

ANEXO 03

TUTORIAL: NORMALIZACIÓN DE LA BASE DE DATOS

Ejercicios

Instrucciones: Las actividades se realizan usando una base de datos que previamente se diseñó para los ejercicios; en el caso de que no se use ésta, se sugiere se vayan anotando, en la línea que aparece a un lado de los nombres (ej. *1_Boriginal* _____), el (los) nombre de la tabla que se van a usar o se vayan generando. También se sugiere ir marcando el recuadro izquierdo de cada uno de los pasos conforme éstos se vayan realizando.


Como se mencionó en el manual, en la sección 'Descripción y organización de la información original', se requiere primero examinar y determinar el contenido de cada campo, seleccionando los que contengan la descripción de localidad y datos referentes al lugar de colecta.

Identificación y selección de los campos que estén relacionados con la descripción de una localidad:

En Access abrir la base de datos llamada *capacitacion* que se encuentra en *C:/capacitacion.mdb*

- ☐ Abrir la tabla *1_Boriginal* _____. Revisar, identificar y seleccionar cuales campos de la tabla contienen la descripción de localidad y otros datos complementarios que se consideren útiles para el proceso de georreferenciación.

Nota: Si la tabla original ya posee un identificador único (autonumérico), éste debe conservarse; por el contrario, si no tiene un identificador único debe agregarse.

- ☐ Ir a 'Vista diseño'  de la tabla y agregar un Identificador único (id autonumérico), nombrar el campo como *Id_Georref*.
- ☐ Guardar los cambios a la tabla.
- ☐ Hacer una copia de la tabla *1_Boriginal*, seleccionar la tabla y dar un clic derecho eligiendo la opción de *copiar*, posteriormente nuevamente dar clic derecho y seleccionar *pegar* nombrándola como *1_Capacitacion_original*, en opción de pegar seleccionar *estructura y datos*. Dar clic en **aceptar**.
- ☐ Mediante una consulta de selección, elegir la tabla *1_Capacitacion_original* _____, seleccionar los campos que contengan información sobre la descripción de localidad: *Pais*, *Estado*, *localidad*, *elevación* y *Des_Localidad*, y el *Id_Georref* o el autonumérico original.

Nota: Los nombres de los campos pueden variar en cada base.

- ☐ Cambiar la consulta de selección a una de creación de tabla, y nombrar la tabla como: *2_Capacitacion_unifica* _____. Ejecutar la consulta con el botón: [!].

Una vez obtenida la tabla con la información necesaria para georreferir se debe hacer una revisión rápida de las localidades para detectar duplicados, es decir, localidades que sean idénticas en información aunque no en sintaxis y orden (registros redundantes); estos registros redundantes deben eliminarse por medio de una agrupación. Las agrupaciones deben realizarse siguiendo una metodología precisa para evitar el problema de la pérdida de relaciones entre los registros de las tablas agrupadas; la unificación y atomización involucran agrupaciones de las tablas, en las cuales se utilizan identificadores únicos llamados llaves primarias (primary key) PK y llaves foráneas (foreign key) FK.


Normalización

La Normalización es un proceso de reestructuración de la información que contiene una tabla de datos, lo cuál permite eliminar registros redundantes o falsos duplicados. Son dos las actividades para normalizar la información: unificación y atomización.

Para realizar la normalización de una tabla se requiere de un conjunto de consultas en el manejador de bases de datos utilizado (*Microsoft Access* en este caso), a continuación se ejemplifica el proceso de normalización.

Primer paso:

Unificación: Sustitución de los errores de captura entre descripciones homónimas.

- ☐ Ir a **Vista diseño**  de la tabla *2_Capacitacion_unifica* y hacer una copia del campo *Des_Localidad*, nombrándolo como *Descripcion_unifica*.
- ☐ Por medio de una consulta de actualización poner NG a las celdas (campos) que no tengan información. (usar en criterio *nulo*)
- ☐ Copiar la información del campo *Desc_Localidad* a *Descripcion_unifica*.

- ☐ Ordenar alfabéticamente el campo correspondiente a la división administrativa junto con el de *Descripcion_unifica*.
- ☐ En el campo *Descripcion_unifica* (referente a ‘unificación de las descripciones de localidad’), se procede a identificar los registros idénticos que difieran entre sí solamente por errores de captura; el registro que presente la sintaxis más adecuada sustituye a los otros. (La sustitución se realiza con las funciones ‘copiar’ (Ctrl + c) y ‘pegar’ (Ctrl + v) de Access).

Nota: Ver criterios de homogeneización, anexo del Manual.

Al finalizar este procedimiento, se realiza una consulta de agrupación de tabla basada en la columna duplicada de la descripción (*Descripcion_unifica*), con el fin de crear una nueva tabla con menos registros.

La consulta de agrupación permite eliminar los registros redundantes que no pudieron desecharse previamente debido a que presentaban errores de captura.

Atención:

Cada vez que se realice una agrupación de una tabla, se debe realizar el procedimiento siguiente:

- 1) Se seleccionan los campos a través de los cuales se quiere realizar la agrupación de una tabla, se agrupa a través de una consulta en Access y se crea una nueva tabla.
- 2) A la nueva tabla creada se le agrega una *llave primaria* (PK), y a la tabla original se le agrega un campo vacío que funcionará como *llave foránea* (FK).
- 3) En una consulta de actualización de Access se relaciona la tabla original y la tabla agrupada a través de los campos por los que se agrupó, y en el campo de la *llave foránea* (en la tabla original) se debe incluir la *llave primaria* de la tabla agrupada.

Cada vez que se realice una agrupación se debe realizar este procedimiento inmediatamente, y así, cuando se quiera integrar la información de la tabla agrupada a la tabla original, la relación entre las dos tablas se realizará solamente a través de los campos de *llave primaria* y de *llave foránea*.

Si a la tabla agrupada se le necesita realizar otra agrupación, el procedimiento se debe repetir nuevamente, tantas veces como se necesite agrupar una tabla.


A continuación se realizará el ejercicio de agrupación con la tabla *2_Capacitacion_unifica* _____.

- ☐ Concluida la etapa de unificación se realiza una consulta de selección para agrupar las descripciones iguales. Se siguen los siguientes pasos:

- En el cuadro de diálogo de consultas de Access se oprime el botón **Nuevo**, aparece un menú de opciones y se selecciona la opción de '**Vista diseño**'. Se agrega la tabla **2_Capacitacion_unifica** _____ (la tabla que será agrupada).
- Se seleccionan todos los campos de la tabla **menos** el *ID_Georref* y *Desc_localidad* y se transfieren a la sección inferior del cuadro de consulta.
- Se oprime el botón de 'Totales' en la barra de herramientas, representado con la letra sigma Σ .
- De esta consulta se debe crear una nueva tabla de trabajo.

- ☐ Seleccionar la opción de '**Consulta de creación de tabla**' ubicada en la barra de herramientas. En el cuadro de diálogo que aparece inmediatamente, nombrar la tabla como: **3_Capacitacion_atomiza** _____. Finalmente se ejecuta la consulta con el botón '**Ejecutar**' [!].

Se realiza el procedimiento para agregar la llave primaria (PK) y la llave foránea (FK):


- ☐ La tabla **3_Capacitacion_atomiza** _____ se abre en **Vista diseño**  y se agrega un nuevo campo *ID_1* para la PK, el campo debe ser de tipo autonumérico. Guardar cambios.
- ☐ En la tabla de **2_Capacitacion_unifica** _____ se agrega también un campo nuevo *ID_1*, de tipo numérico que permita duplicados, en donde posteriormente se le agregará la FK. Guardar cambios.
- ☐ Se abre una consulta de actualización y se agregan ambas tablas. Se relacionan las tablas a través de los campos que comparten.
- ☐ En la sección inferior del cuadro de consulta se agrega únicamente el campo de FK (*ID_1* de tipo numérico de la tabla **2_Capacitacion_unifica** _____), y en el renglón de '**Actualizar**' se introduce el campo PK (*ID_1* de tipo autonumérico de la tabla **3_Capacitacion_atomiza** _____); la llave primaria PK se introduce desde el generador de expresiones de la ventana de consulta.

Segundo paso:

Atomización de las descripciones en un formato definido.

- ☐ Abrir la tabla **3_Capacitacion_atomiza** _____
- ☐ Ordenar alfabéticamente el campo correspondiente a la división administrativa

junto con el de *Descripcion_unifica* (campo de localidades unificadas).

- ☐ En la tabla *3_Capacitacion_atomiza* _____ se agregan los 13 campos que se utilizarán para atomizar, éstos se agregan en la vista del diseño de la tabla (menú 'Ver', opción **Vista diseño** ) , ahí se especifican las características particulares de cada uno, como el nombre, el tipo de campo y el tamaño.

Nombre	Tipo	Tamaño	Descripción del campo.
RefGeoSup	Texto	100	Referencia geográfica de orden superior al rasgo principal de referencia
LocalidadPrincipal	Texto	100	Nombre de la Localidad o rasgo geográfico principal.
Distancia	Texto	100	Distancia recorrida de la localidad o rasgo geográfico principal al sitio, unidades de distancia abreviada (p.e: m,km,mi) y anotaciones adicionales a la distancia (p.e: aproximado, cerca) .
Dirección	Texto	100	Dirección cardinal y anotaciones adicionales a ésta.
ViaDeAcceso	Texto	100	Nombre de la vía de acceso y anotaciones adicionales.
Distancia_ort	Texto	100	Distancias ortogonales: unidades y dirección cardinal (p. ej. 10 km N, 5 km E) .
Referencia_comp.	Texto	250	Referencias complementarias para la ubicación de una descripción de localidad
2aLocalidadRef	Texto	250	Segunda localidad de referencia.
Información_adicional	Texto	255	Información que no aporta elementos a la ubicación geográfica de un sitio.
Altitud_ato	Texto	50	Dato de altitud contenido en la descripción.
Estado_ato	Texto	50	Nombre del estado si está contenido en la descripción.
Mpio_ato	Texto	50	Nombre del municipio si está contenido en la descripción.
Obser_ato	Memo		Observaciones de la atomización

- ☐ Una vez que la tabla *3_Capacitacion_atomiza* _____ tiene los campos para atomizar, desagregar cada descripción de localidad en sus distintos elementos considerando los criterios establecidos para la atomización.

Nota: ver criterios de atomización en el manual: Anexo 03_Criterios para la atomización de las descripciones de localidades

Una vez concluida la atomización los falsos duplicados han sido convertidos en registros redundantes, los cuales pueden ser eliminados con una agrupación sencilla.


- ☐ Después de atomizar todas las localidades hacer una consulta de actualización para poner NG a las celdas de los campos de atomización que no tengan información. (usar en criterio *nulo*).
- ☐ Realizar una consulta de selección para agrupar las descripciones iguales. Seguir los pasos siguientes:

- En el cuadro de diálogo de consultas de Access se oprime el botón **Nuevo**, aparece un menú de opciones y se selecciona la opción de '**Vista diseño**'. Se agrega la tabla **3_Capacitacion_atomiza** (la tabla que será agrupada).
- Se seleccionan todos los campos de la tabla **menos** el *ID_1*, *Descripcion_unifica*, *InformacionAdicional* y *Obser_ato* y se transfieren a la sección inferior del cuadro de consulta.
- Se oprime el botón de 'Totales' en la barra de herramientas, representado con la letra sigma Σ .
- De esta consulta se debe crear una nueva tabla de trabajo.

Nota: Si en la atomización quedaron campos sin información, éstos también pueden omitirse y agregar en la consulta sólo aquellos que contengan datos.

- ☐ Cambiar a la opción '**Consulta de creación de tabla**', nombrar la tabla como: **4_Capacitacion_Trabajo** _____. Se ejecuta la consulta [!].

Se realiza el procedimiento para agregar la llave primaria (PK) y la llave foránea (FK):

- ☐ La tabla **4_Capacitacion_Trabajo** _____ se abre en **Vista diseño**  y se agrega un nuevo campo *ID_2* para la PK, el campo debe ser de tipo autonumérico. Guardar cambios.
- ☐ En la tabla de **3_Capacitacion_atomiza** _____ se agrega también un campo nuevo *ID_2*, de tipo numérico, en donde posteriormente se le agregará la FK. Guardar cambios.

Nota: la tabla con la información atomizada, en este caso **3_Capacitacion_Atomiza** _____, debe quedar con dos identificadores el *ID_1* (autonumérico) y el *ID_2* (numérico)

- ☐ En la sección inferior del cuadro de consulta se agrega únicamente el campo de FK (*ID_2* de la tabla **3_Capacitacion_atomiza** _____), y en el renglón de '**Actualizar**' se introduce el campo PK de la tabla agrupada (*ID_2* de la tabla **4_Capacitacion_Trabajo** _____). El campo *ID_2* se introduce desde el Generador de expresiones de la ventana de consulta.
- ☐ Se ejecuta la consulta [!].

La tabla **3_Capacitacion_atomiza** _____ debe quedar con dos identificadores el *ID_1* (autonumérico) y el *ID_2* (numérico).

Tercer paso:

Creación de la tabla de trabajo: esta tabla debe contener los campos atomizados relacionados con la descripción de un sitio y los campos utilizados en el proceso de georreferenciación.

- ☐ A la tabla *4_Capacitacion_Trabajo* _____ se le agregan los campos utilizados para la georreferenciación.

Nombre	Tipo	Tamaño	Descripción del campo.
Id_formulario	Númerico, Entero largo, Indexado: Si (con duplicados)		identificador único para registrar en el formulario
No_Georref	Texto	250 caracteres	Explicación de la razón por la que el registro no pudo ser georreferenciado.
Analista	Texto	50 caracteres	Persona que hizo la georreferenciación del registro
Fecha	Tipo: fecha/hora (día/mes/año)		Fecha de georreferenciación del registro

- ☐ El campo *Id_formulario* debe llenarse con el identificador del campo *ID_2*. Copiar el campo *ID_2* y pegarlo al campo *Id_formulario*; así, se conserva intacto el *ID_2* mientras que el *Id_formulario* podrá modificarse. Se modifica al momento de georreferenciar cuando se encuentran dos o más descripciones iguales a las que se les asignará la misma coordenada; con el fin de introducir sólo un número al formulario, el cual servirá para todas las localidades que correspondan al mismo lugar.

Nota: Los campos correspondientes a las coordenadas, la incertidumbre, la fuente, etc. Están incluidos en la tabla de el formulario de Access.

Se Inicia la georreferenciación.

Secuencia de las tablas generadas en la organización de la información.

La secuencia de las tablas que se generaron durante la normalización es básicamente la siguiente:

- **Tabla 1.** Copia de la tabla original de registros.
- **Tabla 2.** Primera agrupación de la tabla original (eliminación de registros redundantes idénticos). En esta tabla se duplicó el campo de las descripciones de localidad para unificarlas y eliminar los registros redundantes que tienen errores o diferencias en la captura.
- **Tabla 3.** Segunda agrupación de la tabla (eliminación de localidades redundantes no idénticas). En esta tabla se agregaron los 13 campos de la atomización.
- **Tabla 4.** Tercera agrupación de la tabla (eliminación de falsos duplicados). En esta tabla creada se agregan los campos que se utilizan durante la asignación de coordenadas, y constituye propiamente la tabla de trabajo final.

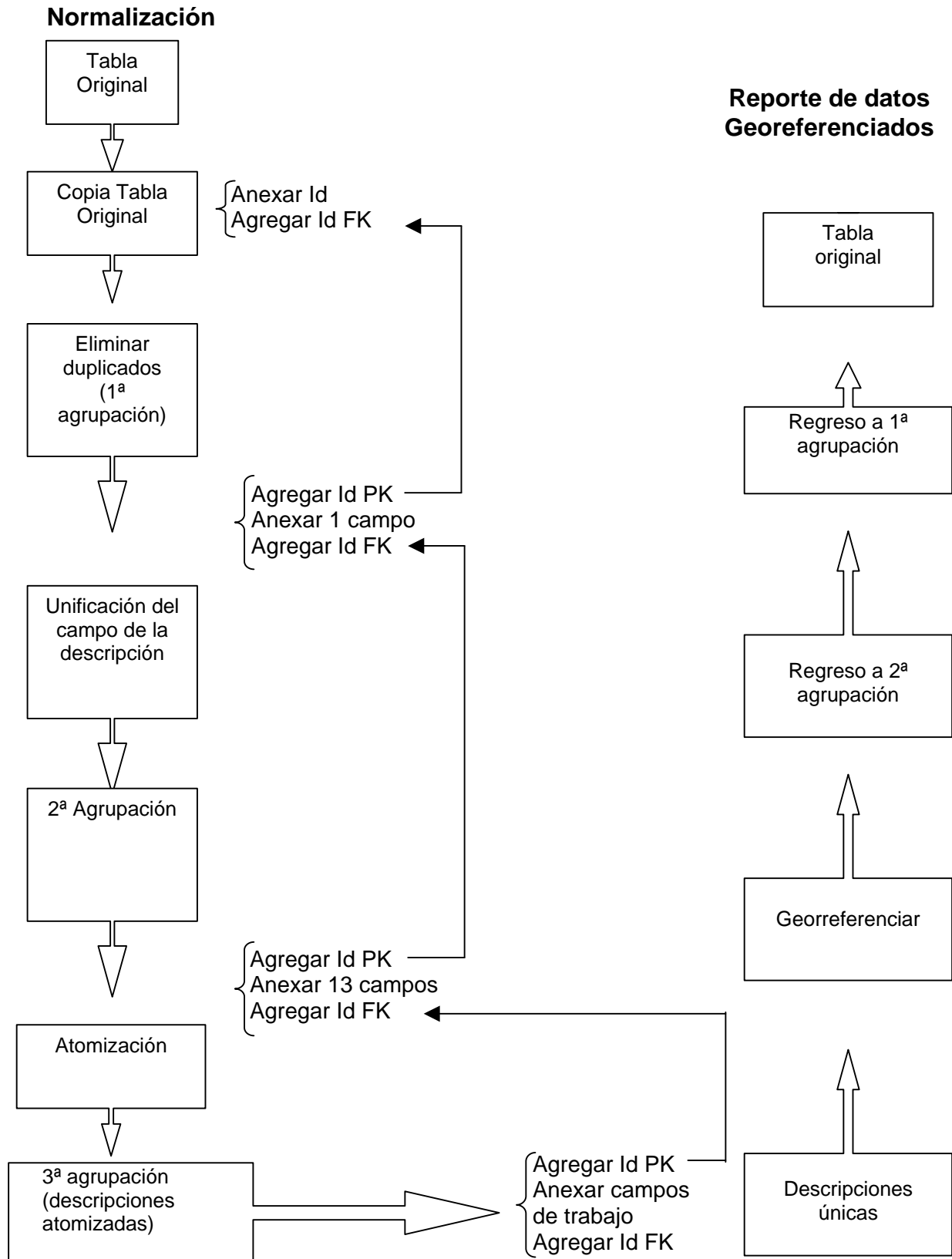


Figura 1. Diagrama de flujo del proceso de depuración de registros.