



**DIRECCIÓN DE REGULACIÓN, PLANEACIÓN,
ESTANDARIZACIÓN Y NORMALIZACIÓN
(DIRPEN)**

**GUÍA CON RECOMENDACIONES PARA
ELABORAR MODELOS ENTIDAD - RELACIÓN**

COORDINACIÓN DE REGULACIÓN ESTADÍSTICA

Diciembre de 2024

Contenido

| | |
|---|-----------|
| Introducción | 3 |
| 1. Objetivo y alcance | 4 |
| 1.1. Objetivo | 4 |
| 1.2. Alcance..... | 4 |
| 2. Conceptos básicos | 5 |
| 3. Importancia del modelo entidad-relación | 7 |
| 4. Aspectos fundamentales | 9 |
| 4.1. Partes de un modelo ER..... | 9 |
| 5. Buenas prácticas y recomendaciones para la elaboración del modelo ER | 23 |
| 6. Actividades para la elaboración del modelo ER | 30 |
| 6.1. Socialización del diseño de la operación estadística | 30 |
| 6.2. Determinar las entidades | 30 |
| 6.3. Identificar las relaciones | 34 |
| 6.4. Establecer los atributos asociados a las entidades o las relaciones..... | 36 |
| 6.5. Determinar los dominios de los atributos..... | 38 |
| 6.6. Determinar los identificadores..... | 40 |
| 6.7. Determinar jerarquías | 42 |
| 6.8. Dibujar el diagrama ER..... | 43 |
| 7. Revisar el esquema conceptual con el equipo temático de la operación estadística | 43 |
| 8. Consolidar documento..... | 45 |
| Bibliografía | 46 |

Introducción

El Departamento Administrativo Nacional de Estadística (DANE) es la entidad rectora del Sistema Estadístico Nacional (SEN) y la autoridad nacional de regulación estadística (Ley 2355 de 2023). Como tal, establece la regulación en materia estadística a través de un conjunto de directrices, lineamientos, estándares y buenas prácticas aplicables al proceso estadístico en todas las entidades productoras de estadísticas oficiales.

Siguiendo lo establecido en los Lineamientos para el Proceso Estadístico del Sistema Estadístico Nacional Versión 2.0 (DANE, 2020), adaptado del estándar *Generic Statistical Business Process Model* (GSBPM) en su versión 5.1, y considerando la importancia para la gestión de la calidad en las operaciones estadísticas (especialmente en relación con la NTC PE 1000), el DANE presenta este documento con recomendaciones para la elaboración de un modelo entidad-relación.

El diseño del modelo entidad-relación corresponde a la fase de diseño dentro del ciclo de vida del desarrollo de software. Este documento brinda los elementos conceptuales necesarios, describe las actividades a realizar y presenta las buenas prácticas que se recomiendan implementar para diseñar y elaborar de forma rigurosa un modelo entidad-relación.

El documento se compone de siete partes: en la primera parte, se presentan el objetivo y el alcance del documento; la segunda parte define los conceptos que se utilizan en el diseño y la elaboración del modelo; la tercera parte explica la importancia de construir el modelo dentro de la implementación del proceso estadístico; la cuarta parte expone los aspectos fundamentales para su elaboración; la quinta parte describe las buenas prácticas y las recomendaciones para su elaboración; la sexta parte contiene una propuesta de actividades para su elaboración, y en la séptima parte se muestra la bibliografía utilizada en la formulación del presente documento.

1. Objetivo y alcance

1.1. Objetivo

Proporcionar directrices claras y estandarizadas para la elaboración de modelos entidad-relación, asegurando la calidad y la consistencia en la fase de diseño para las bases de datos relacionales utilizadas en las operaciones estadísticas por las entidades del SEN.

1.2. Alcance

Este documento se aplica a las operaciones estadísticas generadas por las entidades del SEN y abarca la creación de modelos entidad-relación para bases de datos relacionales, proporcionando una guía estructurada que incluye conceptos clave, buenas prácticas y recomendaciones para asegurar un diseño de alta calidad, para los diseñadores, desarrolladores, administradores y usuarios de las bases de datos.

2. Conceptos básicos

Base de datos: conjunto o colección de datos interrelacionados entre sí, que se utilizan para la obtención de información de acuerdo con el contexto de estos y que son almacenados sistemáticamente para su posterior uso ((DANE), 2021).

Diagrama: representación gráfica, generalmente esquemática, para describir o comunicar algo ((DNP), 2019).

Diseño (Proceso estadístico): fase del proceso estadístico en la que se definen y documentan los aspectos metodológicos y los procedimientos para la construcción, la recolección o el acopio, el procesamiento, el análisis, la difusión y la evaluación ((DANE), Departamento Administrativo Nacional de Estadística, 2024).

Diseño del procesamiento: en este subproceso se establece, a partir de criterios temáticos y estadísticos, las especificaciones para la consolidación e integración de los archivos o bases de datos, el tratamiento, la edición e imputación de datos, con el propósito de disponer una base de datos depurada para la producción de resultados. Aplica para todas las operaciones estadísticas. El responsable del proceso estadístico establece los mecanismos para procesar los datos recolectados o copiados, entre ellos, determina las herramientas tecnológicas para el procesamiento (software y hardware) (DANE, 2020)

Documento metodológico: documento que presenta a los usuarios de manera clara y precisa los métodos y actividades desarrolladas en el proceso estadístico ((DANE), Departamento Administrativo Nacional de Estadística, 2024).

Metadatos: información necesaria para el uso e interpretación de las estadísticas. Los metadatos describen la conceptualización, calidad, generación, cálculo y características de un conjunto de datos estadísticos (DANE, 2023).

Modelo: esquema teórico y formal de un sistema o de una realidad compleja que se elabora para facilitar su comprensión y el estudio de su comportamiento ((DNP), 2019).

Operación estadística: es la aplicación del conjunto de procesos y actividades que comprende la identificación de necesidades, diseño, construcción, recolección o acopio, procesamiento, análisis,

difusión y evaluación, la cual conduce a la producción de información estadística sobre un tema de interés nacional o territorial (DANE, 2023).

Procesamiento: fase del proceso estadístico en la que se consolidan, integran, procesan y depuran los datos, de acuerdo con lo establecido en el diseño ((DANE), Departamento Administrativo Nacional de Estadística, 2024).

Proceso estadístico: es el conjunto sistemático de actividades encaminadas a la producción de estadísticas, entre las cuales están comprendidas: la detección de necesidades de información, el diseño, la construcción, la recolección, el procesamiento, el análisis, la difusión y la evaluación (DANE, 2023)

Sistema Estadístico Nacional (SEN): es el conjunto articulado de componentes que garantizan la producción y difusión de las estadísticas oficiales a nivel nacional y territorial que requiere el país, de manera organizada y sistemática. Sus componentes son las entidades y organizaciones productoras de información estadística y responsables de registros administrativos, las personas y entes usuarias, los procesos e instrumentos técnicos para la coordinación, así como las políticas, principios, fuentes de información, infraestructura tecnológica y talento humano necesarios para su funcionamiento (DANE, 2023).

3. Importancia del modelo entidad-relación

En el contexto del proceso estadístico¹ el diseño y la elaboración del modelo entidad-relación (ER) es un componente crucial que permite representar de manera gráfica y comprensible la estructura de la base de datos utilizada en la operación estadística. Este modelo proporciona claridad visual sobre cómo se organizan los datos y facilita la comunicación entre los diferentes roles involucrados en el desarrollo de la base de datos, tales como diseñadores, desarrolladores, administradores de sistemas y usuarios finales.

El modelo ER es una herramienta visual que representa las entidades y sus relaciones dentro de un sistema de información. A través de un diagrama ER, se puede visualizar de manera clara y precisa la estructura de la base de datos y así asegurar que todos los participantes en el proyecto tengan una comprensión común de los requisitos del sistema; esto es fundamental para evitar malentendidos y certificar que el diseño de la base de datos satisfaga las necesidades de todos los usuarios involucrados.

Un aspecto crítico del modelo ER es su capacidad para organizar los datos de manera coherente, lógica y estructurada. Define los datos que deben ser almacenados y gestionados por la base de datos, facilitando la identificación y el entendimiento de los requisitos del sistema. Esta organización estructurada es esencial para la gestión eficiente de los datos y reafirmar que se almacenen de manera coherente y se puedan recuperar de forma efectiva cuando se necesiten.

Contar con un diseño estándar del modelo ER permite aprovechar conocimientos previos y seguir una forma uniforme de elaborar bases de datos, generando economías de escala en el diseño de operaciones estadísticas, reduciendo costos y mejorando la eficiencia. La estandarización asegura que los modelos sean consistentes y reutilizables, facilitando su integración y mantenimiento a largo plazo.

El modelo ER también juega un papel vital en la identificación de posibles redundancias, incoherencias o problemas en el diseño de la base de datos. Al proporcionar una representación detallada de la estructura de la base de datos, permite realizar ajustes y mejoras antes de que se complete el desarrollo final, lo que contribuye a la creación de bases de datos eficientes y bien diseñadas y minimizar los problemas y los costos asociados con un diseño deficiente.

¹ Lineamientos para el proceso estadístico en el Sistema Estadístico Nacional, versión 2.0. DANE (2020) Este modelo cuenta con ocho fases: detección y análisis de necesidades, diseño, construcción, recolección/acopio, procesamiento, análisis, difusión y evaluación.

En cada operación estadística, se gestionan datos recolectados o acopiados, procesados, analizados y finalmente difundidos². Por lo tanto, en la fase de diseño es crucial generar modelos conceptuales de bases de datos relacionales que puedan sintetizar la realidad del manejo de la información en cada operación estadística. Estos modelos orientan el trabajo de los desarrolladores de aplicaciones y bases de datos, asegurando que se sigan los parámetros y las directrices establecidas por la entidad.

² Gestión de datos: es la actividad que debe asegurar, mantener y proveer instituciones de datos, unificando datos maestros y regularizando registros en los sistemas fuente. Esta actividad involucra la identificación de los requerimientos que mantienen repositorios centrales del dato, los que determinan la asociación con procesos claves que usan el dato y los que definen los tipos de aprovisionamiento de datos a gestionar (reactivo, proactivo, administrado, optimizado y autoservicio) (Mintic., 2014) (DAFP, 2021).

4. Aspectos fundamentales

4.1. Partes de un modelo ER

Para garantizar la uniformidad en la elaboración de los modelos ER en la fase de diseño, del proceso estadístico, se sugieren las siguientes partes mínimas y conceptos básicos del modelo.

Entidad y conjunto de entidades

Se define como entidad cualquier objeto o concepto (real o abstracto) que existe en la realidad y sobre el que se quiere almacenar información en una base de datos (De Miguel, Piattini y Marcos; AIU).

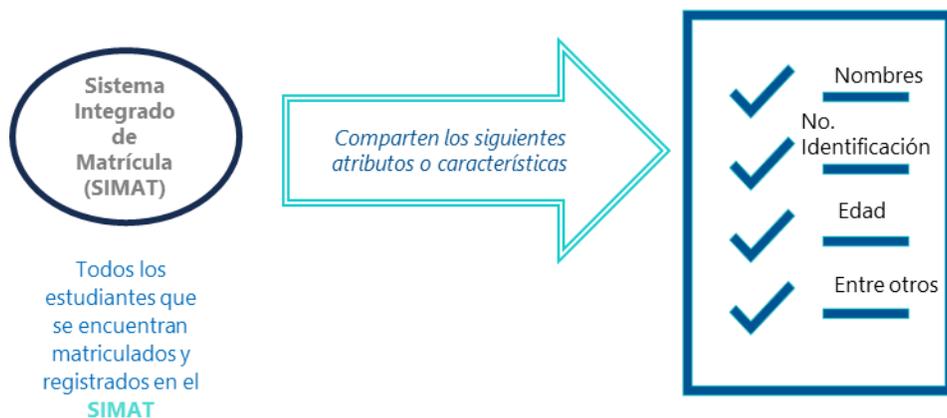
Por ejemplo:



Adicionalmente, se considera que un conjunto de entidades es una colección de entidades que comparten los mismos atributos y características.

Por ejemplo:

Relación y conjunto de relaciones

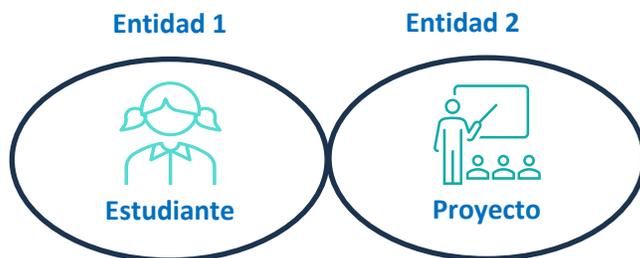


Una relación (o interrelación) es entendida como una asociación, vinculación o correspondencia entre dos o más entidades (De Miguel, Piattini y Marcos; AIU). Cada relación tiene un nombre que describe su función. La dependencia o la asociación entre los conjuntos de entidades es llamada participación. Igualmente, se denomina tipo de relación a una estructura genérica que describe un conjunto de relaciones, mientras que una relación será cada uno de los ejemplares concretos, por lo que el tipo de relación es el resultado de clasificar un conjunto de relaciones (De Miguel, Piattini y Marcos).

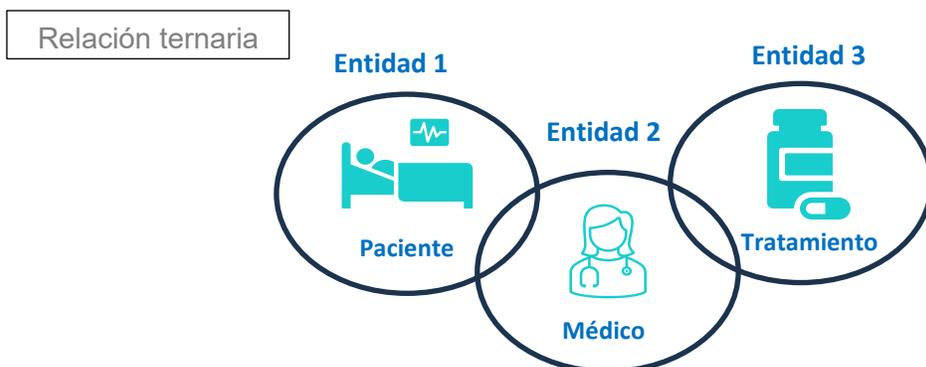
Las entidades que están involucradas en una determinada relación se denominan entidades participantes. El número de participantes en una relación es lo que se denomina grado de la relación. Por ejemplo, una relación en la que participan dos entidades es una relación binaria y si participan tres la relación es ternaria.

Ejemplos

Relación binaria



Un estudiante puede estar inscrito en varios proyectos estudiantiles, y un proyecto estudiantil puede tener varios estudiantes inscritos.



Un paciente puede recibir varios tratamientos, cada tratamiento puede ser administrado por varios médicos, y un médico puede tratar a varios pacientes.

Un conjunto de relaciones es una colección de relaciones de la misma naturaleza. Además, el grado del conjunto de relaciones es la cantidad de conjunto de entidades participantes de la relación.

Para el ejemplo sobre educación, en una relación de la «Secretaría de Educación de Entidades Territoriales Certificadas (SED_ETC)» y «Estudiante» indicaría que:



Atributo

Se considera que un atributo es una característica de interés o un hecho sobre una entidad o una relación. Un atributo representa propiedades básicas o características de entidades y relaciones (De Miguel, Piattini y Marcos; AIU).

Para cada atributo se establece un conjunto de posibles valores llamado dominio. Este define todos los valores posibles que puede tomar un atributo y puede haber varios atributos definidos sobre un mismo dominio (De Miguel, Piattini y Marcos; AIU).

Los atributos pueden ser simples o compuestos. Un atributo simple es un atributo que tiene un solo componente, que no se puede dividir en partes más pequeñas que tengan un significado propio. Un atributo compuesto es un atributo con varios componentes, cada uno con un significado por sí mismo.

Ejemplo:



Los atributos también pueden clasificarse en monovalentes o polivalentes. Un atributo monovalente es aquel que tiene un solo valor para cada ocurrencia de la entidad o la relación a la que pertenece. Un atributo polivalente es aquel que tiene varios valores para cada ocurrencia de la entidad o la relación a la que pertenece y pueden tener un número máximo y un número mínimo de valores (AIU).

Ejemplos:



Identificación

Número que identifica a un estudiante



Matrícula

Número de registro de un estudiante en el SIMAT

Atributos Monovalentes:

Los atributos "identificación" y "matrícula solo puede tener un valor único para cada estudiante.



Entidad Estudiante



Teléfono:

3102222222
3051111111



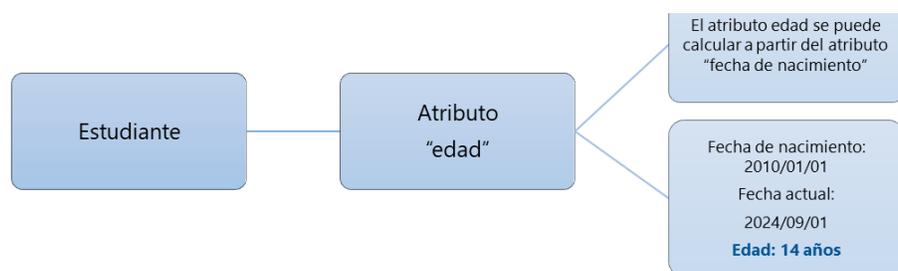
Correo electrónico:

estudiante@ejemplo.co
m.
estudiante2@ejemplo.com

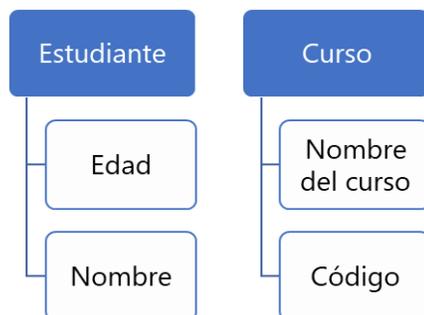
Atributos Polivalentes:

Los atributos "teléfono" y "correo electrónico" pueden contener múltiples valores.

Asimismo, los atributos pueden ser derivados. Un atributo derivado es aquel que representa un valor que se puede obtener a partir del valor de uno o varios atributos y que no necesariamente pertenecen a la misma entidad o relación.



Las relaciones también pueden tener atributos asociados. Se representan igual que los atributos de las entidades. Un ejemplo típico son las relaciones de tipo "histórico" donde debe constar una fecha o una hora (AIU).

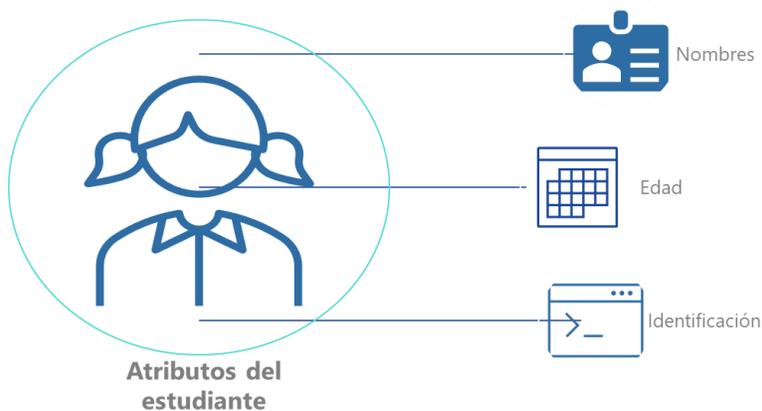


La relación entre las entidades "estudiante" y "curso" es la "matrícula"



La relación "Matricula" entre las entidades "Estudiante" y "Curso" tiene un atributo asociado: "Fecha de inscripción", proporcionado información adicional sobre la inscripción de un estudiante en un curso específico.

Siguiendo con el ejemplo de la operación estadística sobre Estadísticas de Matrícula de Educación Preescolar, Básica y Media, para la entidad «Estudiante» los atributos pueden ser :



Identificador

Un identificador de una entidad es un atributo o conjunto de atributos que determina, de modo único, cada ocurrencia de esta (AIU). Toda entidad tiene al menos un identificador y puede tener varios identificadores alternativos. Un identificador debe cumplir dos condiciones:

1. No pueden existir dos ocurrencias de la entidad con el mismo valor del identificador.
2. Si se omite cualquier atributo del identificador, la condición anterior deja de cumplirse.

En cuanto el ejemplo sobre Estadísticas Matrícula de Educación Preescolar, Básica y Media, se pueden agrupar, así:



Jerarquía de generalización

Una entidad E es una generalización de un grupo de entidades E, E, ... E, si cada ocurrencia de cada una de esas entidades es también una ocurrencia de E. Todas las propiedades de la entidad genérica E son heredadas por las subentidades.

Cada jerarquía es total o parcial y exclusiva o superpuesta. Una jerarquía es total si cada ocurrencia de la entidad genérica corresponde al menos con una ocurrencia de alguna subentidad. Es parcial si existe alguna ocurrencia de la entidad genérica que no corresponde con ninguna ocurrencia de ninguna subentidad. Una jerarquía es exclusiva si cada ocurrencia de la entidad genérica corresponde, como mucho, con una ocurrencia de una sola de las subentidades. Es superpuesta si existe alguna ocurrencia

de la entidad genérica que corresponde a ocurrencias de dos o más subentidades diferentes. Un subconjunto es un caso particular de generalización con una sola entidad como subentidad. Un subconjunto siempre es una jerarquía parcial y exclusiva (AIU).

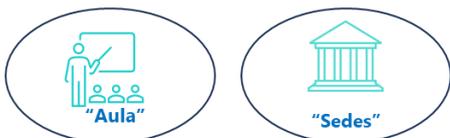
Entidades fuertes y débiles

Al participar una entidad en una relación adquiere un papel fuerte o débil. Se considera que una entidad débil cuando no puede existir sin participar en la relación, esta no puede ser unívocamente identificada solamente por sus atributos.

Ejemplos,

Ejemplo 1

Entidades



"Aula" es la Entidad débil

Explicación: puede haber **aulas identificadas** con la **misma numeración**, pero **en sedes diferentes**, por lo que es **necesario** que para identificar un aula se deba **conocer la localización de la sede**.

Ejemplo 2

Entidades



"Establecimiento educativo" es la Entidad débil

Explicación: un **establecimiento educativo** que **depende** de una secretaría de educación de **Entidad Territorial Certificada (ETC)**, a la que pertenece, y su **código identificativo** incluye al **código** de la secretaría de educación de la **ETC** de la que depende.

Una entidad fuerte sí puede ser identificada unívocamente. Si se requiere una entidad fuerte puede facilitar algunos de sus atributos a una entidad débil para que se pueda identificar (AIU). Continuado con el ejemplo del sector de educación:



“Alumno”



“Calificaciones”



Entidad Fuerte

Explicación: los atributos de entidad “alumno” pueden incluir el ID del alumno, nombre, apellido; por otro lado, la entidad “calificaciones” se utiliza para almacenar las calificaciones de los alumnos. Los atributos pueden incluir el ID del alumno, el ID del curso, la calificación obtenida. Por lo anterior, “alumno” puede ser identificada de manera única por sus propios atributos sin necesidad de depender de la entidad “Calificaciones”.

Cardinalidad

La cardinalidad de un atributo indica el número mínimo y el número máximo de valores que puede tomar para cada ocurrencia de la entidad o la relación a la que pertenece (De Miguel, Piattini y Marcos; AIU). Es decir, el número de elementos o instancias de una entidad A que se relacionan con un elemento de la entidad B y viceversa.

Si se tiene un conjunto de relaciones en el que participan dos o más conjuntos de entidades, la correspondencia de cardinalidad indica el número de entidades con las que puede estar relacionada una entidad dada. Estas pueden ser:

- Uno a uno: una entidad de A se relaciona únicamente con una entidad en B y viceversa.

Ejemplo:



- Uno a varios: una entidad en A se relaciona con cero o muchas entidades en B. Pero una entidad en B se relaciona con una única entidad en A.

Ejemplo:



- Varios a uno: una entidad en A se relaciona exclusivamente con una entidad en B. Pero una entidad en B se puede relacionar con 0 o muchas entidades en A.

Ejemplo:



- Varios a varios: una entidad en A se puede relacionar con 0 o muchas entidades en B y viceversa (De Miguel, Piattini y Marcos; AIU).
Ejemplo:

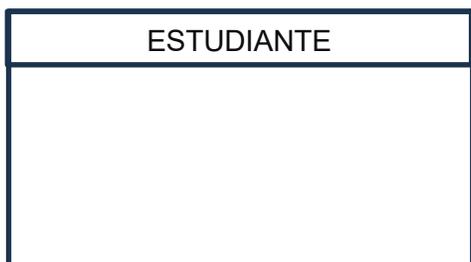


4.2. Notación para la elaboración del modelo ER

En un modelo ER, las notaciones se utilizan para representar a las entidades y sus relaciones. Existen diferentes tipos de notación, a continuación, se presenta la notación *crow's foot* o pata de gallo (Dybka, 2016), la cual ha sido seleccionada por ser una de las más concisas, cercana a la modelo relacional y ampliamente implementada por las herramientas de software para generación de modelos entidad relación. A continuación se presentan las distintas notaciones sugeridas:

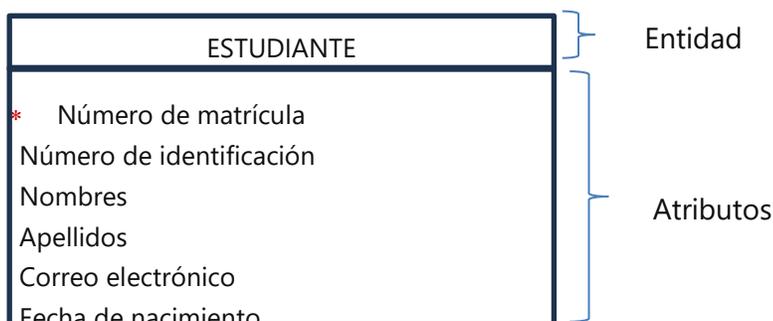
Entidades: una entidad es representada por un rectángulo, con su nombre en la parte superior y redactado en forma singular, por ejemplo:

Imagen 1. Entidad



Atributos: el atributo de una entidad se representa incluyendo su nombre dentro del rectángulo que representa la entidad a la que pertenecen, debajo del nombre de la entidad.

Imagen 2. Atributos



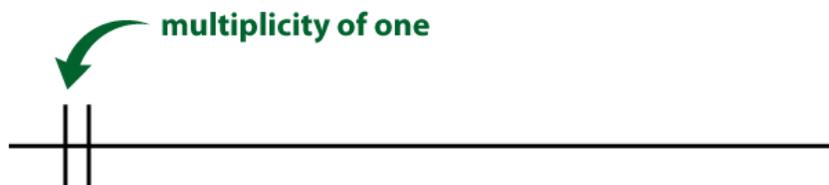
Identificador: el atributo o los atributos que distingue de manera única una instancia de la entidad es el identificador el cual puede estar marcado con un asterisco o el dibujo de una llave. En el ejemplo, el identificador puede ser "Número de matrícula", ya que este número único permite identificar de manera exclusiva a cada estudiante dentro del sistema, tal como lo muestra la imagen 2.

Relaciones: las relaciones ilustran la asociación entre dos entidades y son representadas con una línea recta. Las líneas de conexión de las relaciones suelen tener una apariencia gráfica diferente en función del tipo de relación que exista entre las entidades que conectan (puede ser una línea continua o discontinua).

Cardinalidad: las relaciones tienen dos indicadores que son dibujados en los dos extremos de la línea. El primero (comúnmente llamado multiplicidad) se refiere al número máximo de veces que una instancia de una entidad puede ser asociada con instancias en la entidad relacionada. Podría ser uno o muchos.

La multiplicidad de uno se representa con dos líneas perpendiculares a la línea de la relación como se observa en la siguiente imagen:

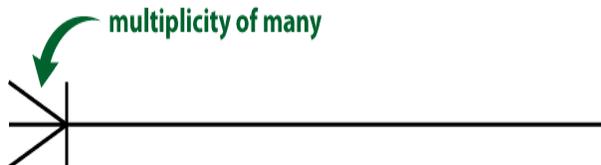
Imagen 3. Relaciones – multiplicidad de uno



Fuente: (Dybka, 2016).

La multiplicidad de muchos se representa con la pata de gallo como se observa en la siguiente imagen:

Imagen 4. Relaciones – multiplicidad de muchos



Fuente: (Dybka, 2016).

El segundo indicador describe el número mínimo de veces que una instancia de una entidad puede estar relacionada con otras entidades. Puede ser cero o uno y en consecuencia describe la relación como opcional u obligatoria. La obligatoria se representa con una línea perpendicular a la línea de la relación y la opcional se representa con un círculo.

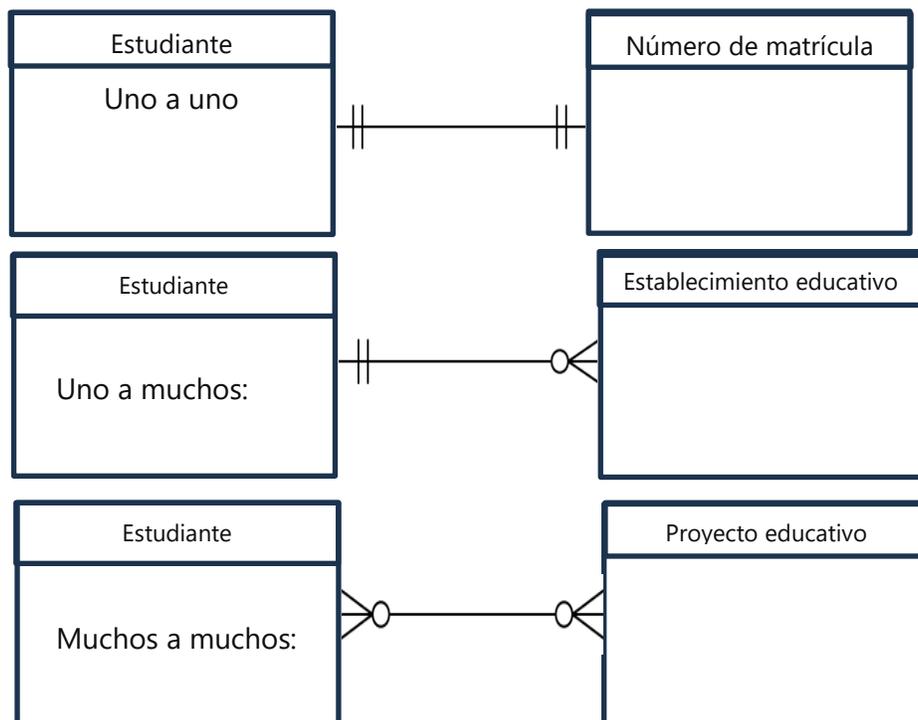
Imagen 5. Relaciones – obligatoria u opcional



Fuente: (Dybka, 2016).

La combinación de estos dos indicadores es siempre en un orden específico. Colocado en el borde exterior de la relación, el símbolo de multiplicidad es lo primero. El símbolo indicando si la relación es obligatoria u opcional es dibujado después del símbolo de multiplicidad. A continuación, se presentan los posibles tipos de relación, teniendo en cuenta el ejemplo relacionado con las estadísticas de educación:

Imagen 6. Tipos de relación



5. Buenas prácticas y recomendaciones para la elaboración del modelo ER

La correcta elaboración de modelos ER es un componente esencial en el desarrollo de bases de datos eficientes y coherentes, especialmente en el ámbito de la producción estadística. Este capítulo presenta un conjunto de buenas prácticas y recomendaciones dirigidas a optimizar el diseño y el uso de modelos ER, asegurando que estos modelos sean técnicamente sólidos, prácticos y adaptables a las necesidades específicas de cada entidad.

Las buenas prácticas aquí descritas se fundamentan en la experiencia y el conocimiento acumulado de diversas entidades y profesionales involucrados en la producción estadística. Estas prácticas buscan promover la claridad y la precisión en la representación de la estructura de datos, facilitando la comunicación entre diseñadores, desarrolladores y usuarios finales. Además, se enfatiza la importancia de identificar y corregir posibles redundancias e incoherencias en el diseño para contribuir a la creación de bases de datos robustas y eficientes.

Las recomendaciones incluidas en este capítulo también destacan la necesidad de estandarizar los modelos ER, aprovechando conocimientos previos y así asegurar una metodología uniforme en su elaboración. Este enfoque no solo reduce costos y esfuerzos, sino que también mejora la integridad y la reutilización de los modelos a lo largo del tiempo. Al implementar estas buenas prácticas y seguir las recomendaciones propuestas, las entidades del SEN pueden mejorar significativamente la calidad y la eficiencia de sus procesos estadísticos y garantizar la generación de datos precisos y confiables.

El diseño del modelo ER debería ser realizado por el equipo técnico de la oficina de sistemas o áreas afines con el entrenamiento y experiencia adecuados, dada la complejidad y la especificidad de esta tarea.

- Se recomienda que el equipo técnico asignado tenga formación y experiencia en diseño de modelos ER para asegurar la calidad del modelo.
- Implementar programas de capacitación continua en metodologías y herramientas actuales de diseño de modelos ER.
- Realizar evaluaciones periódicas de competencias para identificar necesidades de actualización y capacitación en el equipo técnico.
- Fomentar la obtención de certificaciones profesionales en áreas relacionadas con el diseño de bases de datos y modelos ER.

El diseño del modelo debería comenzar con la descripción de las actividades que se llevarán a cabo en la operación estadística, proporcionada por el equipo temático responsable.

- Organizar reuniones entre el equipo técnico de sistemas y el equipo temático para asegurar una comprensión clara de las actividades de las operaciones estadísticas.
- Mantener disponible la documentación detallada de la operación estadística donde se describan las actividades del proceso de producción definidas.
- Utilizar herramientas colaborativas para compartir y actualizar la documentación de manera eficiente.
- Validar los entregables con el equipo temático en cada etapa del proceso de diseño.

Un insumo para la elaboración del modelo es el documento de especificaciones de requerimientos, que debería ser validado por el equipo responsable de la operación estadística.

- Implementar un proceso de validación exhaustivo para el documento de especificaciones de requerimientos, involucrando a todos los responsables de la operación estadística.
- Realizar revisiones iterativas del documento de especificaciones para incorporar *feedback* y mantenerlo actualizado.
- Involucrar a los usuarios clave en el proceso de validación para asegurar que sus necesidades y expectativas sean consideradas.
- Realizar pruebas de concepto basadas en el documento de especificaciones para identificar posibles problemas antes de la implementación completa.

El modelo ER debería ser representado en un diagrama ER y utilizar conceptos y gráficos estandarizados.

- Utilizar estándares gráficos reconocidos para la representación de entidades y relaciones en el diagrama.
- Emplear software especializado para la creación y el manejo de diagramas ER.
- Capacitar al equipo en el uso de herramientas de diagramación para asegurar que todos puedan utilizarlas eficientemente.
- Documentar y mantener un historial de versiones de los diagramas ER, incluyendo cambios y justificaciones.

El documento del modelo debería describir detalladamente cada una de sus partes: entidades, relaciones, atributos, jerarquías, identificadores y cardinalidad.

- Proveer descripciones detalladas para cada componente del modelo, incluyendo ejemplos y casos de uso.
- Actualizar regularmente la documentación del modelo para reflejar cualquier cambio o mejora realizada.
- Implementar revisiones cruzadas de la documentación por diferentes miembros del equipo.
- Utilizar plantillas estandarizadas para la documentación de las diferentes partes del modelo.
- Incorporar *feedback* de los usuarios o los responsables de la operación estadística en la documentación para asegurar que sea útil y relevante.

El equipo responsable del diseño o rediseño de la operación estadística debería socializar los aspectos metodológicos y temáticos con los responsables de la elaboración del modelo ER.

- Realizar presentaciones informativas detalladas sobre la operación estadística, incluyendo descripción de las fuentes de datos, los subprocesos, los instrumentos y las aplicaciones.
- Mantener y compartir documentos metodológicos y especificaciones de requerimientos actualizados con el equipo técnico.
- Organizar reuniones de seguimiento periódicas para revisar el progreso y resolver cualquier duda o problema.
- Mantener un registro detallado de las reuniones, incluyendo acuerdos, acciones pendientes y responsabilidades asignadas.

Identificar y nombrar correctamente las entidades en el modelo ER es crucial para asegurar la claridad y la funcionalidad del sistema.

- Seguir un proceso sistemático para identificar entidades a partir de las especificaciones de requerimientos funcionales.
- Aplicar normas de nombramiento claras y consistentes para las entidades para evitar ambigüedades.
- Validar el listado de entidades con el equipo temático para asegurar que todas las entidades identificadas son relevantes.
- Documentar todas las entidades en el diccionario de datos e incluir sus nombres, descripciones y cualquier alias o sinónimo utilizado.

Las relaciones entre entidades deberían ser claramente identificadas y documentadas, incluyendo la cardinalidad y los nombres de las relaciones.

- Identificar todas las relaciones entre entidades y utilizar especificaciones de requerimientos y realizar iteraciones de análisis.
- Documentar detalladamente las relaciones e incluir nombres, descripciones y cardinalidades.
- Validar las relaciones identificadas con el equipo temático para asegurar que todas las interacciones entre entidades sean correctas.
- Asignar nombres claros y descriptivos a las relaciones y evitar términos ambiguos.

Definir claramente los atributos y sus dominios para cada entidad y relación en el modelo ER.

- Identificar todos los atributos asociados a cada entidad y relación y asegurarse de que cada atributo sea único y esté correctamente asociado.
- Establecer los dominios de los atributos e incluir valores permitidos, tamaño y formato.
- Documentar cada atributo en el diccionario de datos e incluir su nombre, descripción, alias, tipo de dato, longitud, valores por defecto y cualquier otra característica relevante.
- Revisar los atributos con el equipo temático para asegurar que todos los atributos necesarios estén presentes y correctamente definidos.
- Implementar un proceso de gestión de cambios para actualizar los atributos y sus dominios cuando sea necesario.
- Promover que los atributos que tienen el mismo dominio y descriptor conserven el mismo nombre en todas las entidades que los contengan.

Revisar y validar el modelo ER con el equipo temático de la operación estadística para asegurar que cumple con todos los requisitos y las especificaciones.

- Realizar revisiones conjuntas del modelo con el equipo temático y asegurarse de que todos los aspectos del modelo sean revisados y validados.
- Implementar iteraciones de validación y ajuste del modelo y permitir el *feedback* y realizar correcciones necesarias.
- Realizar pruebas de concepto con el modelo para identificar posibles problemas y áreas de mejora antes de la implementación completa.
- Documentar todos los procesos de validación e incluir los resultados, los comentarios y cualquier acción tomada como resultado de la validación.

Utilizar herramientas de diagramación eficientes y seguir normas de diseño para crear diagramas ER claros y fáciles de entender.

- Seleccionar herramientas de diagramación que ofrezcan funcionalidades avanzadas para la creación y la gestión de diagramas ER.
- Seguir normas de diseño para evitar congestión en los diagramas, como evitar líneas cruzadas y utilizar ángulos rectos para las conexiones.
- Capacitar al equipo en el uso de herramientas de diagramación para asegurar que todos puedan utilizarlas eficientemente.
- Revisar y actualizar los diagramas regularmente para reflejar cualquier cambio o mejora en el modelo.
- Documentar y mantener un historial de versiones de los diagramas ER e incluir cambios y justificaciones.

Consolidar toda la documentación del modelo ER e incluir diagramas y descripciones detalladas, en un documento final comprensible y navegable.

- Estructurar la documentación final de manera clara y lógica para facilitar la navegación y la comprensión de los diferentes componentes del modelo.
- Realizar una revisión y aprobación final de la documentación con todos los responsables de la operación estadística.
- Asegurarse de que todos los responsables de la operación estadística tengan acceso a la documentación final.
- Implementar un plan de mantenimiento para la documentación y asegurarse de que se mantenga actualizada con cualquier cambio o mejora en el modelo.

Asegurar la protección de los datos utilizados en el modelo ER mediante la implementación de medidas de seguridad adecuadas.

- Se recomienda realizar auditorías de seguridad periódicas para identificar y mitigar posibles vulnerabilidades. Las auditorías periódicas ayudan a mantener la integridad y la seguridad de los datos.

- Es fundamental asegurar que todos los datos sensibles estén cifrados, tanto en tránsito como en reposo. El cifrado protege los datos contra accesos no autorizados.
- Implementar controles de acceso basados en roles para limitar el acceso a los datos solo a personas autorizadas. Los controles de acceso reducen el riesgo de brechas de seguridad.
- Establecer sistemas de monitoreo y registro de actividades para detectar y responder rápidamente a incidentes de seguridad. El monitoreo continuo ayuda a identificar y responder a amenazas en tiempo real.

Diseñar el modelo ER para asegurar su compatibilidad e integración con otros sistemas y bases de datos.

- Se recomienda utilizar estándares abiertos y protocolos de integración para asegurar la interoperabilidad. Los estándares abiertos facilitan la integración con otros sistemas.
- Realizar pruebas de compatibilidad con otros sistemas y bases de datos para asegurar una integración fluida. Las pruebas de compatibilidad ayudan a identificar y resolver problemas de integración.
- Es fundamental documentar detalladamente los procedimientos de integración e incluir configuraciones y dependencias. La documentación clara facilita el proceso de integración y la resolución de problemas.
- Capacitar al equipo técnico en los procedimientos y las herramientas de integración para asegurar una implementación exitosa. La capacitación adecuada asegura que el equipo esté preparado para manejar integraciones complejas.

Asegurarse de que el modelo ER esté optimizado para rendimiento y escalabilidad y que pueda soportar tanto las cargas de trabajo actuales como futuras.

- Se recomienda realizar pruebas de rendimiento bajo diferentes escenarios de carga para identificar posibles cuellos de botella. Las pruebas de rendimiento aseguran que el modelo pueda manejar las cargas de trabajo esperadas.
- Optimizar las consultas y el diseño de la base de datos para mejorar el rendimiento general. La optimización de consultas reduce el tiempo de respuesta y mejora la eficiencia del sistema.
- Planificar para la escalabilidad futura del sistema, considerando el crecimiento de los datos y usuarios. La planificación anticipada para la escalabilidad asegura que el sistema pueda crecer sin problemas.
- Implementar herramientas de monitoreo del rendimiento para identificar y resolver problemas de manera proactiva. El monitoreo continuo permite detectar problemas de rendimiento antes de que afecten a los usuarios.

Desarrollar procedimientos claros y sistemáticos para el mantenimiento y la actualización del modelo ER.

- Se recomienda implementar un plan de mantenimiento regular para revisar y actualizar el modelo, según sea necesario. El mantenimiento regular asegura que el modelo se mantenga actualizado y funcional.
- Capacitar al equipo en las mejores prácticas de mantenimiento y actualización del modelo. La capacitación asegura que el equipo pueda manejar tareas de mantenimiento de manera eficiente.
- Realizar evaluaciones de impacto antes de implementar cambios significativos en el modelo. Las evaluaciones de impacto ayudan a identificar posibles riesgos y mitigar problemas antes de que ocurran.
- Establecer un procedimiento claro para la gestión de incidentes relacionados con el modelo ER. Un procedimiento de gestión de incidentes asegura una respuesta rápida y eficiente ante cualquier problema que surja.

6. Actividades para la elaboración del modelo ER

La creación de un modelo entidad-relación ER es crucial para estructurar y asegurar la calidad de los datos en bases de datos utilizadas en operaciones estadísticas. Este capítulo presenta las actividades clave para desarrollar un modelo ER, desde la socialización del diseño de la operación estadística y la identificación inicial de entidades y sus atributos hasta la determinación de relaciones y cardinalidades. Cada paso está diseñado para guiar al equipo técnico y a los expertos temáticos en la captura precisa de las características necesarias de los datos, basándose en especificaciones de requerimientos funcionales y documentos relacionados.

Además, se subraya la importancia de iterar y validar el modelo con el equipo temático para asegurar que todas las relaciones sean adecuadamente documentadas. Este enfoque garantiza la creación de un modelo ER robusto y eficiente, que refleje fielmente las necesidades del sistema y mantenga la calidad de los datos en las operaciones estadísticas.

6.1. Socialización del diseño de la operación estadística

El equipo responsable del diseño o el rediseño de la operación estadística presenta o socializa a los responsables de la elaboración del modelo ER, los siguientes aspectos metodológicos y temáticos del proceso estadístico:

- La descripción de las fuentes de datos (primaria y secundarias).
- La manera cómo se llevará a cabo cada uno de los subprocesos.
- Los instrumentos que se utilizan y las aplicaciones que son requeridas en las fases de recolección o acopio, procesamiento, análisis y difusión.

Estos aspectos metodológicos se encuentran definidas en el documento metodológico y en las especificaciones de requerimientos de instrumentos y aplicativos.

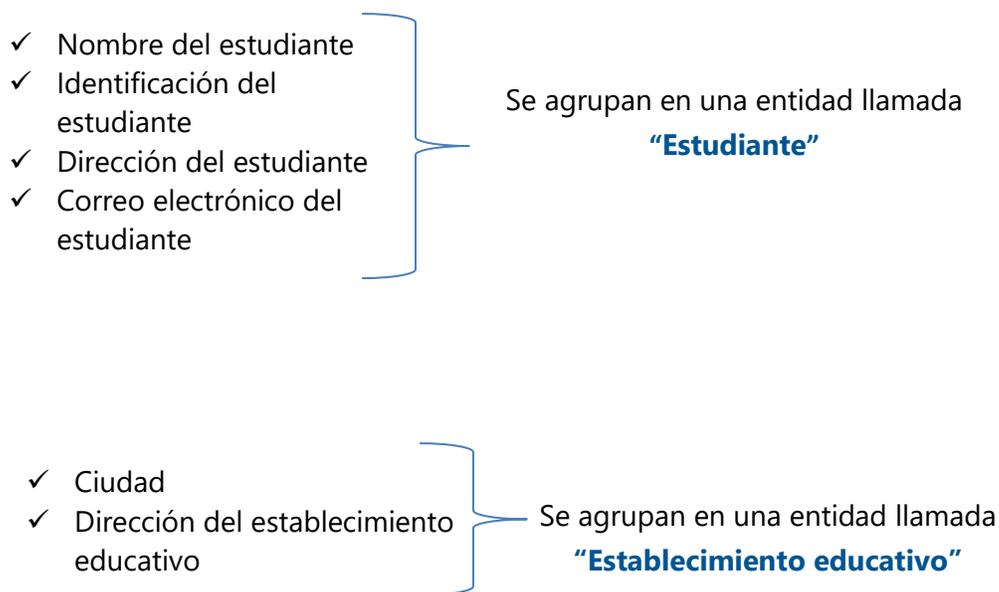
6.2. Determinar las entidades

Una forma de identificar las entidades es examinar las especificaciones de requerimientos funcionales para la generación de los instrumentos o los aplicativos necesarios para el desarrollo de las distintas fases en las operaciones estadísticas.

Dado que las entidades generalmente son sustantivos u objetos, se recomienda enlistarlos para ser analizados, definiendo los principales objetos de la operación estadística.

Para facilitar la identificación de las entidades de manera correcta, se recomienda que las descripciones de las especificaciones de requerimientos funcionales estén redactadas con un lenguaje técnico, riguroso y claro, evitando el uso de ejemplos o analogías, así como de sinónimos y homónimos³; en caso de que las especificaciones funcionales sean confusas, es necesario aclararlas con los responsables del diseño de la operación (AIU).

Del listado de entidades identificadas, se eliminan aquellas que son simplemente propiedades de otras. Por ejemplo:

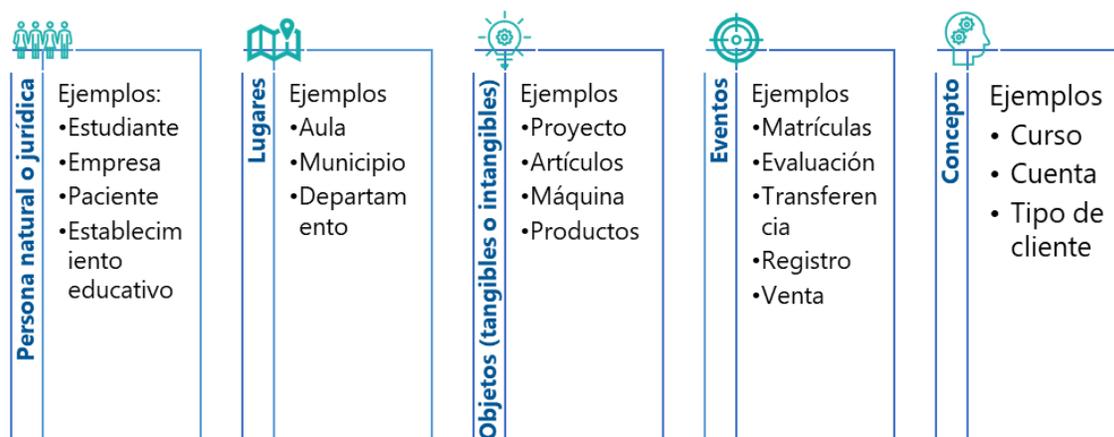


Puesto que el desarrollo de las actividades para obtener un modelo ER implica un alto grado de subjetividad y puede dar lugar a diferentes resultados por parte del personal técnico a cargo, es crucial que el listado de entidades sea validado y revisado por el equipo responsable del diseño de la operación estadística. Si las especificaciones de requerimientos funcionales y los documentos suministrados no

³ Dos palabras son sinónimos cuando tienen el mismo significado. Los homónimos ocurren cuando la misma palabra puede tener distintos significados dependiendo del contexto (AIU).

permiten definir un conjunto único de entidades, es necesario realizar múltiples iteraciones de análisis de los sustantivos para obtener las entidades adecuadas para el sistema que se desea construir (AIU).

A continuación, se presentan algunos ejemplos de entidades:



Es importante asegurarse que la entidad es relevante para la base de datos y su inclusión es necesaria para cumplir con sus requisitos; de igual manera que tienen existencia independiente

Cuando se identifican las entidades, se les da un nombre que tenga un significado claro y esté relacionado con la operación estadística. Es crucial que cada nombre se registre una sola vez en el esquema conceptual.

Los nombres de las entidades y sus descripciones se incluyen en la descripción del modelo y en el diccionario de datos. Cuando sea posible, se registra el número aproximado de ocurrencias de cada entidad.

A continuación, se presenta recomendaciones para nombrar a la entidad:

Nombre entidad: es el nombre que se le asigna a la entidad. Se sugiere que tenga las siguientes características:

- Generalmente los nombres de las entidades se convertirán en nombres de tablas, por lo tanto, es importante tener nombres entendibles, claros y de fácil identificación por los usuarios.

- Estar en singular.
- Ser únicos y no exceder los 30 caracteres.
- Los nombres cortos de las entidades son utilizados para crear restricciones y llaves foraneas y nombres de secuencias. Por lo anterior, se aconseja que tengan una longitud de 4 caracteres de acuerdo con la siguiente convención:
 - Si el nombre tiene 4 sustantivos, utilizar la primera letra de cada uno de ellos.
 - Si el nombre tiene 3 sustantivos, utilizar la primera letra de los dos primeros y las 2 primeras letras del tercero.
 - Si el nombre tiene 2 sustantivos, utilizar las 2 primeras letras de cada uno de ellos.
 - Si el nombre tiene 1 sustantivo, utilizar las 4 primeras letras de este sustantivo.
 - Si después de aplicar las reglas anteriores, la longitud total es menor a 4, completar con "S".

Ejemplos:

| CANTIDAD DE SUSTANTIVOS | NOMBRE DE LA ENTIDAD | NOMBRE CORTO |
|-------------------------|--|--------------|
| 4 sustantivos | INSTITUCIÓN EDUCACIÓN SUPERIOR PÚBLICA | IESP |
| 3 sustantivos | ENTIDAD TERRITORIAL CERTIFICADA | ETCE |
| 2 sustantivos | ÁREA INVESTIGACIÓN | ARIN |
| 1 sustantivo | ESTUDIANTE | ESTU |

Adicionalmente:

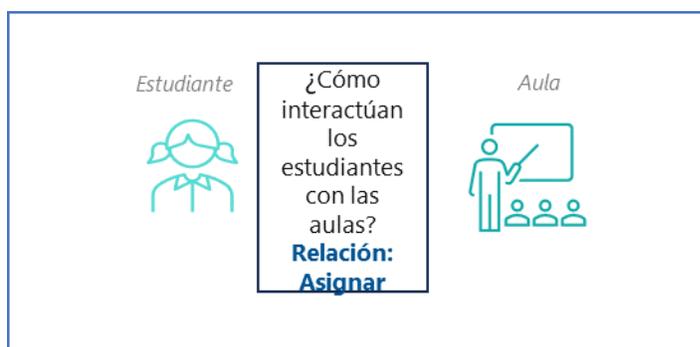
- El nombre plural de las entidades tendrá un máximo de 26 caracteres.

- Los nombres de las entidades no contienen carácter especial como: @, #, \$, %, *, o /.
- Los nombres de las entidades no contienen preposición.
- Los nombres de las entidades no contienen artículos como: un, una, el.
- Utilizar en la medida de lo posible abreviaciones estándar.

6.3. Identificar las relaciones

Luego de determinar las entidades, se procede a identificar las relaciones que existen entre ellas. Al igual que con la identificación de las entidades, se buscarán nombres en los instrumentos establecidos o en sus especificaciones de requerimientos⁴ para identificar las relaciones entre las entidades. Se suelen buscar las expresiones verbales relacionadas con las entidades, prestando atención solo a las relaciones que son necesarias.

Ejemplo:



La mayoría de las relaciones son binarias (entre dos entidades), aunque también puede haber relaciones en las que participen más de dos entidades. Es muy importante retomar las especificaciones de requerimientos funcionales y la documentación relacionada para comprobar que todas las relaciones, explícitas o implícitas, se han encontrado (AIU). De todos modos, la totalidad de las relaciones que no se identifican inmediatamente, sino son el resultado de varias iteraciones o se suelen encontrarse cuando se valida el esquema con el equipo temático.

⁴ Si las especificaciones de requisitos reflejan estas relaciones es porque son claras y esenciales para cada operación estadística y, por lo ende, deben reflejarse en el modelo ER.

Una vez identificadas todas las relaciones, hay que determinar la cardinalidad mínima y máxima con la que participa cada entidad en cada una de ellas. De este modo, el esquema representa de un modo más explícito la semántica de las relaciones. La cardinalidad es un tipo de restricción que se utiliza para comprobar y mantener la calidad de los datos (AIU).

Conforme se van identificando las relaciones, se les van asignando nombres que estén de acuerdo con la operación estadística. Se registra en la descripción del modelo y en el diccionario de datos, tanto los nombres de las relaciones, y su descripción, como las cardinalidades con las que participan las entidades en ellas.

Debido a que las relaciones son bidireccionales, al nombrar las relaciones se requiere de dos nombres. De aquí que la relación es legible en ambas direcciones.

Ejemplos:

- Cada persona puede estar localizada en una o más direcciones.
- Cada dirección debe ser la localización para una y solo una persona.
- Cada departamento debe ser responsable para uno o más empleados.
- Cada empleado debe estar asignado a uno y solo un departamento.

No se debería utilizar nombres de relaciones débiles, como: "asociado con" o "relacionado con". Las relaciones documentan reglas en la base de datos de la operación estadística y serán utilizadas para explicar los modelos a los usuarios finales. Cuando sea posible se recomienda utilizar terminología de la operación.

Por ejemplo, la forma de documentar una persona y su dirección sería:

| Forma correcta | No utilizar |
|--|--|
|  Un estudiante puede estar <i>asignado a</i> una o más aulas. |  Un estudiante puede estar asociado a una o más aulas. |

Algunos pares de nombres de relaciones útiles son:

- Identifica - Identificada por
- Define - Definida por
- Mantiene - Mantenida por
- Crea - Creada por
- Sirve - Servida por
- Patrocina - Patrocinada por
- Especifica - Especificada por
- Cita - Citado por
- Detalla - Detallado por
- Describe - Descrita por
- Relaciona - Relacionado por
- Clasifica - Clasificado por
- Utiliza - Utilizado por
- Autoriza - Autorizado por
- Transfiere - Transferido por

6.4. Establecer los atributos asociados a las entidades o las relaciones

Como se ha indicado, los atributos son las características que describen a las entidades o las relaciones. Al igual que con las entidades, se buscan nombres en las especificaciones de requerimientos funcionales y documentos relacionados. Los atributos comprenden nombres que identifican propiedades, cualidades, identificadores o características de entidades o relaciones.

Al identificar los atributos es importante considerar si son simples o compuestos. La elección entre un atributo simple o compuesto depende de los requisitos de la operación estadística. También deben identificarse los atributos derivados o calculados, que son aquellos cuyo valor se puede obtener a partir de los valores de otros atributos (AIU).

Al establecer los atributos, es posible identificar entidades que no se habían definido previamente. En tal caso, es necesario determinar nuevamente estas entidades e incluirlas en el listado y revisar si se relacionan con otras entidades.

Para asegurar una correcta asociación de atributos, es útil crear una lista de ellos y eliminarlos conforme se asignan a una entidad o una relación para garantizar que cada atributo se asocia únicamente a una entidad o una relación. Al finalizar la lista todos los atributos están asociados, sin embargo, a veces

parece que un atributo se asocia a múltiples entidades y esto puede ocurrir por alguna de las siguientes razones (AIU):

- Se han identificado varias entidades, no obstante, pueden representarse en una sola entidad.
- Se ha identificado una relación entre entidades. En este caso, se asocia el atributo a una sola de las entidades y hay que asegurarse de que la relación ya se había identificado previamente. Si no es así, se actualiza la documentación para recoger la nueva relación.

Conforme se van identificando los atributos, se les asignan nombres que sean claros para el equipo responsable de la operación estadística. De cada atributo se anota la siguiente información (AIU):

- Nombre y descripción del atributo.
- Alias o sinónimos por los que se conoce al atributo.
- Tipo de dato y longitud.
- Valores por defecto del atributo (si se especifican).
- Si el atributo siempre va a tener un valor (si admite o no nulos).
- Si el atributo es compuesto y, en su caso, qué atributos simples lo forman.
- Si el atributo es derivado y, en su caso, cómo se calcula su valor.
- Si el atributo es multi evaluado.

Para definir los nombres de los atributos es importante tener en cuenta:

- Ser lo más descriptivos posibles, utilizar términos de fácil reconocimiento por los usuarios y evitar el uso de acrónimos y abreviaturas.
 - Su longitud no exceda los 30 caracteres.
 - Los atributos utilizados en la identificación de registros y que sean denotados con la palabra ID, se les agrega el nombre completo de la entidad en singular.
 - No crear nombres de atributos o entidades que contengan espacios entre sus caracteres
 - No crear nombres de atributos o entidades que contengan caracteres especiales como tildes , o "\$#%&/
-
- Por ejemplo:

| ENTIDAD | NOMBRE CORTO DE ENTIDAD | ATRIBUTO |
|------------------|-------------------------|-------------------|
| TIPO INDICADORES | TIIN | ID TIPO INDICADOR |

- El nombre corto de los atributos (alias) pueden tener una longitud de 4 caracteres.

6.5. Determinar los dominios de los atributos

Un modelo ER no está completo si excluye los dominios⁵ de cada atributo: los valores permitidos para cada atributo, su tamaño y su formato. El equipo técnico de sistemas o los profesionales en sistemas deben determinarlos para cada uno de ellos. Adicionalmente, si se considera necesario, pueden incluir información adicional sobre los dominios como los atributos que pueden compararse entre sí o los atributos pueden combinarse con otros (AIU). Toda la información sobre los dominios se registran en la descripción del modelo y en el diccionario de datos.

Algunas recomendaciones para definir dominios son:

- Dominios asociados con un código de atributo finalizan con la palabra CD. Cuando se nombran estos códigos de dominio la práctica general es utilizar el mismo formato de nombrado de la entidad/atributo que es originalmente modificado.

Ejemplo:

| Entidad | Atributo | Dominio |
|---------------------------|---|------------------------|
| Establecimiento educativo | Código Único de Establecimientos Educativos | ESTABLECIMIENTOCUEE CD |

- Cualquier atributo que posea un rango de 20 o menos valores predefinidos se asocia con un dominio dinámico que describa el grupo de valores.

Ejemplo: Ejemplo de atributo que posee un rango de 20 o menos valores

⁵ El dominio de un atributo es el conjunto de valores que puede tomar el atributo.

| Entidad | Atributos | Dominio Dinámico | |
|-------------------|--|------------------|-----------------|
| Estudiante | Nombre | | |
| | Fecha de nacimiento | | |
| | Grado académico (Atributo con un rango de valores predefinidos) | | Preescolar |
| | | | Primaria |
| | | | Secundaria |
| | | | Bachillerato |
| | | | Técnico |
| | | | Especialización |
| | | | Maestría |
| | | | Doctorado |
| | | | Postdoctorado |
| | | | Diplomado |
| | | | Seminario |
| | Curso corto | | |

- Cualquier atributo que posea un rango de más de 21 valores o que se espera posea un grupo de valores cambiantes en el tiempo se colocan en una entidad de referencia, de esta forma los valores del dominio pueden ser mantenidos a través de un formulario.
Ejemplo de atributo que posee un rango de más de 21 valores:

| Entidad | Atributos | Rango |
|-------------------|--|---|
| Estudiante | Nombre | |
| | Fecha de nacimiento | |
| | Especialidad (rango de más de 21 valores) | Medicina, derecho, economía, ingeniería industrial, Biología, odontología, física, química, matemática, relaciones internacionales, ciencias políticas, contaduría, arquitectura, filosofía, sociología, antropología, comunicación social, teología, Ingeniería de Software, música, artes plásticas |

- Dominios estáticos sólo se utilizan para listas de valores que nunca cambiarán. Por ejemplo: Yes/No, True/False. Los otros dominios son dinámicos.

| Entidad | Atributos | Dominio estático |
|-------------------|--------------------------|---|
| Estudiante | Nombre | |
| | Fecha de nacimiento | |
| | Tipo de documento | <ul style="list-style-type: none"> • Registro civil • Tarjeta de identidad • Cédula de ciudadanía • Pasaporte |

6.6. Determinar los identificadores

El equipo técnico de sistemas o los profesionales en sistemas determinarán los identificadores del modelo ER. Cada entidad tiene al menos un identificador. En esta actividad se encuentran todos los identificadores de cada una de las entidades identificadas. Al determinar los identificadores se establece si una entidad es fuerte⁶ o débil⁷. Todos los identificadores de las entidades se registran en la descripción del modelo y en el diccionario de datos.

⁶ Una entidad es fuerte cuando tiene al menos un identificador.

⁷ Una entidad es débil cuando no tiene atributos que le sirvan de identificador.

6.7. Normalización del modelo ER

La normalización del modelo entidad-relación (ER) es un paso esencial para optimizar la estructura de los datos en una base de datos destinada a operaciones estadísticas. Este proceso tiene como objetivo minimizar redundancias, eliminar inconsistencias y garantizar que los datos sean eficientes y consistentes.

La normalización se lleva a cabo mediante la aplicación de las primeras tres formas normales (1FN, 2FN y 3FN), que son las siguientes:

- **Primera forma normal (1FN):** Garantiza que no existan grupos repetitivos o atributos multivaluados, asegurando que cada celda de la tabla contenga un solo valor.
- **Segunda forma normal (2FN):** Asegura que todos los atributos sean completamente dependientes de la clave primaria, eliminando dependencias parciales.
- **Tercera forma normal (3FN):** Elimina las dependencias transitivas, garantizando que ningún atributo no clave dependa de otro atributo no clave.

Actividades relacionadas con la normalización:

1. Revisar las entidades y atributos identificados para identificar posibles redundancias o inconsistencias.
2. Aplicar las reglas de normalización para estructurar correctamente las entidades y relaciones.
3. Realizar iteraciones para ajustar el modelo y asegurar su cumplimiento con las normas de normalización.
4. Documentar los cambios realizados durante el proceso de normalización y validar los resultados con el equipo temático.

Resultados esperados:

- Un modelo ER optimizado, con entidades y relaciones que cumplan con los principios de normalización.
- Documentación completa y detallada del proceso de normalización.

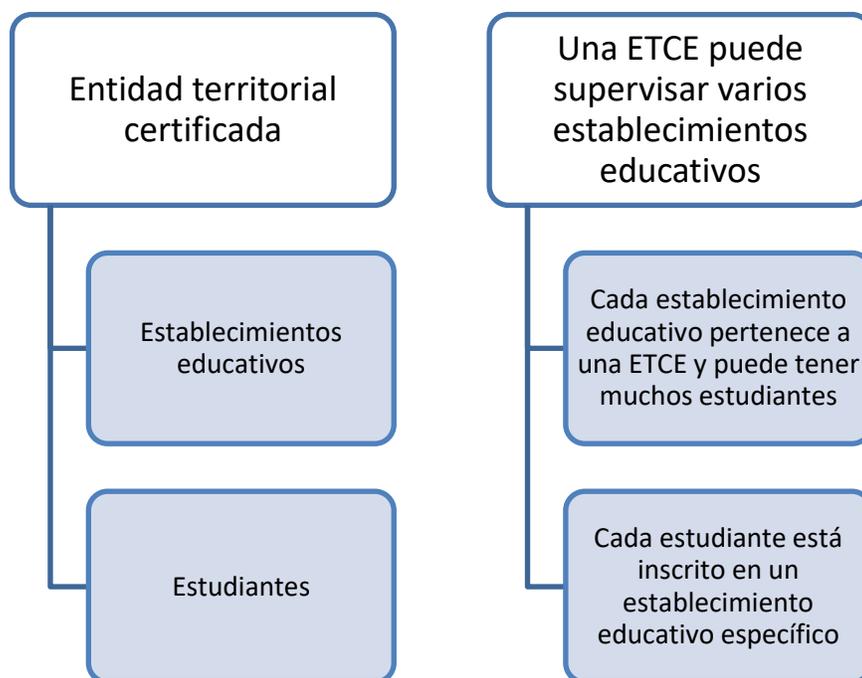
Esta actividad debe ser realizada antes de determinar las jerarquías y de dibujar el diagrama ER, ya que la normalización proporciona una base sólida para estas etapas posteriores.

6.8. Determinar jerarquías

Esta actividad parte del listado de entidades que se han identificado. Inicialmente, el equipo de sistemas evalúa si es necesario reflejar las diferencias entre distintas ocurrencias de una entidad, en el caso positivo surgirán nuevas subentidades de las entidades que se consideren genéricas. Luego, se revisa si existen entidades que compartan características comunes y que se consideraran como subentidades de una nueva entidad genérica. En cada jerarquía hay que determinar si es total o parcial y exclusiva o superpuesta (AIU).

Los dominios, los identificadores y las jerarquías obedecen a reglas funcionales o de negocio y validaciones de la operación y son identificadas y definidas por los usuarios junto con el equipo técnico en la fase de diseño del sistema de información.

Ejemplo de jerarquía:



6.9. Dibujar el diagrama ER

Al contar con todos los conceptos o partes del modelo ER, el equipo de sistemas es el encargado de elaborar el diagrama ER correspondiente a la base de datos para la operación estadística.

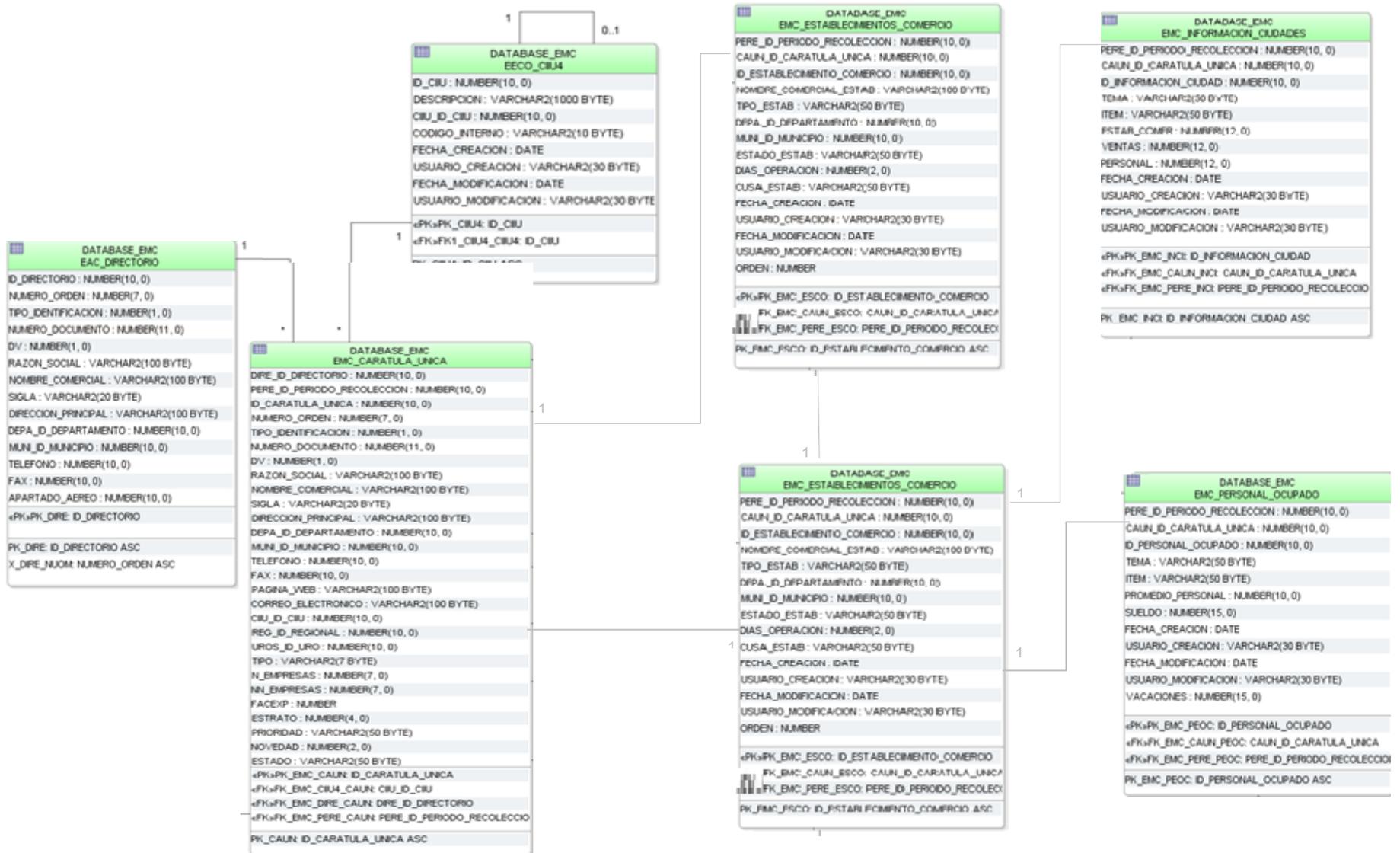
Con el fin de facilitar la lectura y la comprensión de los diagramas del modelo ER se sugiere considerar las siguientes normas de diseño (ver diagrama 1):

- Procurar dibujar el menor número de líneas cruzadas, paralelas juntas o muy cercanas, esto con el fin de evitar congestión.
- Las líneas deberían ser trazadas en un ángulo de 90° rectas en la medida de lo posible.
- Etiquetar las versiones de los diagramas añadiendo título, fecha e identificar el autor o los autores de cada diagrama.
- El texto no debe ser ambiguo, evitar abreviaturas y estar alineado horizontalmente para facilitar la lectura.
- El grado de relación entre entidades se ubica a la izquierda o en la parte superior de la línea de relación.
- Las entidades débiles pueden representarse por un rectángulo de doble pared.
- Las entidades hijas se ubican en la parte inferior de las entidades padres de las cuales heredan atributos.

6.10. Revisar el esquema conceptual con el equipo temático de la operación estadística

Antes de finalizar el diseño del modelo ER o el diseño conceptual, el equipo temático de la operación estadística revisa el diagrama y los contenidos que lo describen, con el fin de identificar alguna anomalía, error o nuevo concepto y realizar los cambios necesarios y oportunos. Es posible que sea necesario repetir alguna de las anteriores actividades, así como su revisión, hasta que el esquema conceptual, representación gráfica de la operación estadística y el manejo de los datos sean idóneos.

Diagrama 1. Modelo Entidad Relación Encuesta Mensual de Comercio – EMC: Módulo de captura



Fuente: DANE.

6.10. Consolidar documento

Teniendo en cuenta los ajustes sugeridos, el equipo de sistemas consolida con el equipo temático el documento del modelo ER, con su diagrama y las descripciones de las entidades, las relaciones, los atributos, las jerarquías, los identificadores y las cardinalidades, que ayuden a comprender y navegar el mismo.

Bibliografía

- Atlantic International University (AIU). Bases de Datos. Obtenido de:
<http://cursos.aiu.edu/Base%20de%20Datos/pdf/Tema%203.pdf>
- Departamento Administrativo de la Función Pública (DAFP) (2021) Manual Gobierno de Datos. Obtenido de:
https://www1.funcionpublica.gov.co/documents/34645357/0/2021-12-15_Manual_gobierno_datos_v1.pdf/9109ab27-4769-ff70-819d-fa1d30ac5318?t=1639612799087
- Departamento Administrativo Nacional de Estadística (DANE), D. A. (03 de 10 de 2023). *Ley 2335*. Obtenido de: Por la cual se expiden disposiciones sobre las estadísticas oficiales en el país:
<https://www.suin-juriscol.gov.co/viewDocument.asp?ruta=Leyes/30050183>
- Departamento Administrativo Nacional de Estadística (DANE). *Sistema de consulta de conceptos estandarizados*. Obtenido de: https://conceptos.dane.gov.co/conceptos/conceptos_catalogo
- Departamento Administrativo Nacional de Estadística (DANE). (2020). Lineamiento para el proceso estadístico en el Sistema Estadístico Nacional. Obtenido de:
https://www.sen.gov.co/sites/default/files/migracion-files/sen/normatividad/Lineamientos_Proceso_Estad%3%ADstico_v2.pdf
- Departamento Administrativo Nacional de Estadística (DANE) (2020). Norma Técnica de la Calidad del Proceso Estadístico.
- Departamento Administrativo Nacional de Estadística (DANE), D. A. (16 de 11 de 2021). *Resolución 1553. Por la cual se reglamenta la difusión de la información estadística*. Obtenido de https://www.dane.gov.co/files/acerca/Normatividad/Resolucion1503_2011.pdf
- De Miguel, A., Piattini, M., y Marcos, E. (2000). *Diseño de Bases de Datos Relacionales*. México D.F., México: Alfaomega.
- Departamento Nacional de Planeación (DNP), D. N. (2019). *Guía modelos entidad relación*. Obtenido de:
<https://colaboracion.dnp.gov.co/CDTI/Oficina%20Informatica/Sistemas%20de%20informaci%C3%B3n/Gu%C3%ADas%20Formatos%20Plantillas/Lineamientos%20Modelos%20Entidad%20Relaci%C3%B3n.pdf>
- Dypka, P. (2016). Crow's Foot Notation. Vertabelo. Obtenido de:
<https://www.vertabelo.com/blog/crow-s-foot-notation/>
- Ministerio de Tecnologías de la Información y las Comunicaciones (Mintic) (2014). Guía Técnica de la información - Gobierno del dato G.INF.06. Obtenido de:
https://www.cvc.gov.co/sites/default/files/Sistema_Gestion_de_Calidad/Procesos%20y%20procedimientos%20Vigente/Normatividad_Gnl/G.Inf.06%20Guia%20Tecnica%20-%20Gobierno%20del%20dato%20V1%202014-Dic-30.pdf