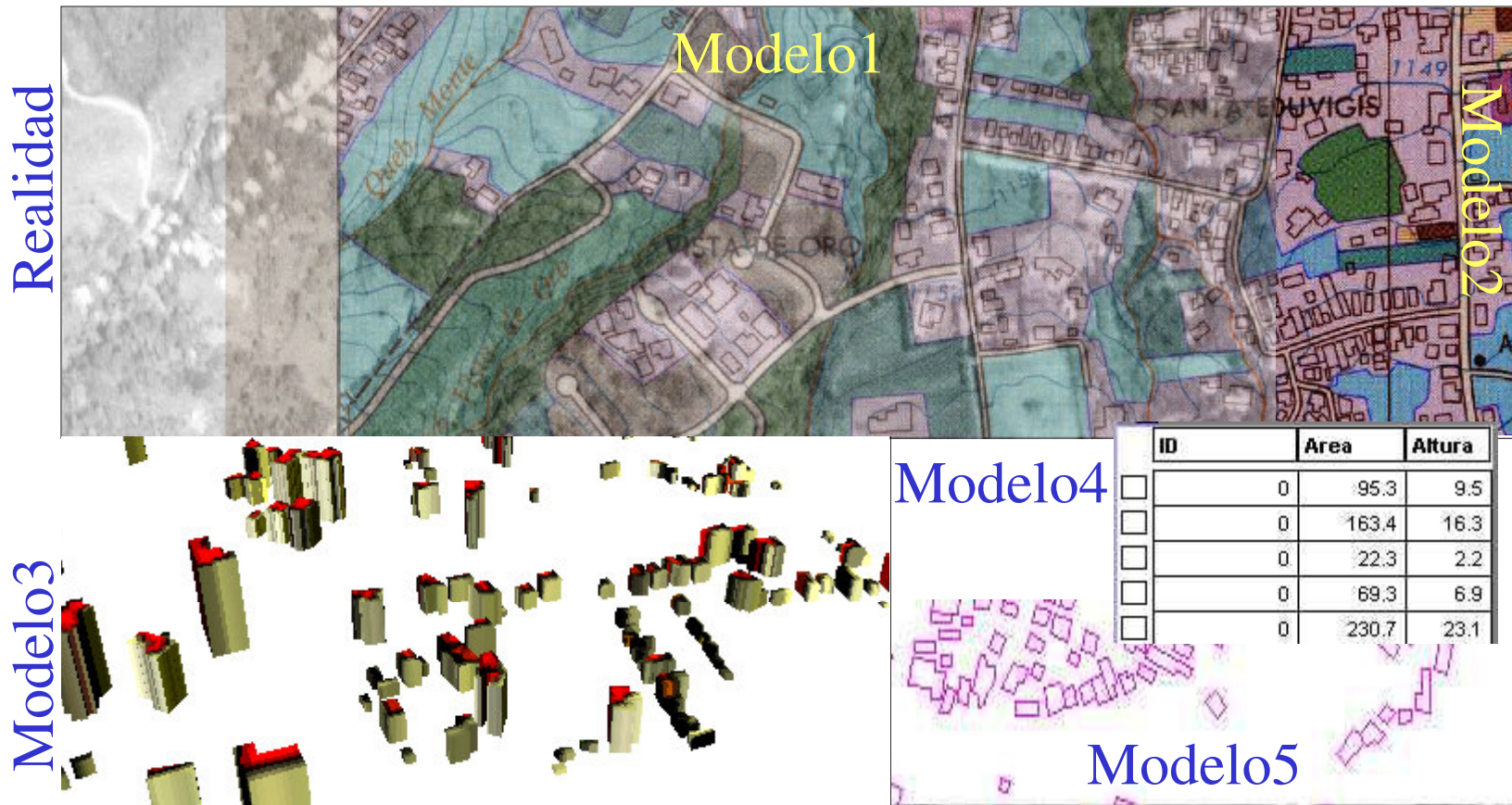
Un modelo se representa mediante un conjunto de símbolos

- Estos símbolos, su significado y las reglas de uso constituyen un lenguaje
- Un lenguaje formal voluntariamente restringido en función de un objetivo preciso de modelado se llama un formalismo.



Diversidad de objetivos = muchos modelos

- Una misma realidad puede ser representada simultáneamente por muchos modelos
- Cada modelo es construido con un objetivo preciso



Limitaciones ligadas al proceso de modelado

- La pérdida de detalles
- La dependencia con relación al objetivo perseguido
- La dependencia con relación al contexto
- Las limitaciones ligadas a las estimaciones:
 - 1. Identificación
 - 2. Medida
 - 2.1 Cualitativa
 - 2.2 Cuantitativa.

Limitaciones ligadas al modelador

- Experiencia
- Conocimientos
- Valores
- Personalidad
- Normas profesionales
- Otras

Homomorfismo de los modelos versus la realidad

- Debido a los marcos de referencia diferentes de dos personas, es casi imposible que dos personas diferentes creen dos modelos M_1 y M_2 perfectamente homomórficos entre ellos a partir de una misma realidad.
- Sin embargo, cuanto más pertinente sea la formación del modelador con relación a la aplicación modelada, más homomórfico será el modelo con respecto a la realidad.

Escalas de medida de las propiedades

- 1. Escala nominal.
 - Los valores son nombres en los cuales el orden no tiene incidencia sobre el valor de la propiedad.
 - Ejemplo: residencial, comercial, industrial, agrícola
- 2. Escala Ordinal.
 - Los valores son nombres clasificados según un orden riguroso que corresponde a una progresión del valor tomado por la propiedad.
 - Ejemplo: clasificación de los suelos agrícolas en pobres, medios, buenos, ricos.
- 3. Escala de intervalo.
 - Esta escala da un valor numérico utilizando una unidad de medida referida a un cero arbitrario.
 - Ejemplo: elevación "h" sobre el nivel medio del mar "nmm".
- 4. Escala de razón o proporción.
 - Esta escala, con relación a la precedente, agrega un cero significativo, no arbitrario, con lo cual todas las operaciones aritméticas (+,-,*,/) sobre los valores de las propiedades son posibles y el resultado es independiente de la unidad de medida utilizada.
 - Ejemplo: valor de una casa. Casa A = \$30,000.00; casa B = \$25,000.00.

Modelado: datos versus tratamientos

- Objetivos:
- Definir lo que es un dato.
- Definir lo que es un tratamiento.
- Comprender porque se modelan los datos
- Comprender lo que es un formalismo

Dato versus tratamiento

- Dato: contenido de una base de datos.
- Tratamiento: lo que se hace con el contenido de la base de datos.
- Tres tipos de tratamientos sobre los datos:
- 1. El ensamblaje
 - Los reagrupamientos de los datos individuales con información transpuesta
 - Ejemplo: el lote 121 del municipio X vale \$10,000.00.
- 2. La transformación
 - La utilización de un dato y de ciertos procedimientos para obtener una información derivada que no contiene al dato en bruto
 - Ejemplo: cálculo de la tasa de crecimiento de la población a partir de dos datos de población tomado en diferente año. $Tasa = (Pob_año2 - Pob_año1) / Pob_año1 * 100.$
- 3. Puesta en contexto
 - La comunicación de un dato suficientemente significativo para que el usuario deduzca de él una información que depende del contexto (una interpretación)
 - Ejemplo: Clase de edificio = 4. Donde 4 significa "3 o más pisos, paredes y divisiones de cemento, menos de 10 años".

¿Para qué modelar los datos?

- Para comprender el conjunto de los datos con los cuales el usuario desea trabajar.
- Un medio de comunicación entre los actores de un proyecto.
- Para documentar bien el sistema a fin de poderlo hacer evolucionar y actualizar más fácilmente.
- Para facilitar la programación.
- Sirve igualmente para reanalizar los datos con los cuales trabaja la organización.
- Los tipos de datos utilizados por la organización representan los elementos más estables (más que los tratamientos, los equipamientos y las personas).

Formalismo

- Un formalismo es un lenguaje
 - 1. Riguroso
 - 2. Restringido a algunas nociones.
 - 3. Semi-gráfico o exclusivamente literario.
- Un formalismo comprende:
 - 1. Componentes semánticas
 - 2. Notaciones
 - 3. Reglas de utilización

Formalismos gráficos versus formalismos literales

- Necesidades expresadas en formato literal
 - 1. Tengo datos sobre los hidrantes y las casas.
 - 2. Para los hidrantes necesito el número y la presión.
 - 3. Para las casas necesito su dirección y el número de pisos.
 - 4. Necesito saber si hay un hidrante cerca de una casa.
- Necesidades expresadas en formato gráfico

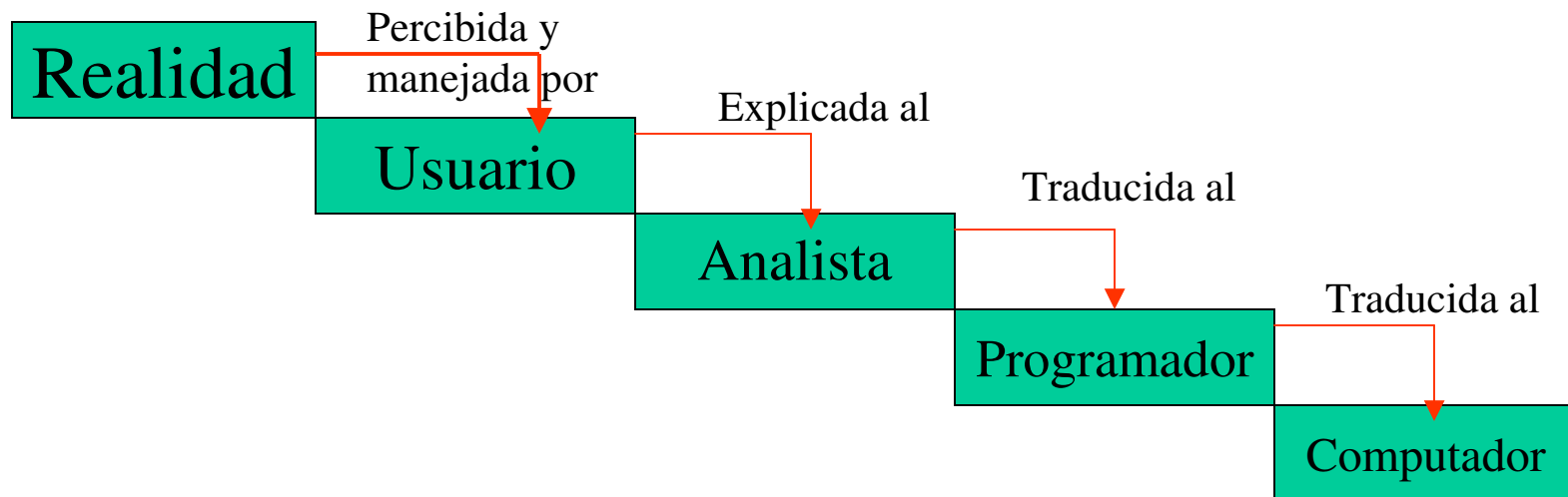


Los tipos de modelos de datos utilizados en los SIG

- Objetivos:
- Identificar los diferentes niveles de modelado
- Enunciar las características y los objetivos de los modelos utilizados a cada nivel
- Identificar los formalismos utilizables a cada nivel
- Establecer un paralelo entre los niveles del modelado de datos y del modelado de tratamientos

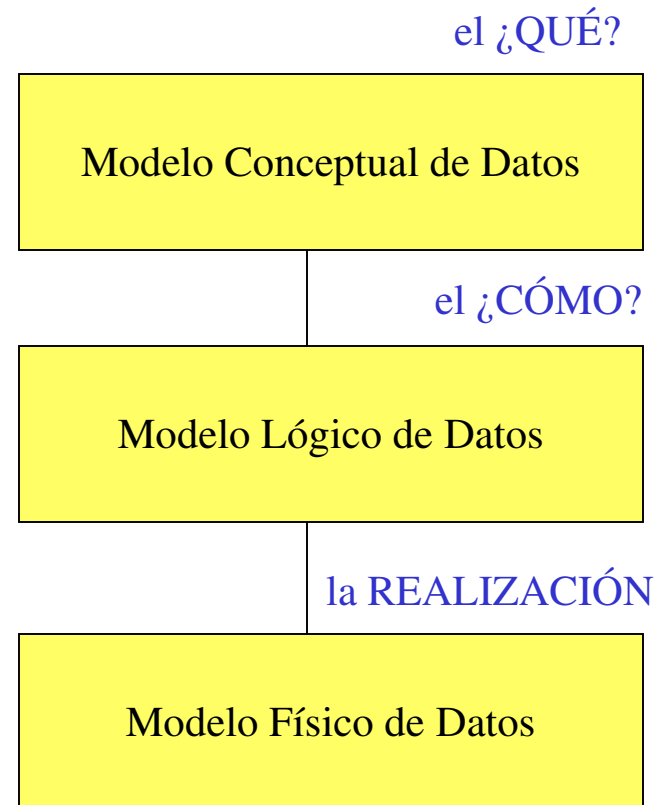
Diferentes necesidades de comunicación

- Durante el desarrollo de un SIG, la comunicación debe establecerse a varios niveles:
 - 1. Comunicación entre el diseñador o analista y los futuros usuarios del SIG
 - 2. Comunicación entre el analista y el programador
 - 3. Comunicación entre el programador y el sistema informático

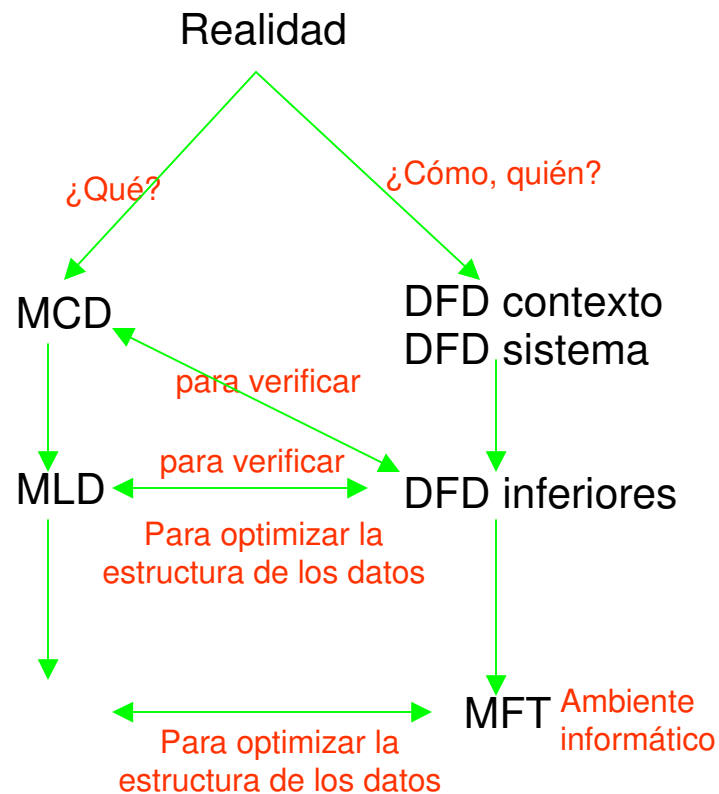


La jerarquía de los modelos de datos

- Nivel conceptual:
 - Responde a la pregunta del ¿Qué?
 - Es independiente de la tecnología.
 - Minimiza la redundancia de los datos.
 - Herramienta de trabajo del analista o diseñador.
- Nivel lógico:
 - Responde a la pregunta del ¿Cómo?
 - Es dependiente del tipo de tecnología escogida.
 - Es un modelo optimizado para obtener el máximo rendimiento.
 - Herramienta de trabajo para el programador.
- Nivel físico:
 - Consiste en la realización.
 - Depende de la herramienta escogida y de su lenguaje de programación
 - Herramienta de trabajo para el computador



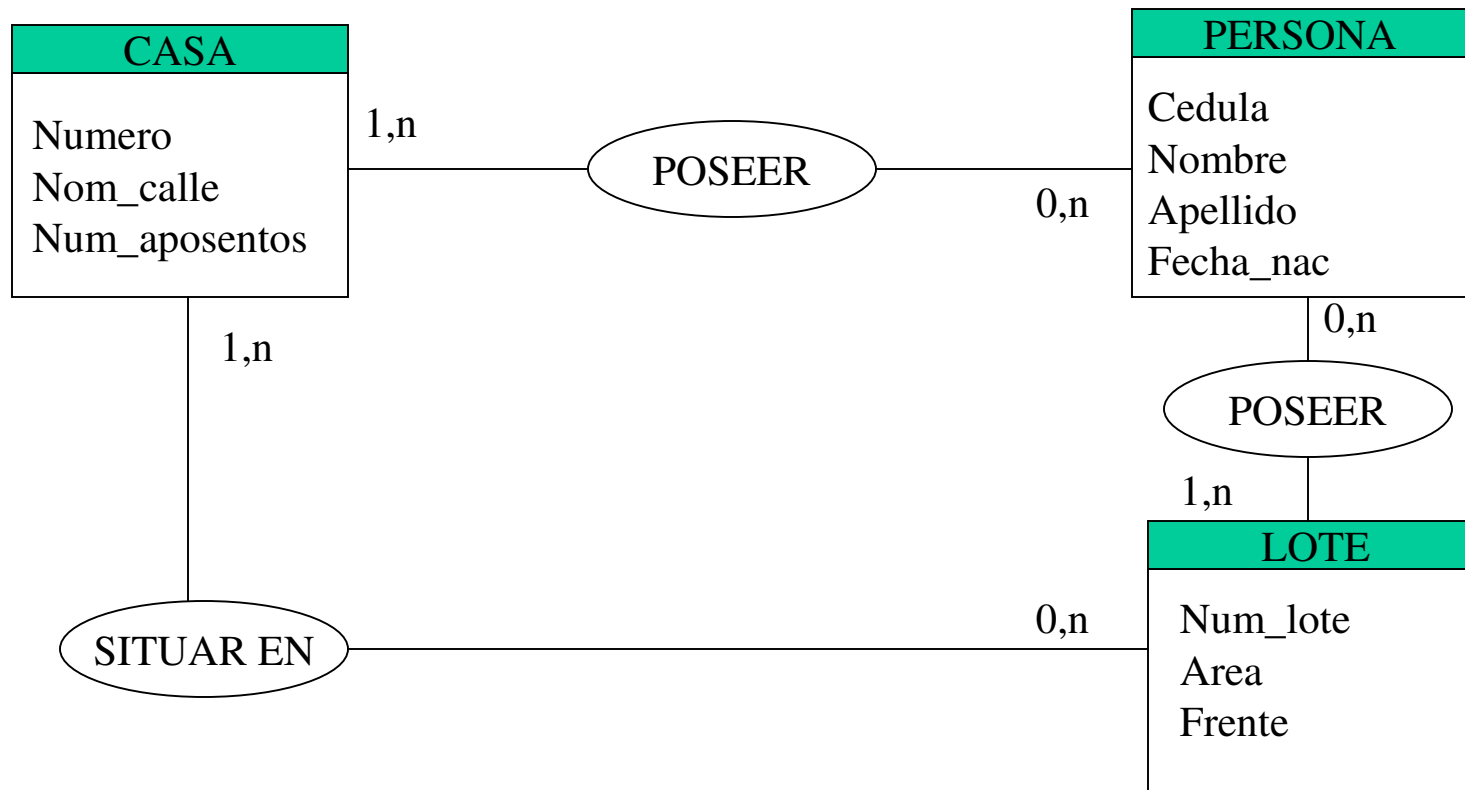
Ligamen entre el modelado de los datos y el modelado de los tratamientos



El Modelo Conceptual de Datos (MCD)

- El Modelo Conceptual de Datos (MCD):
- El MCD es una representación de la realidad en función de un propósito establecido, es independiente del tipo de tecnología que será utilizada durante la implantación del SIG y no está optimizado desde el punto de vista del rendimiento de los tratamientos.
- Este modelo es ante todo una herramienta de diálogo entre el diseñador o analista y los diversos actores de un proyecto.
- Se construye a partir de un lenguaje riguroso y muy restringido que resulta intuitivo y fácil de comprender.
- No incluye las limitaciones técnicas que harían pesada la lectura del modelo.

Ejemplo de un MCD (formalismo individual)



El Modelo Lógico de Datos (MLD)

- Es un modelo intermedio entre el MCD y el MFD. (A veces se produce más de un MLD).
- Es dependiente del tipo de "software" escogido (SGBD relacional, GIS orientado a los objetos, CAD ligado a un SGBD, ...) o del programa particular seleccionado.
- Permite efectuar una optimización de la estructura de datos elaborada en el MCD teniendo en cuenta los principales tratamientos, los puntos fuertes y las debilidades de un tipo de "software", así como la arquitectura informática de la organización.

Ejemplo de un MLD relacional

- CASA (Num_casa, Nom_calle, Num_aposentos)
- PERSONA_POSEER_CASA (Cédula, Num_casa, Nom_calle)
- PERSONA (Cédula, Apellido1, Apellido2, Nombre, Fecha_nacim)
- PERSONA_POSEER_LOTE (Cédula, Num_lote)
- LOTE (Num_lote, Superficie, Frente)
- CASA_SITUAR_SOBRE_LOTE (Num_casa, Nom_calle, Num_lote)

El Modelo Físico de Datos (MFD)

- El MFD es una herramienta esencialmente informática
- Representa, en el sistema informático, el contenido del MLD
- Es específico de la herramienta "software" escogida.

Ejemplo de un MFD sobre Informix (SGBD relacional)

- MAIN
- CREATE DATABASE model_1
- create table casa
- (num_casa Smallint ,
- nom_calle Char (20) ,
- num_aposentos Smallint ,
-)
- create UNIQUE INDEX i_num_casa ON casa (num_casa)
- create UNIQUE INDEX i_nom_calle ON casa (nom_calle)
- create table lote
- (
- num_lote Char (10) ,
- superficie Doble ,
- frente Doble ,
-)
- create UNIQUE INDEX i_num_lote ON lote(num_lote)
- create table casa_sobre_lote
- (
- num_casa Smallint ,
- nom_calle Char (20) ,
- num_lote Char (10) ,
-)
- END MAIN

El formalismo individual

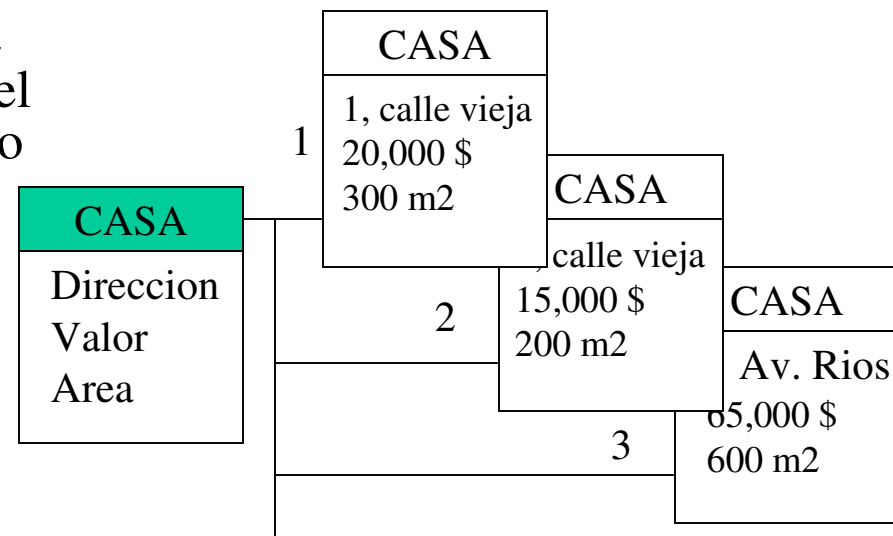
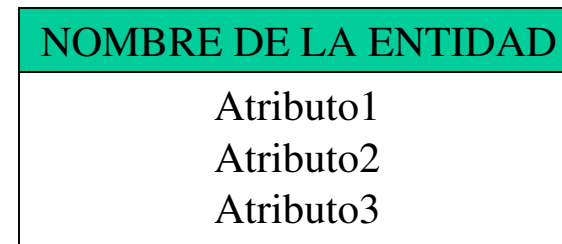
- Objetivos:
 1. Conocer las componentes del formalismo individual
 2. Saber manipular estas componentes
 3. Aprender a leer un modelo conceptual de datos utilizando este formalismo
- El formalismo individual es una variante del formalismo Entidad-Relación, variante desarrollada en Francia por los diseñadores del método MERISE. La familia de formalismos de tipo Entidad-Relación (E/R) es la más utilizada para modelar los datos a nivel conceptual.

Los componentes del formalismo individual

- Las entidades o individuos:
- Todo objeto, persona, concepto o evento de la realidad sobre el cual se desea información.
- Los atributos o propiedades:
- Las características de una entidad o de una relación que nos interesan en función del objetivo de nuestra aplicación. Existe un tipo particular de atributo: el identificador.
- Las relaciones:
- Asociaciones entre las entidades.
- Las cardinalidades:
- Los número máximo y mínimo de veces que una relación puede producirse entre las ocurrencias de dos entidades.

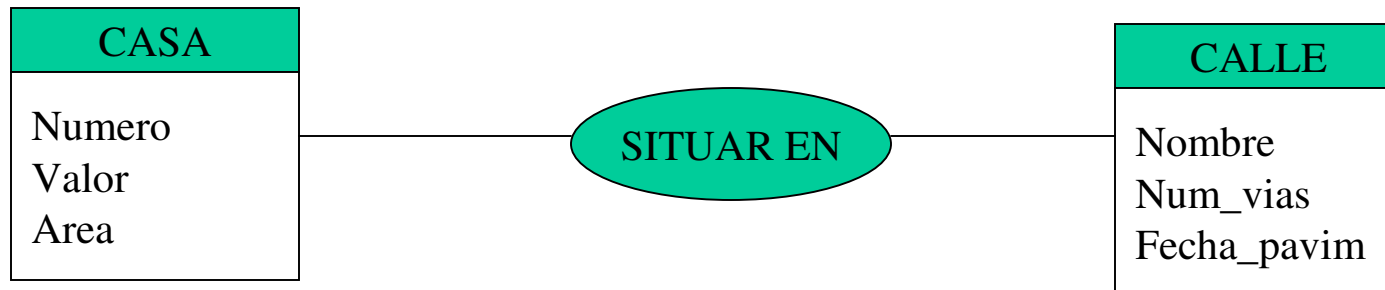
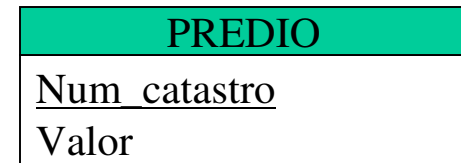
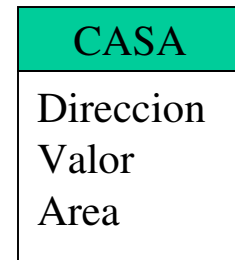
Las notaciones del formalismo individual 1

- Entidad ó individuo
- Un rectángulo que contiene el nombre único de la entidad en letras mayúsculas.
- Distinción importante entre *entidad* y *ocurrencia*
- Una entidad está compuesta de muchos atributos y representa el conjunto de ocurrencias del mismo tipo.
- Una ocurrencia es una entidad específica que posee un conjunto de valores que describen a esta entidad en particular.



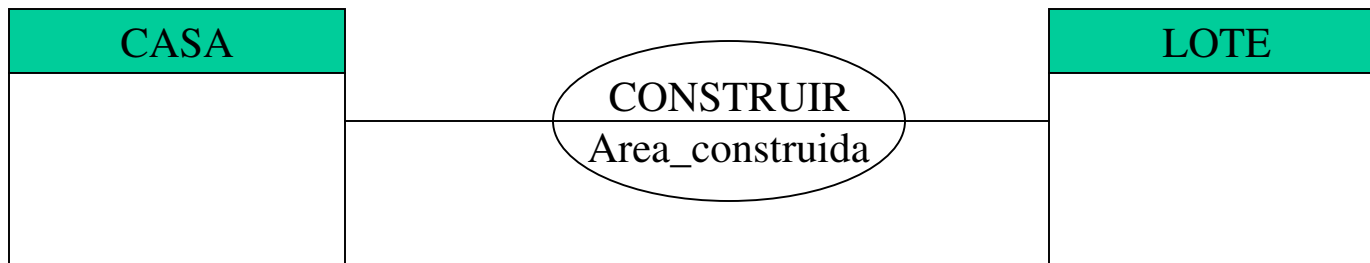
Las notaciones del formalismo individual 2

- Atributo o propiedad:
- Los atributos sirven para describir o para localizar una entidad. Uno representa un atributo donándole un nombre único en el seno de la entidad que él caracteriza.
- El *identificador* es un tipo de atributo particular el cual permite identificar de forma única (o casi única) toda ocurrencia de una entidad particular.
- Cuando la identificación no es única sino que depende del identificador de otra entidad, se habla entonces de *dependencia funcional*.



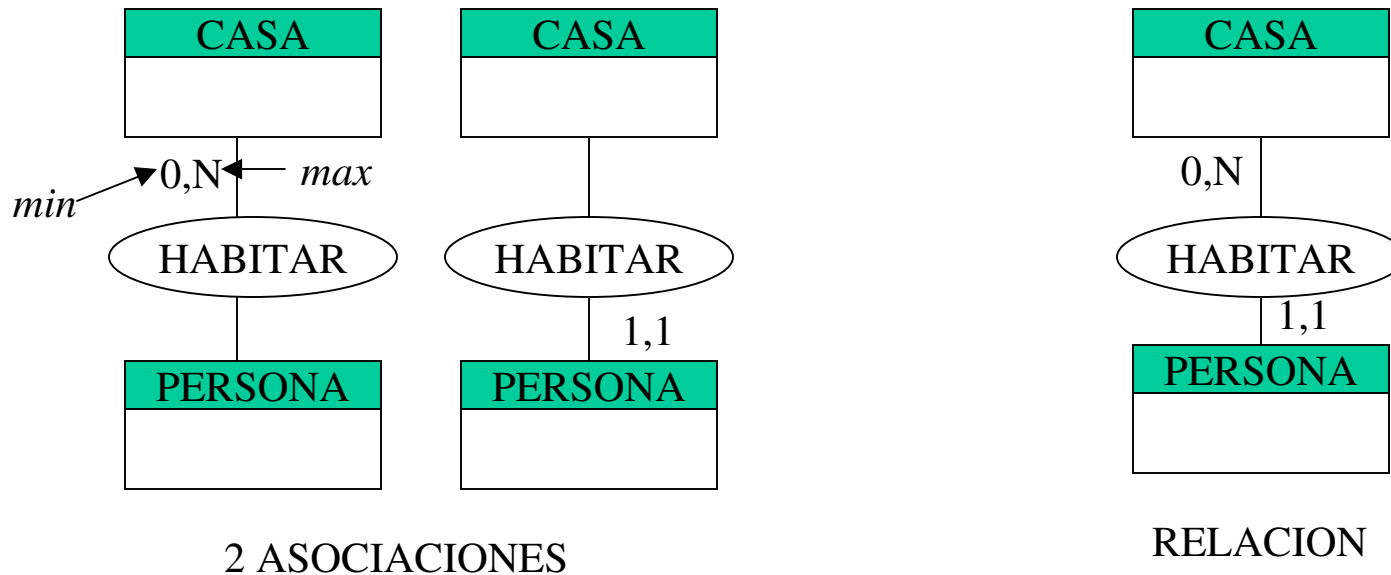
Las notaciones del formalismo individual 3

- Relación:
- Una relación se representa mediante una elipse en la cual se sitúa el nombre de la relación en mayúscula y sus atributos en minúscula si fuera del caso.



El formalismo individual (cardinalidad)

- Cardinalidad
- Una relación posee cardinalidades. Para comprender bien la noción de cardinalidad, uno puede descomponer la relación en dos asociaciones, una en cada dirección de la relación. Cada asociación posee una cardinalidad, i.e. el número mínimo y máximo de veces que la asociación puede producirse entre dos entidades.



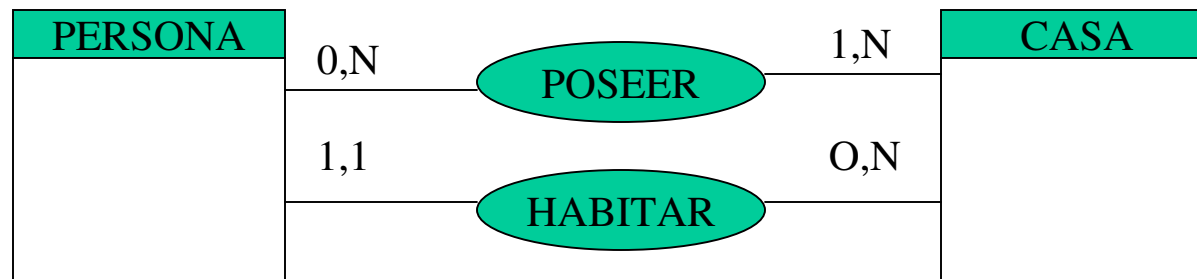
La cardinalidad total de una relación se obtiene a partir de los máximos de las dos asociaciones. En el ejemplo precedente, la cardinalidad total de la relación **HABITAR** es $1,N$. Es esta cardinalidad total la que servirá para construir el MLD.

El formalismo individual (tipos de asociaciones y cardinalidad total)

- 1. Asociación simple: sólo una pero es obligatoria (1,1)
 - 2. Asociación condicional: sólo una, pero es facultativa (0,1)
 - 3. Asociación múltiple: muchas posibles, pero al menos una es obligatoria (1,N)
 - 4. Asociación múltiple condicional: muchas posibles, pero podría que halla ninguna (0,N)
 - 5. Asociación especial: i.e. con un número mínimo específico y un número máximo específico (diferente de 1) (2,4).
-
- La cardinalidad total de una relación puede entonces ser (1,1), (1,N), (N,N) ó especial (ej.: 4,N).

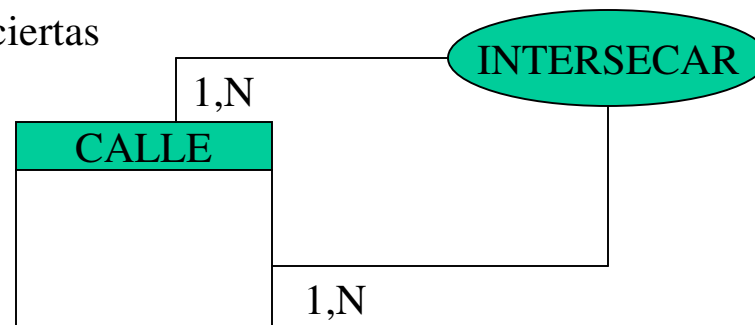
El formalismo individual (número de relaciones)

- Relaciones:
- Pueden existir dos o aún muchas relaciones de diferente naturaleza entre dos entidades.
- Ej.: las relaciones entre PERSONA y CASA.



El formalismo individual acepta ciertas relaciones particulares.

Ej. Relaciones recursivas.

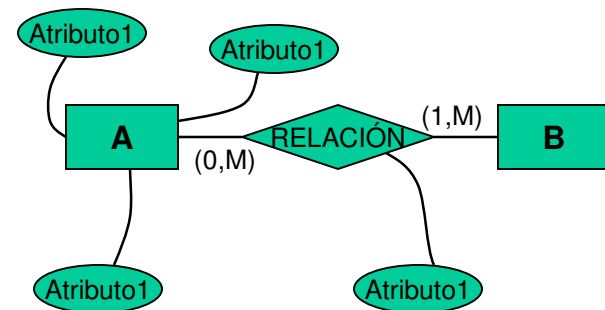


El formalismo entidad – relación de Chen

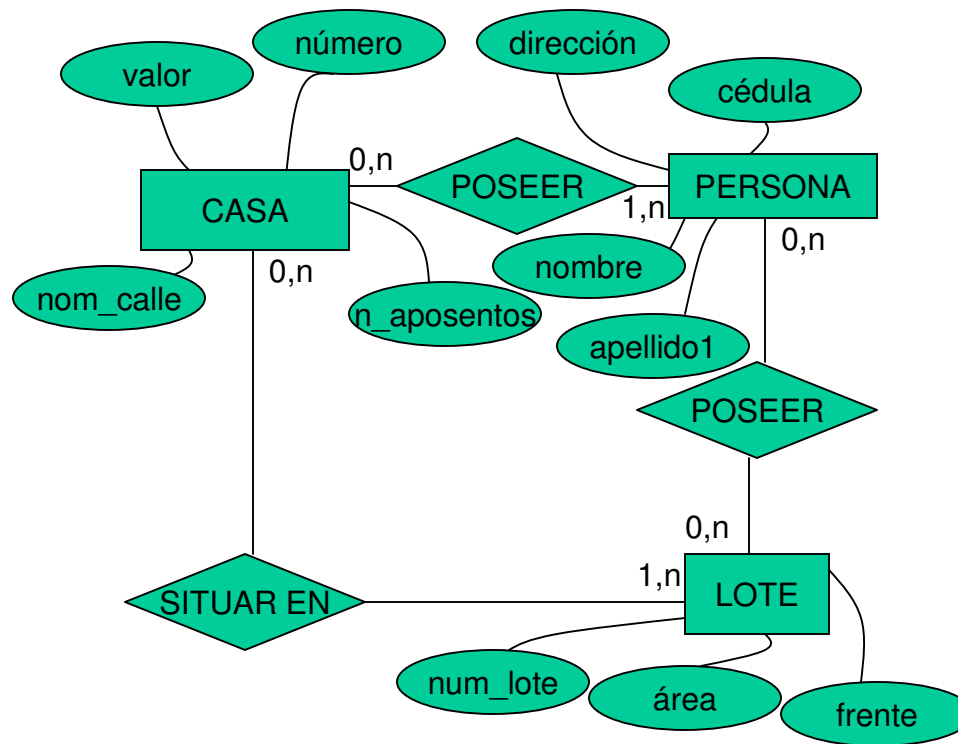
- Existen diferentes formalismos de tipo E/R, uno de entre ellos es el formalismo individual. Otro formalismo muy utilizado es el de Chen.

Las componentes del formalismo individual de Chen

- La notación gráfica de sus componentes:
- Las entidades:
 - Un rectángulo dentro del cual se coloca únicamente el nombre en letras mayúsculas.
- Los atributos:
 - Una elipse que contiene el nombre del atributo en minúsculas la cual está ligada por una curva o una recta al rectángulo que corresponde a la entidad caracterizada por tal atributo.
- Las relaciones:
 - Un romboide conteniendo el nombre de la relación en mayúsculas.
- Las cardinalidades:
 - Las cardinalidades se leen al inverso que en el formalismo individual, i.e., la primera entidad, la relación, la cardinalidad y la segunda entidad.



Ejemplo de un MCD, formalismo ER de Chen



Normalización del MCD 1

- Objetivos:
 - 1. Aprender la necesidad de normalización de un MCD
 - 2. Conocer las cuatro formas normales
 - 3. Conocer los peligros de la sobrenormalización
 - 4. Aplicar estas reglas a un MCD

Primera forma normal

- Con el propósito de eliminar la redundancia de los datos, el MCD debe respetar cuatro formas normales.
- Carácter elemental de las propiedades o atributos:
- Todos los atributos deben ser elementales (significativos e indivisibles) con relación a los deseos del cliente.
- Ejemplo, para un edificio:

EDIFICIO
<u>Numero</u>
<u>Nombre calle</u>

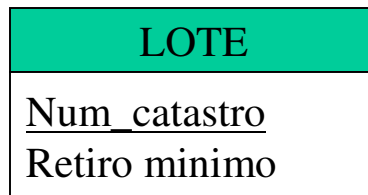
versus

EDIFICIO
Direccion

Segunda forma normal

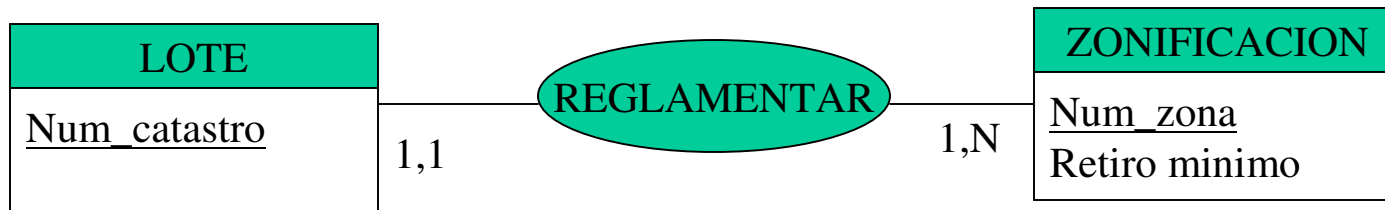
- Dependencia plena del identificador:
- No deben haber valores de atributos que no dependan de la ocurrencia de la entidad a la cual ellos son atribuidos, i.e. que sean automáticamente los mismos para algunas ocurrencias diferentes.

Ejemplo:



El atributo "margen de retiro mínimo" no depende del lote, sino de una zona definida por el reglamento de zonificación municipal.

Solucion:



Tercera forma normal

- Dependencia transitiva.
- Si un atributo C depende del atributo A por medio del atributo B, este no está conforme con la tercera forma normal.

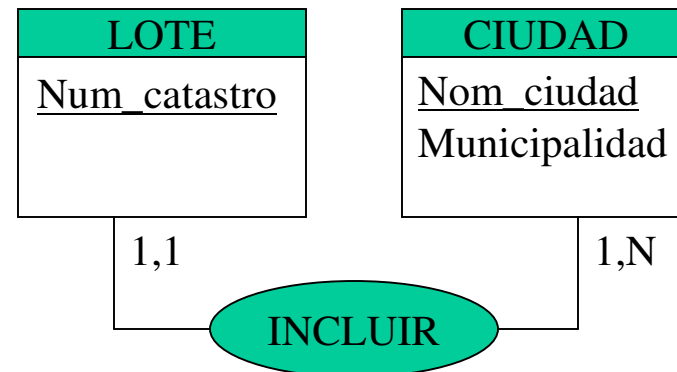
Ejemplo:

LOTE
<u>Num_catastro</u> (Atributo A)
Nombre_ciudad (Atributo B)
Municipalidad (Atributo C)

El condado (atributo C) es siempre el mismo para una ciudad específica (atributo B)

Solución:

Crear una entidad CIUDAD donde el atributo *Condado* estaría incluido. Crear una relación entre CIUDAD y LOTE.



Cuarta forma normal

- Forma normal de Boyce-Codd:
- Si, para un identificador compuesto de muchos atributos, uno de estos atributos depende directamente de una propiedad que no es el identificador, la forma normal de Boyce-Codd no es respetada.

Ejemplo:

EDIFICIO
<u>Codigo_municipalidad</u>
Direccion
Nombre de la ciudad

CIUDAD
<u>Codigo_municipalidad</u>
Nombre de la ciudad

El Código de la municipalidad es siempre el mismo para una ciudad específica.

Solucion:

Crear una entidad CIUDAD donde el atributo *Código de municipalidad* estaría incluido.
Crear una relación entre CIUDAD y UNIDAD DE EVALUACIÓN lo que agrega una dependencia funcional.

Cuidado con la sobrenormalización

- Las reglas de normalización tienen un carácter obligatorio más o menos marcado, i.e. que en la práctica, uno deja generalmente de aplicarlas cuando ellas hacen la lectura de un modelo menos intuitiva.
- "Es asunto del analista percibir hasta donde conviene normalizar el modelo (...). Un modelo muy normalizado permitirá iniciar la validación sobre una base clara; el esquema más descompuesto será menos legible." Collonge et al., 1986. Merise, méthode de conception.

El Modelo Lógico de Datos (formalismo relacional)

- Objetivos:
- Comprender el propósito de la modelado lógico
- Conocer las componentes del formalismo relacional
- Comprender el paso de un MCD formalismo individual a un MLD relacional
- Comprender como traducir un MCD formalismo MODUL-R al nivel lógico
- Adquirir algunas nociones de optimización

El formalismo relacional de nivel lógico

- Sus componentes:
- Entidad ó tabla
- Nombre en mayúsculas, a veces con recuadro, situado por encima y a la izquierda del rectángulo que contiene los atributos
- Atributos ó columnas
- Nombre en minúsculas, todos los atributos en recuadro y adyacentes horizontalmente; los identificadores se convierten en las llaves de acceso y se subrayan.
- Relación ó tabla
- Similar a las entidades.

TABLA				
<u>Llave de acceso</u>	Atributo1	Atributo2	Atributo3	Atributo4

Existe igualmente una versión no gráfica del formalismo:

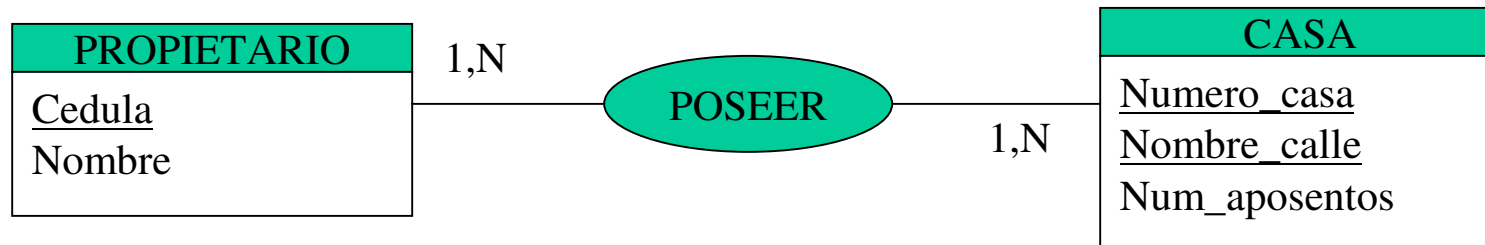
TABLA (Llave de acceso, Atributo1, Atributo2, Atributo3, Atributo4,...)

Pasaje del MCD al MLD relacional (sin optimización)

MCD	MLD
Entidad	Tabla
Identificador	Llave de acceso
Atributo	Columna
Relación:	
N á N	Una nueva tabla que tiene como llave de acceso los identificadores de las entidades que la relación liga.
1 á N	La llave de acceso de la tabla padre es agregada a la tabla hija.
1 á 1	Dos opciones se ofrecen: Migrar los atributos de una hacia la otra ó, copiar la llave de acceso de una en la otra.

Ejemplos de traducción 1

Relacion N a N



PROPIETARIO (Cedula, nombre)

PROPIETARIO_POSEER_CASA (Cedula, numero_casa, nombre_calle)

CASA (Numero_casa, nombre_calle, num_aposentos)

Relacion 1 a N

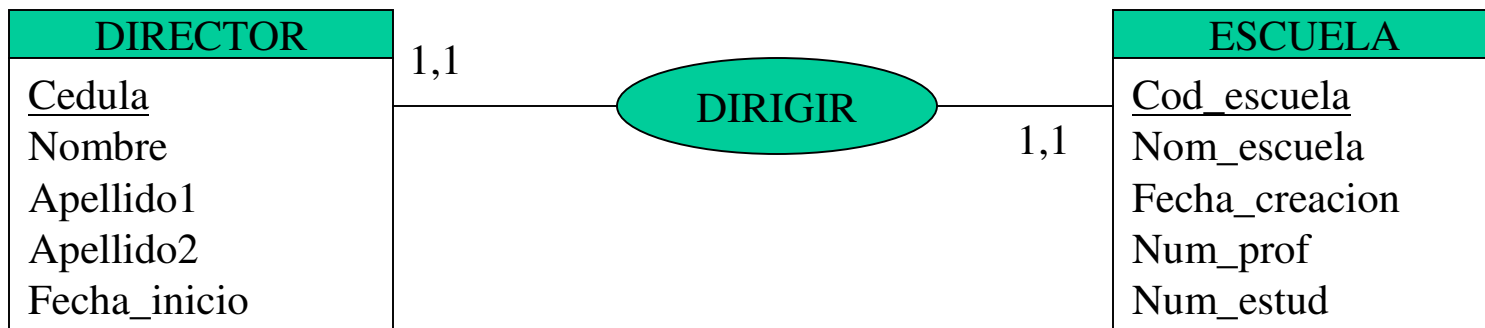


PERSONA (Cedula, numero_casa, nombre_calle, nombre)

CASA (Numero_casa, nombre_calle, num_aposentos)

Ejemplos de traducción 2

Relacion 1 a 1



Solucion 1:

DIRECTOR-DIRIGIR-ESCUELA (Cedula, cod_escuela, nombre, apellido1, apellido2, fecha_inicio, nom_escuela, fecha_creacion, num_prof, num_estud)

Solucion 2:

DIRECTOR (cedula, cod_escuela, nombre, apellido1, apellido2, fecha_inicio)

ESCUELA (cod_escuela, cedula, nom_escuela, fecha_creacion, num_prof, num_estud)

Optimización de un MLD 1

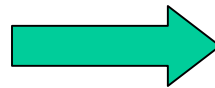
- Una vez que se ha realizado el MLD, nosotros podemos modificar su estructura a fin de mejorar el rendimiento de ciertos tratamientos muy importantes. Para hacerlo, se reintroduce generalmente (de forma controlada) alguna redundancia de los datos. Se habla entonces de "denormalización".
- Ejemplo 1: A menudo se necesita el valor total de una propiedad, por lo general este valor es calculado a partir del valor de todos los terrenos y todos los edificios que constituyen cada propiedad. Con el fin de acelerar los tratamientos que hacen referencia al valor total, uno puede decidir agregar un campo "valor total", contraviniendo, a propósito, las formas normales.

Optimización de un MLD 2

ELEMENTO VIAL
Atributo#1
Atributo#1
Atributo#1
Atributo#1
Atributo#1

INTERSECCION
Atributo#1
Atributo#1
Atributo#1

Ejemplo 2



TRAMO VIAL
Atributo#1
Atributo#1
Atributo#1
Atributo#1
Atributo#1

Donde los atributos #4 y #5 estaran vacios para las intersecciones de calle.

En este caso no se agrega redundancia, sino campos vacíos para los casos donde se maneja una INTERSECCIÓN:

- Esto reserva espacio inútilmente, pero
- Permite crear una sola tabla en vez de dos

En general, es más rápido trabajar sobre una sola tabla que explotar las relaciones entre dos tablas, de ahí el interés de hacer este tipo de compromiso espacio/rendimiento.

N.B.: Estas operaciones dependen mucho del "software" utilizado.