



SISTEMAS INTEGRADOS DE INFORMACIÓN GEOGRÁFICA

DISEÑO DE BASES DE DATOS



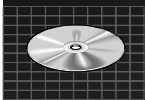
Jorge Fallas

Laboratorio de Teledetección y Sistemas de Información Geográfica

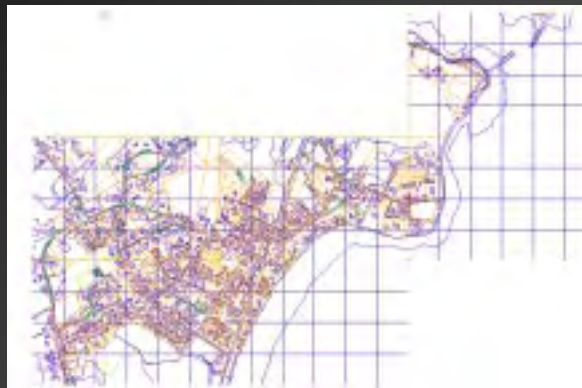
**Programa Regional en Manejo de Vida Silvestre y Escuela de Ciencias Ambientales.
Universidad Nacional. Heredia. Costa Rica.**

Email: jfallas@racsa.co.cr www.una.ac.cr/ambi/telesig/index.htm

1997



DISEÑO DE BASES DE DATOS



INDICE

INTRODUCCIÓN	<u>1</u>
ELEMENTOS A CONSIDERAR EN EL DISEÑO DE UNA BASE DE DATOS	<u>1</u>
OBJETIVOS DEL DISEÑO	<u>2</u>
ELEMENTOS PARA UN DISEÑO EXITOSO	<u>4</u>
ETAPAS EN EL PROCESO DE DISEÑO	<u>6</u>
A. DISEÑO CONCEPTUAL	<u>6</u>
1. Modelar la visión o perspectiva del usuario	<u>6</u>
2. Definir entidades y sus relaciones	<u>8</u>
Normalizar los datos	<u>13</u>
Matriz de diseño	<u>16</u>
3. Identificar representaciones de las entidades	<u>16</u>
B. DISEÑO LÓGICO	<u>18</u>
4. Asociar entidades con modelo de datos geoespacial en uso	<u>18</u>
5. Organizar las entidades en conjuntos o sets de datos geográficos	<u>19</u>
C. AUTOMATIZACIÓN	<u>23</u>
D. PROYECTO PILOTO	<u>23</u>
LITERATURA	<u>24</u>

INTRODUCCIÓN

Su institución dedicará hasta un 80% del presupuesto destinado a implementar el Sistema de Información Geográfica en elaborar y mantener la base de datos. Por esta razón es esencial que la organización dedique tiempo y esfuerzo a la fase de diseño, con el propósito de asegurarse que el modelo de SIG a implementar cumpla con las exigencias de la compañía y de sus usuarios. Los pasos que se presentan en el presente documento tienen como objetivo guiar al equipo que enfrentará el reto de diseñar e implementar la base de datos como parte del Sistema de Información Geográfica. Durante la fase de prueba inicial y aun durante la fase operacional del SIG surgirán otras dudas que tendrán que ser resueltas y documentadas. La guía no debe percibirse como un documento rígido, sino más bien como un listado de aspectos que deben considerarse y documentarse durante el proceso de diseño.

El primer paso en el proceso de diseño debe ser convocar a un pequeño grupo de colegas (usuarios reales y potenciales de la base de datos) para discutir sobre los siguientes aspectos:

1. Identifique lo que su empresa, departamento, sección o proyecto desea lograr a través de la base de datos y del uso de un Sistema de Información Geográfica.
2. Identifique los datos que su organización requiere para lograr los objetivos planteados.
3. Identifique las características de los datos (atributos).
4. Identifique el modelo de datos geoespacial que mejor se ajusta a sus datos. El modelo permite convertir los elementos y propiedades del mundo real al mundo digital de la computadora.

El producto de esta reunión debe ser una visión de conjunto de la institución, así como un primer listado de las funciones que debería cumplir el SIG. Otros subproductos útiles son listados de funciones, datos disponibles, formatos, necesidades de información, etc.

ELEMENTOS A CONSIDERAR EN EL DISEÑO DE UNA BASE DE DATOS

El Sistema de Información Geográfica (SIG) le permite a la institución realizar una serie de tareas que van desde las operaciones cotidianas hasta la planificación estratégica. Sin embargo, esta flexibilidad intrínseca del sistema sólo se logra cuando este se implementa de forma eficiente. El SIG le permitirá almacenar, analizar y compartir datos con diversos departamentos o unidades en su institución; así como integrar el uso de otras tecnologías como los Sistema de Posicionamiento Global (SPG), el procesamiento digital de imágenes y los sistemas de multimedia. EL diseño e implementación de una base de datos sólida y eficiente es un requisito para lograr un SIG exitoso y saludable. Normalmente la fase de diseño y elaboración de la base de datos georeferenciada puede consumir hasta un 80% del costo total del SIG (ESRI,1994). La clave para diseñar una buena base de datos es hacer las preguntas correctas a los individuos apropiados de la compañía. Por ejemplo, para iniciar el proceso puede responderse a las siguientes preguntas:

- ☐ Cómo puede implementarse el SIG para optimizar las operaciones y procesos que actualmente se realizan? o cambiar la forma en que se logra una meta particular?
- ☐ Cuáles datos son de más beneficio para la institución/empresa?

- ☐ Cuáles datos pueden compartirse entre departamentos?
- ☐ Quién es o deberá ser responsable por su mantenimiento?

Para responder a las preguntas anteriores se requiere de un buen conocimiento de la organización, de sus funciones o servicios que provee y de la tecnología SIG. La implementación de un SIG es similar a cualquier otra actividad que involucre la toma de una decisión. El proceso inicia con la definición o clarificación de las metas de la institución, luego de los procesos a diferentes niveles de detalle mediante los cuales se obtiene la información y finalmente la forma en que las funciones o servicios son implementados. Lo anterior nos lleva a definir:

- ▶ Metas
- ▶ Manejo de datos
- ▶ Requerimientos de uso de datos
- ▶ Proceso de colecta de datos
- ▶ Implementación de servicios

En la mayoría de los proyectos esta fase recibe poca o ninguna atención; sin embargo las razones para dedicar un pequeño porcentaje del tiempo y de los recursos humanos y financieros de la institución a esta fase son:

- ▶ El producto de su esfuerzo es una base de datos pobremente diseñada y que no responde a las necesidades de sus usuarios
- ▶ Los datos incluidos en la base de datos pueden estar duplicados; datos necesarios no existen o por el contrario contienen datos que no son necesarios.
- ▶ La representación de los datos es inapropiada.
- ▶ Las técnicas en uso para la gestión de la base de datos son inapropiadas.
- ▶ Pérdida de credibilidad del sistema
- ▶ Finalmente, fracaso de la iniciativa, pérdida de la inversión y eliminación del SIG.

Un diseño exitoso de la base de datos involucra no solo aquellos aspectos propiamente relacionados con la base de datos sino también con la aplicación en la cual será utilizada.

OBJETIVOS DEL DISEÑO

El diseño es un proceso que involucra tanto a los administradores como a los usuarios y al personal técnico que utilizará el SIG. A lo largo del proceso se definen los objetivos y metas, se estudian las alternativas de diseño y se prepara un plan de implementación. La meta final del proceso es generar un diseño que asegure flexibilidad, fiabilidad y consistencia en la base de datos. En términos generales, el diseño provee a la compañía con una imagen de donde se encuentran, a donde se dirigen y como llegar al estado deseado. Conforme se avanza en el proceso de diseño se definen los datos requeridos y la

estructura de datos geoespaciales que mejor se ajusta a los usos que se derán a la base de datos y al SIG. Una base de datos bien diseñada debe:

- ▶ Cumple con los objetivos para los cuales fue creada y apoya el logro de las metas de la empresa (apoya su desarrollo institucional).
- ▶ Contiene sólo los datos necesarios para logra las metas de la empresa. Datos redundantes no son permitidos en una base de datos, excepto cuando forman parte explícita del diseño de la base de datos.
- ▶ Los datos están organizados de tal forma que todos los usuarios tienen acceso al mismo set de datos. Esto asegura la integridad de la información que se utiliza en la compañía o empresa.
- ▶ El diseño es lo suficientemente flexible como para suplir la necesidades de múltiples usuarios.
- ▶ Permite mantener organizados tanto el componente espacial como los atributos de los elementos que utiliza la empresa o institución.

Una base de datos correctamente diseñada ofrece los siguientes beneficios:

- ▶ Mayor flexibilidad en la recuperación y análisis de los datos.
- ▶ Incrementa la posibilidad de que los usuarios desarrollen aplicaciones utilizando los datos disponibles en la base de datos. Esto se logra cuando la base de datos almacena datos que pueden ser utilizados por diversos usuarios.
- ▶ Los costos de captura, almacenamiento y uso es compartido por diversos usuarios o departamentos. Esto racionaliza el uso de recursos en la institución.
- ▶ Una base de datos integrada facilita su mantenimiento y por lo tanto asegura la integridad de la información utiliza por los diferentes usuarios (facilita el gestión de las transacciones).
- ▶ El diseño es lo suficientemente flexible como para acomodar actualizaciones o modificaciones en el futuro.
- ▶ Minimiza los datos reduntantes y por lo tanto hace más eficiente la creación, mantenimiento y uso de la base de datos.

El diseño e implementación de la base de datos involucra tres fases o momentos bien definidos en el proceso, a saber:

Modelo conceptual: Datos necesarios para lograr los objetivos y metas de la empresa
Los datos seran utilizados para modelar datos geográficos y no geográficos y las relaciones que existen entre ellos.

Modelo lógico: Convergencia del modelo de datos geospaciales a utilizar con los requerimientos de datos por parte de la empresa o compañía.

Modelo físico: En esta fase se implementa y ajusta el diseño de la base de datos para optimizar su rendimiento considerando el software, equipo de cómputo y la configuración de la red de la institución.

ELEMENTOS PARA UN DISEÑO EXITOSO

A continuación se brindan algunas sugerencias para lograr un diseño armonioso y con un alto grado de aceptación de la base de datos:



Involucre a los usuarios: Recuerde que los usuarios son los que hacen al sistema exitoso. Además, ellos pueden proveer información de los procesos que realizan actualmente, así como sugerir cómo la base de datos podría facilitar su trabajo, su eficiencia y por ende redundar en ahorros económicos para la empresa. Finalmente, cuando el usuario se involucra en el proceso de diseño desde las fases tempranas se crea una sensación de pertenencia o propiedad del proyecto y por ende los comprometerá a utilizarlo una vez que esté en operación.



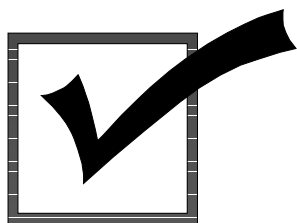
Focalice su esfuerzo: Aun cuando la meta del proceso de diseño es finalizar con una base de datos que cumpla los objetivos de la empresa; no es necesario crear un diseño detallado de todo el sistema en un solo proyecto. Al inicio, sólo es necesario tener una idea clara del diseño conceptual de la base de datos. Posteriormente se puede ejecutar cada componente por etapas, asignando los recursos necesarios a cada una de las tareas.



Forme un equipo de trabajo: Durante el proceso de diseño e implementación de la base de datos se requiere de información muy diversa y del concurso de individuos con múltiples disciplinas. El equipo de trabajo debe estar formado por individuos que conozcan las funciones de la organización que se modela, que posean conocimientos sobre entrevistas, sobre modelado, conocimientos sobre SIG y que además puede comunicarse con los administradores de la compañía.



Sea creativo (a): El proceso de diseño de la base de datos es un excelente momento para identificar y planificar aquellos mecanismos necesarios para optimizar los objetivos y metas de la organización. La tecnología y los medios para capturar, almacenar, visualizar y comunicar información cambian constantemente y por lo tanto el equipo de trabajo debe utilizar toda su creatividad para obtener un producto novedoso y eficiente.



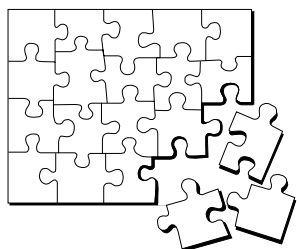
Genere productos: Elabore un plan de tal forma que los productos de las diferentes fases estén claramente especificados. Defina quién o quienes son los responsables, establezca fechas de culminación de cada proceso y asigne los recursos requeridos en cada fase. Por ejemplo, un producto de la primer fase de diseño debería ser la matriz de datos y funciones. Los productos son los logros por medio de los cuales se evalúa el avance en el proceso de diseño. Cuanto más pronto se identifiquen los errores o debilidades en el proceso de diseño de la base de datos menor será el costo requerido para sanear la base de datos.



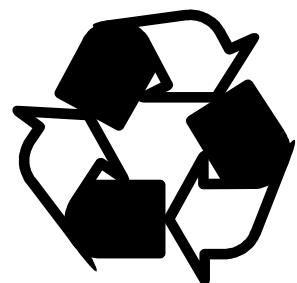
No adicione detalle innecesario: Los detalles tienen como objetivo proveer información para que el equipo tome las decisiones correctas en el momento correcto. No trate de definir elementos no requeridos en las primeras fases del diseño. Por ejemplo, no defina coberturas o atributos durante la primera fase del proceso. En esta fase sólo es necesario definir los objetivos que tendrá la base de datos.



Documente todos los pasos y decisiones durante el proceso de diseño: EL objetivo básico de documentar cada paso en el proceso de diseño es proveer la información requerida por el equipo de trabajo para tomar las decisiones correctas en el momento apropiado. EL uso de diagramas y tablas facilitará el proceso de comunicación y a la vez permitirá sintetizar en pocos documentos los elementos esenciales del proceso de diseño.



Mantenga el foco en sus objetivos y metas: El diseño de la base de datos y de los productos generados deben apuntar a cumplir con los objetivos y metas de la organización. En caso de dudas o confusión, consulte sus notas sobre objetivos y prioridades. Otra alternativa es convocar nuevamente a los usuarios del sistema y jerarcas de la empresa para clarificar dudas y redefinir objetivos y metas.



Sea flexible en el diseño: EL diseño esbozado por el equipo de trabajo es una guía y por lo tanto no debe utilizarse como una 'camisa de fuerza'. El diseño debe percibirse como un documento cambiante que debe adecuarse a las necesidades de la organización, a las nuevas tecnologías y a la mejor comprensión del uso de un SIG por parte del personal involucrado en el proyecto.



Planifique la implementación del modelo: Recuerde que el modelo es sólo una aproximación a la realidad y que además se requiere de recursos (humanos, tiempo, financieros) para implementarlo. Por ejemplo, debe establecerse prioridades, la forma en que se implementará cada etapa y los datos requeridos por cada aplicación.

ETAPAS EN EL PROCESO DE DISEÑO

El diseño de la base de datos se presenta a continuación como una serie de pasos. Algunos subproductos del proceso de diseño pueden incluir el identificar aplicaciones requeridas por la compañía, definir las necesidades de capacitación y el establecer estándares de calidad para la adquisición y mantenimiento de los datos. Los pasos son una guía y se dividen en:

A. DISEÑO CONCEPTUAL

1. Modelar la visión o perspectiva del usuario

El objetivo de esta fase es crear un lenguaje común de comunicación entre todas aquellos usuarios de la base de datos y a la vez crear una visión de conjunto de las funciones y operaciones que ejecutará el SIG. Durante esta fase usted debe:

► **Identificar las funciones que permiten a la empresa cumplir con sus objetivos y metas.**

El propósito del SIG es mejorar la eficiencia de su empresa y por lo tanto es esencial esbozar claramente cuales son los objetivos y metas de la compañía. La empresa implementa sus objetivos y metas a través de una serie de funciones que pueden involucrar a varios departamentos o unidades. Dados que las funciones de la empresa son más estables que su estructura organizativa se recomienda utilizarlas como guía en el proceso de diseño de la base de datos. Los pasos a seguir son:

1. Identifique las funciones de interés para el proyecto. Los diferentes departamentos pueden ser el punto de inicio para esta actividad.
2. Para cada función provee una breve descripción de las actividades que comprende. Algunas de ellas puede que no requieran de un SIG, sin embargo no deben excluirse pues se perderá la visión de conjunto.

3. Someta a evaluación el producto generado a la jerarquía de la institución. Esto permitirá aclarar dudas sobre el diseño, establecer las perspectivas correctas sobre beneficios de la base de datos y definir posibles fechas de conclusión del proyecto.

► **Identificar los datos requeridos para cumplir las funciones de la empresa.**

Una vez definidas las funciones identifique los datos requeridos para llevar a cabo dichas funciones. Determine cuáles funciones utilizan los datos y cuáles los generan; esto le será útil durante la fase de implementación de la base de datos. En esta fase concentre su atención en aquellas funciones que requieren o tienen un componente espacial. Un diagrama que muestre el flujo de datos entre funciones es esencial para determinar el flujo de información en la empresa. Por ejemplo, cuando los datos son generados por una función es necesario definir su naturaleza, colectarlos, almacenarlos y distribuirlos. Los datos que ingresan como insumos a una función son el producto de otra función o de datos recibidos de una entidad fuera de la compañía (Ej. proveedor de datos). En resumen, las preguntas claves a responder en esta fase son:

¿Quién es responsable por cada función? Esta debe ser la persona a entrevistar.

¿Con qué otras funciones interaccione cada función? ¿Quién es el responsable de esas funciones?

En cada caso debe definirse ¿Cuál es la naturaleza de la interacción?

Al responder a estas preguntas usted puede encontrar dos o más nombres o códigos que significan los mismo en diferentes funciones o con funciones que duplican la captura y almacenamiento de los datos. Durante el proceso de diseño esta situación de redundancia debe eliminarse o si se mantiene debe justificarse las razones para hacerlo. Al finalizar esta fase envíe una copia de los diagramas y del informe a los involucrados en las diversas tareas y funciones para su validación y comentarios.

► **Organizar los datos en sets o grupos de elementos lógicos o geográficos.**

Una vez definidas las funciones y los datos requeridos para llevar a cabo cada función organice dicha información en una matriz de funciones y datos. Ubique las funciones en el eje 'X' y los datos en el eje 'Y'. Reordene las funciones de tal forma que aquellas que utilizan datos similares esten juntas; de igual manera, aquellos datos con características similares también deberán estar juntos. Para cada celda de la matriz utilice una "C" para indicar que se crean datos y una "U" para indicar que sólo se utilizan los datos. Reorganice nuevamente la matriz de tal forma que las casillas con la letra "C" se alinean lo más cercano posible a la diagonal. La tarea de refinar esta matriz es una labor del equipo de trabajo involucrado en el diseño de la base de datos y puede incluir un largo proceso de negociación. Documente lo mejor que pueda cada área de discrepancia y la forma en que se llegó a la solución. La matriz de datos y funciones muestra:

- Clasificación de los datos a un alto nivel de organización de la compañía.
- Interdependencia entre los datos y las funciones que ejecuta la empresa.
- Funciones que crean datos utilizados frecuentemente por otras funciones.
- Interdependencia entre las funciones de la empresa.

- ▶ Funciones no interdependientes (no comparten datos con otras funciones).
- ▶ Datos redundantes. Seleccione la función que creará, mantendrá y compartirá los datos con las otras funciones. Esto asegura una sola fuente de datos y por lo tanto la integridad de la información.
- ▶ Funciones que sólo utilizan datos y por lo tanto dependen de otras funciones para operar.

La entrevista puede requerir de cuatro horas para entrevistar a cada grupo de personas que ejecutan una determinada función. Otras cuatro horas son requeridas para documentar las metas, objetivos y funciones de cada departamento o sección en la empresa o organización. La meta del proceso debe ser visualizar la matriz en una sola hoja de papel. En aquellos casos en que existe demasiado detalle, tanto las funciones como los datos deben generalizarse para cumplir con esta meta.

▶ **Definir un plan de implementación inicial**

Una vez visualizada las funciones y los datos requeridos por la compañía puede decirse cómo se implementará la base de datos. Por ejemplo, puede decirse que el proyecto es muy complejo y que ende debe dividirse en subcomponentes con diferentes fechas de conclusión. Otra posibilidad es que la matriz sugiera un orden de ejecución. En este momento también pueden discutirse otros aspectos tales como la necesidad de conformar un grupo de coordinación dado el volumen de datos a compartir por dos o tres funciones o departamentos; la necesidad de reorganizar algunas de las funciones; ó la capacidad física de la red de cómputo para responder a los requerimientos de información.

2. Definir entidades y sus relaciones

Una vez identificadas las funciones, los datos y sus interrelaciones se deben identificar objetos o elementos denominados *entidades* en el Sistema de Información Geográfica. En esta fase del diseño usted debe:

▶ **Identifique y describa las entidades**

La identificación y descripción de entidades debe hacerse para cada una de las funciones identificadas en la fase anterior. Basados en el trabajo desarrollado hasta el momento usted puede crear una primer versión de la definición de entidades. Esta primera aproximación debe validarse con los usuarios de la base de datos. Una vez creada la lista final de entidades y sus respectivas descripciones continúe con el siguiente paso.

▶ **Identifique y describa las relaciones entre las entidades**

A partir de la lista de entidades por función debe prepararse un diagrama que muestre las relaciones entre las entidades. Valide sus diagramas con los usuarios de la base de datos. Ellos le ayudarán a verificar las relaciones y a descubrir relaciones aún no documentadas. Algunos ejemplos de entidades y relaciones son:

- "Un interruptor controla el paso de energía eléctrica". Esta frase describe a una entidad que podemos denominar *interruptor*.

- ❑ "Una línea de baja tensión se conecta a un transformador". Esta frase describe una relación estructural entre dos entidades (*línea de baja tensión* y *transformador*).
- ❑ "Una subestación eléctrica está formada por transformadores, líneas de alto voltaje y líneas de bajo voltaje". Esta frase describe la agregación de entidades para formar una nueva entidad de mayor complejidad.

Esta fase del diseño es crucial porque se adiciona detalle a la visión que tiene el usuario sobre la base de datos. En esta fase es esencial la participación de los usuarios tanto en el proceso de definición como de validación de las entidades y sus interrelaciones. Un subproducto de esta fase puede ser la redefinición de la forma en que se ejecutan las tareas en la compañía.

- ❑ **Inicie la creación del diccionario de datos.** Este documento describe las características de la base de datos y es el producto de documentar las diferentes etapas en el proceso de diseño. El diccionario de datos es necesario para guiar la creación de aplicaciones, expandir la base de datos y asegurar su integridad. Inicie su diccionario de datos con la elaboración de diagramas que describan cada entidad y que a la vez muestren sus relaciones con otras entidades. Cada entidad debe contener la siguiente información:

- ✓ Nombre
- ✓ Definición
- ✓ Identificador único
- ✓ Atributos relevantes

Si se desea se pueden adicionar otros elementos tales como:

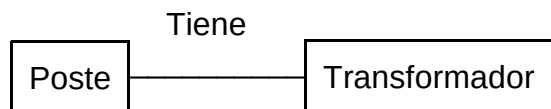
- ✓ Numero de entidades
- ✓ Fuente de datos
- ✓ Tasa de crecimiento esperada

Tabla de definición de entidades

Nombre:	Subestación
Definición:	Una estructura con transformadores, cables de alta tensión y un sistema de control.
Identificador:	Sub123
Atributos:	Area Capacidad No. abonados

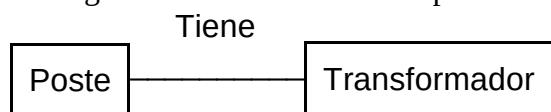
Tabla de relaciones	
Relación:	Transformación se conecta a líneas de baja tensión y de alta tensión.
Definición:	Comunica una línea de alta tensión con otra de baja tensión. Un transformador puede conectar una línea de alta tensión con varias líneas de baja tensión o estar ubicada al final de una línea de transmisión.

Las relaciones puede expresarse visualmente utilizando diagramas elaborados a partir de rectángulos y líneas. Los rectángulos representan las entidades y la línea el tipo de relación que existe en las entidades. Veamos un ejemplo:

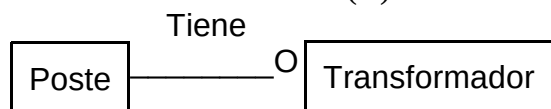


Este diagrama muestra que un poste posee un único transformador y que el transformador está asociado a un único poste. La simbología utilizada para representar relaciones es la siguiente:

- Una línea sin ningún otro símbolo al final representa una relación de uno a uno en la base de datos.

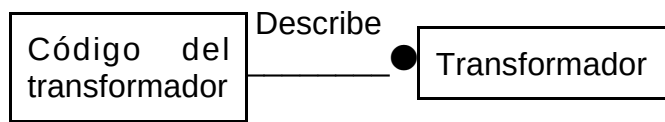


- Una línea con un **círculo en blanco** (O) al final de la línea indica una relación de cero o uno.



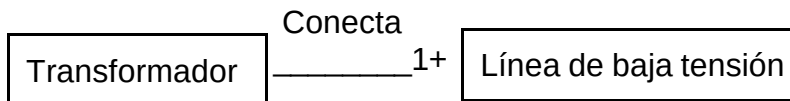
En este caso todo transformador está asociado a un poste pero no todos los postes están asociados a un transformador.

- Una línea con un **círculo en negro** (●) indica una relación de cero o más que uno.



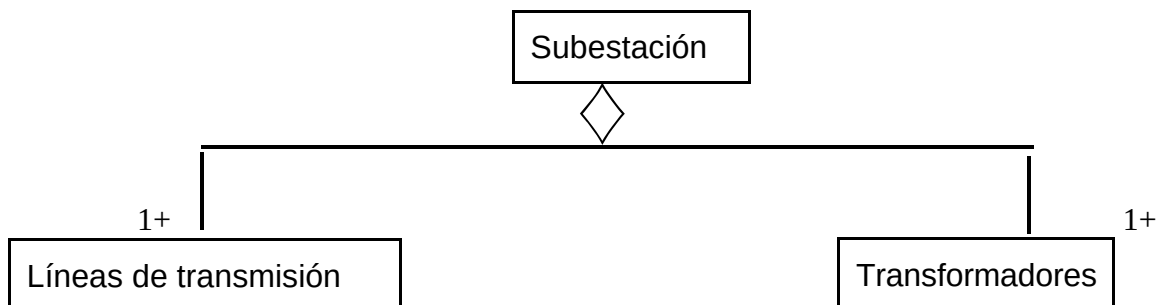
En este caso todo transformador está asociado a un código en el sistema de códigos utilizados para describir a los transformadores; sin embargo no todos los código deben estar asociados a un tipo de transformador. Por ejemplo, cuando se utiliza un sistema de códigos nacionales o internacionales algunos de ellos (tipos de transformadores) puede que no esten representados en el área de estudio.

- Una línea con un **1+** puede representar una relación de uno a uno así como una relación de una entidad a muchas entidades.



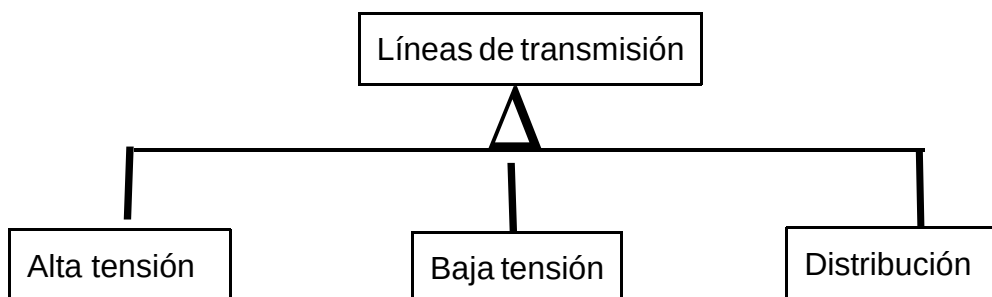
En este caso todo transformador debe estar conectado a menos una línea de baja tensión; sin embargo también puede estar conectada a varias. Este es un ejemplo de una relación de "muchos a muchos" en la base de datos.

- El **diamante** simboliza **agregación**, o sea, la integración o ensamblado de varias entidades con sus respectivas relaciones para formar una nueva entidad de mayor complejidad.



En este caso una subestación está formada por diversos tipos de líneas de transmisión y de transformadores.

- El **triángulo** simboliza **subclasificación** de entidades a partir de un elemento o servicio común.



En este diagrama no se cualifica o define el tipo de asociación entre las entidades ya que su principal función es permitir recrear la estructura funcional de la organización. El desarrollo de los diagramas de relaciones es un proceso dinámico, que usualmente requiere de varias repeticiones; por lo tanto esté preparado para modificar su diagrama a lo largo del proceso. Trate de visualizar las entidades y sus relaciones mediante diagramas o mapas conceptuales. Par facilitar el proceso comunicación prepare una lista de verbos que serán utilizados para definir las relaciones entre las entidades. Esta lista normalizada evita ambigüedades, el uso de términos duplicados (sinónimos) y racionaliza el proceso de diseño. Algunas preguntas que pueden ayudar a cualificar las relaciones son:

- ☐ Es posible tener (por ejemplo un poste) sin tener (un transformador)?
- ☐ Es posible para una entidad (Ej. línea de alta tensión) tener más que una relación (Ej. varios transformadores)?

Adicione palabras a la descripción de la relación de tal forma que reflejen las anteriores cualificaciones. Por ejemplo, "Un poste puede tener un transformador" o "Una línea de alta tensión puede estar conectada a uno o más transformadores". Para adicionar claridad a las relaciones utilice verbos que tengan sentido cuando se leen de izquierda a derecha (la forma tradicional de lectura occidental). Cuando esto no sea posible adicione una flecha la final de la línea para indicar la dirección en que la acción debe leerse. Otra recomendación es mostrar agregación o relaciones jerárquicas de arriba hacia abajo. Por ejemplo, la entidad compuesta o de mayor jerarquía debe ubicarse en la parte superior del diagrama.

Para facilitar la comunicación y evaluación del diagrama de relaciones no debe tratar de incluirlo todo en una página. Para fines de análisis segregue cada diagrama de acuerdo a sus clasificaciones mayores (Ej. subestación, sistema distribución, etc). Como regla general las entidades no deben repetirse en los diagramas; sin embargo cuando esto sea necesario por razones de claridad debe resaltarse el cuadro donde se encuentran para indicar que son la misma entidad. En esta fase del diseño no trate de resolver relaciones de 'muchos a muchos'; esta situación se resolverá más tarde en el proceso de diseño. A este nivel del diseño debe incluir todos los datos, aún cuando no tengan una representación espacial.

Normalizar los datos

La fase de normalización de la base de datos tiene como meta asegurarse de que cada descripción de entidades en la base de datos corresponde a una única entidad. Esto redundará en claridad en el diseño de la base de datos y facilitará su mantenimiento (Ej. adición, remoción, actualización) en el futuro. El proceso de normalización trata de buscar y eliminar *redundancias* en la base de datos; a la vez que asegura que el equipo de trabajo entiende las entidades con que trabaja y sus relaciones.

Durante el proceso de normalización debe seleccionarse un atributo que funcionará como identificador único o 'llave' en la base de datos y que permitirá identificar cualquier otro atributo asociada con dicho llave. En el mundo real normalmente la llave está formada por dos o más atributos y recibe el nombre de '*llave compuesta*'. Por ejemplo, el número de su cuenta corriente es una llave que identifica expresiones de su chequera (Ej, diferentes cheques). Esta es una relación de 'uno a muchos'. La tarea de normalización es un proceso que involucra varias repeticiones hasta llegar a un punto donde mayor desagregación de la base de datos es ineficiente o innecesario para los objetivos del proyecto. Para ilustrar el proceso utilizaremos la siguiente tabla de atributos de la entidad llamada *finca*. El código del registro es un valor único para cada finca.

CR-finca	Descripción legal	Ubicación	Valor \$	Dueño 1	Dirección del dueño 1	Porcentaje del dueño 1

Dueño 2	Dirección del dueño 2	Porcentaje del dueño 2	Area (Has)	Uso1	% de la finca

Uso2	% de la finca	Uso3	% de la finca

Primer paso en el proceso de normalización (1PN): En este primer paso se eliminan los grupos que se repiten y se descomponen los atributos compuestos. En nuestro caso los *grupos repetidos* son:

- ▶ Nombre del dueño
- ▶ Dirección del dueño
- ▶ Porcentaje del dueño
- ▶ Uso
- ▶ Porcentaje de uso

La forma en que está diseñada la tabla presupone que una finca sólo puede tener dos dueños y que cada finca sólo puede tener tres usos de la tierra. El incluir más que dos dueños o más que tres usos de la tierra sería muy complicado en esta tabla. La solución es separar estos atributos de la finca y crear dos nuevas tablas denominadas DUEÑO_CRFINCA y USO_CRFINCA. Estas tablas requieren de una llave para asociarse a las diferentes fincas. Para la primer tabla ésta llave puede ser *CódigoRegistro* y *nombre_dueño* y para la segunda el *CódigoRegistro* y *nombre_dueño*.

La ubicación de la finca y la dirección de los dueños son *atributos compuestos* que podrían segregarse en apartado postal, ciudad, distrito, cantón y provincia. Estas nuevas tablas podrían llamarse DIRECCIÓN_CRFINCA y DIRECCIÓN_DUEÑO y las llaves serían nuevos atributos (códigos) creados en la tabla principal. Al final de este primer paso las tablas resultantes serían:

FINCAS: Los grupos repetidos han sido removidos. Esta tabla solo contiene aquellos atributos propios de cada una de las fincas. Observe que la dirección de la finca y el uso de la tierra son utilizados como llaves para acceder las tablas de direcciones y de usos, respectivamente. Bajo esta estructura cuando una finca cambia de uso de la tierra sólo es necesario editar la tabla correspondiente a USO_FINCA.

Código_finca	Descripción legal	Ubicación (llave)	Valor \$	Area (has)	Uso_ llave
123TJL	Texto	10	1000	10	1

DIRECCIÓN_CRFINCA: El atributo dirección se ha descompuesto en sus elementos unitarios. La llave es el código de ubicación creado en la tabla anterior.

Ubicación_llave	Cacerío	Distrito	Cantón	Provincia
10	Sabanilla	San Pedro	Central	San José

DUEÑO_FINCA: Los nombres de los dueños se repiten en la tabla original y por lo tanto se separan para conformar una nueva tabla. Esta tabla sólo contiene atributos propios de los dueños de las fincas. La llave es el código de la finca.

CRF	Nombre Dueño	% de propiedad	Dirección
123TJL	Juan Pérez Villalobos	100	1200-3000 Heredia

USO_FINCA: Los usos de la tierra se repiten en la tabla original y por lo tanto se separan para conformar una nueva tabla. Esta tabla sólo contiene atributos propios del uso de la tierra para cada finca. La llave es el código de la finca.

Finca_Código	Uso1	% uso1	Uso2	% uso2	Uso3	% uso3
123TJL1	cafe	25	pasto	50	bosque	25

Segundo paso en el proceso de normalización (2PN): El objetivo de este segundo paso es asegurarse de que los atributos dependen únicamente de la llave principal. Algunas veces se requieren dos o más atributos para identificar entidades o relaciones únicas. Este caso se presenta cuando se tiene una relación de '*muchos a muchos*'. Por ejemplo, una finca puede tener varios dueños y un dueño puede tener varias fincas. En este ejemplo el atributo '*dirección del dueño*' califica a '*nombre del dueño*' y es independiente de la finca (*Finca_Código*). Sin embargo el atributo '*% de propiedad*' depende tanto del

nombre del dueño como de la finca '*Código_finca*'. Nuevamente los atributos de las tablas son segregados utilizando el criterio de dependencia única de la llave primaria.

Dueño-Código_finca: Los atributos de la tabla sólo dependen de la llave primaria (código de finca)

Código_finca	Nombre Dueño	% de propiedad
123TJL	Juan Pérez Villalobos	100

Dueño_finca: Los atributos que no dependen de la llave primaria (código de finca) forman una nueva tabla.

Nombre Dueño	Dirección
Juan Pérez Villalobos	1200-3000 Heredia

Tercer paso en el proceso de normalización (3PN): El objetivo de este tercer paso es asegurarse de que los atributos remanentes en la tabla principal solo describen aspectos relacionados con la llave principal y no con otros atributos de la tabla. Por ejemplo, la tabla que se muestra a continuación solo contiene atributos que corresponden a la finca y por lo tanto el proceso de normalización termina en este punto.

Código_finca	Descripción legal	Ubicación_llave	Valor \$	Area (has)	Uso_llave
123TJL	Texto	10	1000	10	1

La llave externa es un atributo de una entidad que sirve como identificador único para otra. Por ejemplo, el atributo Uso_llave sirve como liga o relación entre la tabla principal y la tabla sobre uso de la tierra para cada finca. Las llaves externas pueden estar ubicados en tablas que cumplen con las siguientes condiciones:

- ▶ Cuando la relación es '*uno a uno*', la llave externa puede formar parte de la tabla de atributos de cualquiera de las dos entidades. Por ejemplo, cada finca tiene sólo una dirección y cada dirección está asociada con sólo una finca.
- ▶ Cuando la relación es '*una a muchas*' (Ej. un distrito tiene muchos segmentos sensales, sin embargo un segmento sensal sólo puede pertenecer a un distrito), la entidad que exhibe la relación una a muchos (Ej. segmento sensal) debe contener la llave externa.
- ▶ Cuando la relación es de '*muchos a muchos*' (Ej. una finca puede tener varios dueños y cada dueño puede poseer varias fincas) se debe crear una tabla cruzada. En esta nueva tabla las entidades tendrán los dos identificadores, uno proveniente de cada una de los atributos que lo generan (Ej. Código de finca y código de dueño).

En algunos casos, el equipo que diseña la base de datos puede decidir no segregar un determinado atributo compuesto como por ejemplo ubicación_finca; ya que se espera que este atributo no cambie durante la vida útil de la base de datos. En caso de duda sobre si mantener o no los atributos como parte de la tabla principal, utilice como guía los siguientes criterios: la rapidez con que cambiarán los atributos; la reducción en el tiempo de acceso a la base de datos; el tamaño de la base de datos y el tiempo requerido para llevar adelante el proceso de normalización. Una vez finalizado el proceso de normalización debe actualizarse el diccionario de datos.

Matriz de diseño

Una vez concluida la tarea de normalizar y actualizar el diccionario de datos se debe elaborar una matriz que permita visualizar las entidades que contienen su diccionario. Es importante anotar todas las entidades, aún aquellas que no tienen una representación espacial. Cuando el diseño esté concluido esta matriz tendrá siete columnas, por ahora denomina la primera 'entidad' como se muestra a continuación:

Definición de entidades y sus relaciones

Entidad
Propiedad No. finca Código catastro Propietario Vías terrestres Calle Segmento División político-administrativo Segmento censal Distrito Cantón Provincia

Nota: Usted puede adicionar tantas entidades como sean necesarias en su proyecto.

3. Identificar representaciones de las entidades

En esta fase del diseño se responde a la siguiente pregunta ¿Como se representa la entidad en el mundo real? Por ejemplo, en el caso de entidades geométricas se utilizan puntos, líneas y polígonos para su representación; en tanto que en el caso de entidades no geométricas pueden utilizarse imágenes, dibujos o códigos alfanuméricos. Al establecer la forma en que se representará la entidad considere los siguientes aspectos:

- ¿Puede el elemento ser representada en un mapa?
 - ¿Es la forma del elemento importante para realizar análisis posteriores?
 - ¿Pueden los atributos del elemento accederse mediante su asocio con otro elemento en la base de datos?
- El elemento puede tener diferentes representaciones a diferentes escalas.

El tipo de elemento puede estar asociado a cualquiera de los siguientes términos:

- ☐ **Punto:** Ilustra la ubicación de un elemento cuya área es demasiado pequeña como para ser definida en un mapa a la escala de trabajo.
- ☐ **Línea:** Ilustra la ubicación de un elemento cuya forma es demasiado angosta como para ser definida en un mapa a la escala de trabajo.
- ☐ **Área:** Ilustra la ubicación de un elemento cuya forma está definida por un polígono en un mapa a la escala de trabajo.
- ☐ **Superficie:** Ilustra la forma de un elemento como un área o continuo; sin embargo también incluye forma derivada de cambios en elevación. En el SIG estos datos pueden representarse utilizando un TIN (Triángulos irregulares) o una lattice (puntos equidistantes).
- ☐ **Raster:** Representación de un área utilizando celdas rectangulares (Ej. imágenes de satélite, fotos escaneadas, etc.). Las imágenes raster son utilizadas con fines analíticos.
- ☐ **Imagen, foto, dibujo:** Representaciones gráficas de elementos en la base de datos cuyo uso es solo para fines visuales (sin capacidad analítica).
- ☐ **Alfa:** Identifique un elemento que no posee una representación geométrica o gráfica (Ej. dueño, código de la finca. etc.).

Además, podemos tener los siguientes elementos compuestos:

- ☐ **Región:** Este es un elemento compuesto. Una región puede formarse a partir de una serie de polígonos que se traslapan o simplemente de un conjunto de áreas no contiguas (Ej. archipiélago).
- ☐ **Rutas:** Es un evento lineal compuesto de uno o más arcos o partes de arcos. Las secciones (arcos o partes de arcos) son los componentes básicos de las rutas.

Cuando por efecto de escala un elemento puede representarse utilizando dos formas (Ej. punto y área) documente dicha posibilidad en el diccionario de datos y utilice la forma más compleja (área) en las siguientes fases del diseño. Al finalizar esta fase adicione una columna denominada '*representación espacial*' a su matriz de diseño y llene los respectivos campos.

Identificación de la representación espacial de los elementos

Entidad	Representación espacial
Propiedad	
Finca	área
No. finca	alfa
Codigo catastro	alfa
Propietario	alfa
Dirección	alfa
límite de la finca	línea
Vías terrestres	
Calle	línea
Sección	línea
Segmento	línea
Ruta de bus	línea
Parada de bus	línea
División político y administrativo	
Segmento sensal	área
Distrito	área
Cantón	área
Provincia	área
Medio natural y antrópico	
Monumento histórico	punto
Parque público	área
Cuerpo de agua	área
Cobertura vegetal	área
Imagen de satélite	imagen

B. DISEÑO LÓGICO

4. Asociar entidades con modelo de datos geoespacial en uso

El objetivo de esta fase es determinar la forma en que los datos serán representados en el Sistema de Información Geográfica. Para cada representación espacial (Ej. punto, línea, área) debe seleccionarse la correspondiente representación en el SIG. La atención en esta fase se torna a cómo diseñar una base de datos que responda a la necesidades del usuario en forma eficiente y efectiva. Durante esta fase usted debe:

- ▶ Determinar cuál es la forma óptima de representar cada entidad en el SIG
- ▶ Asegurarse que el diseño es compatible con elementos complejos (El. regiones, sistemas de rutas, eventos puntuales, eventos lineales). Esto elementos son formados a partir de elementos simples.

Identificación de la representación espacial en el SIG.

Entidad	Representación espacial	SIG
Propiedad		
Finca	área	región
No. finca	alfa	Base datos
Codigo catastro	alfa	Base datos
Propietario	alfa	Base datos
Dirección	alfa	Base datos
límite de la finca	línea	arcos
Vías terrestres		
Calle	línea	ruta
Sección	línea	sección
Segmento	línea	arco
Ruta de bus	línea	ruta
Parada de bus	línea	evento-punto
División político y administrativo		
Segmento censal	área	polígono
Distrito	área	región
Cantón	área	región
Provincia	área	región
Medio natural y antrópico		
Monumento histórico	punto	punto
Parque público	área	polígono
Cuerpo de agua	área	polígono
Cobertura vegetal	área	polígono
Imagen de satélite	imagen	grilla

5. Organizar las entidades en conjuntos o sets de datos geográficos

El propósito de esta etapa es agrupar las entidades en conjuntos de datos geográficos. En fase del diseño usted debe:

- ▶ Definir el contenido del conjunto de datos geoespaciales (Ej. coberturas, grillas, tins, dibujos-CADs)
- ▶ Dar un nombre a los espacios de trabajo, los conjuntos de datos geográficos, a las entidades y a los atributos.
- ▶ Completar la definición de las entidades
- ▶ Adicionar anotaciones y tablas auxiliares o de visualización.

Normalmente, todo Sistema de Información Geográfico puede operar con coberturas (mapas temáticos), grillas (estructura raster), Tins (vector) , imágenes (raster) y dibujos (vector-CADs). Estos datos pueden coexistir en un mismo directorio de trabajo, aunque en diferentes subdirectorios. Los datos

alfanumérico son almacenados en un base de datos de tipo relacional (Ej. Oracle, DB2, SQLserver, Sybase, Access, etc). Al agrupar entidades en conjuntos similares considere los siguientes aspectos:

- ▶ Identifique entidades que se refieren a un tópico similar. Por ejemplo, cuerpos de agua, sitios arqueológicos, sitios históricos, uso/cobertura de la tierra, red de distribución de electricidad, etc.
- ▶ Los puntos pueden representar tanto un polígono (centroide o identificador) como una entidad puntual (Ej. pozo, poste, etc.). Los puntos y los polígonos no deben coexistir en una misma cobertura.
- ▶ Una cobertura puede tener varias rutas basados en los arcos de una misma clase de elemento (Ej. diez rutas de buses). Sin embargo no puede mezclar dos tipos de clases de elementos (Ej. rutas de buses con rutas de tranvia).
- ▶ Una cobertura puede tener un número ilimitado de polígonos o líneas.
- ▶ Una cobertura puede tener múltiples regiones; sin embargo cada región debe estar formada por polígonos de un mismo tipo de elemento (Ej. parches de bosque, parcelas de café, áreas protegidas, etc.).

Entidad	Representación espacial	Representación SIG	Cobertura
Propiedad			
Finca	área	región	Finca
No. finca	alfa	Base datos	-----
Codigo catastro	alfa	Base datos	-----
Propietario	alfa	Base datos	-----
Dirección	alfa	Base datos	-----
límite de la finca	línea	arcos	Finca
Vías terrestres			
Calle	línea	ruta	Calle
Sección	línea	sección	Calle
Segmento	línea	arco	Calle
Ruta de bus	línea	ruta	Calle
Parada de bus	línea	evento-punto	Calle
División político y administrativo			
Segmento sensal	área	polígono	Pol_Adm
Distrito	área	región	Pol_Adm
Cantón	área	región	Pol_Adm
Provincia	área	región	Pol_Adm
Medio natural y antrópico			
Monumento histórico	punto	punto	Cultural
Parque público	área	polígono	Recreación
Cuerpo de agua	área	polígono	Ambiente
Cobertura vegetal	área	polígono	Ambiente
Imagen de satélite	imagen	grilla	Satélite (raster)

Una vez concluida esta etapa adicione las definiciones de las coberturas a su diccionario de datos. La definición de cada una de las coberturas brinda un resumen de la información almacenada en la base de datos. Información referente a sistemas de coordenadas, precisión de los datos y unidades de medición forman parte del resumen.

Definición de la cobertura	
Nombre de la cobertura:	Finca
Descripción:	La cobertura 'finca' contiene la información legal y financiera requerida por el departamento de avaluos para realizar sus funciones de cobro tributario. Esta información es también utilizada por el departamento legal y de permisos.
Entidades:	Finca Limites de la finca
Entidades relacionadas:	Foto de la finca Duenos Dirección
Precisión de las coordenadas:	Doble
Proyección:	Geográfica
Sistema de coordenadas:	Lambert Norte Costa Rica
Unidades:	metros

Para asegurar la compatibilidad de los nombres de los archivos entre diferentes sistemas utilice el sistema de nomenclatura 8.3 (ocho caracteres para el nombre y tres para la extensión). Las anotaciones constituyen archivos de texto a ser utilizados durante la fase de producción cartográfica. En los sistemas que operan en Windows (NT, 95) el usuario tiene acceso a los diferentes tipos instalados en el sistema, así como a las propiedades de cada tipo: color, tamaño, forma y orientación. Otro componente asociado a las entidades son las tablas de códigos para visualización. Estas tablas permiten definir códigos o colores para visualizar o imprimir los mapas. Por ejemplo, si se desea imprimir una cobertura con 200 polígonos que tiene sólo 5 usos de la tierra, no es necesario indicar el color asociado a cada uso en la tabla principal; en este caso sólo es necesario crear una tabla que asocie un color a cada uso de la tierra. Los archivos de atributos y de anotación, así como las tablas de visualización deben adicionarse a la matriz de diseño. A continuación se ilustra como documentar cada entidad en la base de datos.

Definición de entidades

Nombre del elemento:	bordefinca
Descripción:	límites de la finca levantados por el topógrafo
Tabla asociada:	bordes.aat
Atributos:	Para cada atributo utilice el siguiente formato. Nombre: descripción, formato (espacios, producto, tipo y decimales). A continuación se brinda un ejemplo para el atributo 'EjeX'
Coordenadas planas en metros; proyección Lambert Norte de Costa Rica; 8,18,F,3	
Cobertura:	Fincas
Tipo de elemento:	arcos
Precisión:	doble
Tabla para visualización:	bodefinca.lut (llave tipo línea)
Anotación:	fincas.anot (llave código finca)
Fuente de datos:	Plano suministrado por topógrafo (Benito Diaz, regente No.00001)
Escala original:	No aplica. Datos se entraron utilizando COGO.
Responsable:	Sección de Catastro

C. AUTOMATIZACIÓN

Una vez concluidas las fases de diseño conceptual y lógico debe diseñarse un plan para automatizar los datos. El objetivo de este plan es determinar la forma más efectiva y eficiente de introducir los datos (geoespaciales y atributos) al SIG. El equipo de trabajo debe considerar aspectos tales como volumen de datos, formatos existentes, automatización propia versus contratación de una compañía externa, equipo requerido, software, personal, control de calidad y métodos para el seguimiento de proyectos.

D. PROYECTO PILOTO

El proyecto piloto tiene como objetivo probar el diseño de la base de datos y del SIG bajo condiciones reales de trabajo pero en una pequeña área representativa del proyecto total a implementar. Esto permite refinar aspectos de diseño de la base de datos, optimizar los flujos de producción y finalmente dar por aprobado la implementación del SIG. La fase piloto es ejecutada mediante un proceso de prueba y error y por lo tanto la información generada no siempre puede utilizarse en la implementación real del proyecto. La duración del proyecto piloto puede ser variable, sin embargo puede extenderse desde 3 hasta 12 meses. Algunas organizaciones muy grandes y complejas pueden tener una fase piloto de hasta dos años. El proyecto piloto ejecuta cada operación/función del SIG solo una vez y normalmente trata de probar todas los aspectos involucrados en la operación del sistema (Ej. equipo, software, administración, datos, productos, usuarios). La fase de *prototipo* se puede ejecutar en cualquier etapa del diseño de la base de datos y, a diferencia del proyecto piloto, tiene como objetivo ejecutar una

y otra vez las funciones-operaciones que se quieren probar. Después de cada prueba se recibe retroalimentación de los involucrados y se vuelve a ejecutar el proceso. El prototipo es creado, probado y modificado tantas veces como sea necesario hasta que se logren los resultados deseados.

LITERATURA

ESRI. 1994. ArcInfo Data Base Design. Environmental Systems Research Institute. Educational Services. Redlands California, USA. 590p.

ESRI. 1994. ARCINFO Data Management: Concepts, data models, database design, and storage. Environmental Systems Research Institute. Redlands California, USA. Chapter 5:1-48.