Tabla de Contenido

[Insumos 2](#_Toc117097752)

[Preparación de insumos 2](#_Toc117097753)

[Limpieza y depuración de datos 2](#_Toc117097754)

[Insumos alfanuméricos 2](#_Toc117097755)

[Insumos geográficos 3](#_Toc117097756)

[Normalización 4](#_Toc117097757)

[Implementación ETL 4](#_Toc117097758)

# Insumos

Los insumos implementados para el ejercicio fuero obtenidos en los portales web de Colombia en Mapas y Datos abiertos, el tipo de datos que se implementaran son de tipo alfanumérico y espacial, correspondientes a la zona rural del municipio de Girón del departamento de Santander, cada uno de los insumos fueron personalizados con el fin de poder cumplir el ejercicio de transferencia de información a modelo RIC (Registro de información catastral).

## Preparación de insumos

Cada uno de los archivos utilizado para la construcción del ejercicio, siendo el atributo conocido como como numero predial nacional, el cual es el código de identificación otorgado a cada uno de los predios del territorio nacional, el cual tiene una estructura de 30 dígitos e indica diferentes condiciones de cada una de los bienes inmuebles, este atributo permitirá entablar la relación tanto de los insumos alfanuméricos como geográficos.



Figura 1 Relación de insumos alfanuméricos y geográficos

## Limpieza y depuración de datos

Los diferentes insumos de información implementados para el ejercicio requieren un proceso preliminar de exploración, esto con el objetivo de identificar las diferentes tipos de datos y las posibles inconsistencias que estos puedan llegar a tener, posterior a esto se inicia un proceso de limpieza y depuración de los datos, esto con el fin de garantizar la correcta migración de la información a la estructura del modelo LADM.

### Insumos alfanuméricos

* **Espacios duplicados:** Se debe corregir aquellos registros los cuales, en sus atributos por error de tabulación, puedan tener presencia de espaciado doble, ya sea el inicio o al final de dato.
* **Estructura de atributos:** En algunos atributos el modelo define un estándar para su registro, un ejemplo de esto es el código predial nacional, el cual tiene una estructura de 30 dígitos, en el cual cada posición indica un dato asociado al predio, un caso que se identifica es la presencia de registros los cuales no corresponden a la zona de trabajo, esto debido a que en sus primeras 5 posiciones no corresponden al código “68307”, el cual es asociado al municipio de Girón del departamento de Santander, todos aquellos registros con código deben ser depurados.
* **Atributos no esenciales:** Se borra aquellos atributos entregados en el insumo, el cual su uso no se contempla en la estructura del modelo LADM, lo cual permite garantizar la calidad de la información manejada.
* **Corrección de caracteres:** En algunos atributos los cuales su estructura está definida por un cierto tipo de caracteres, se debe validar la presencia de caracteres que no violen este tipo de definición, ejemplo la presencia de caracteres alfabéticos en atributos como número de identificación y número predial.
* **Estado jurídico de la información**: Verificar el estado de la información jurídica asociada al predio, en el cual su estado de folio de matrícula estuviera catalogado como “Abierto”, de lo contrario se procede a borrar aquellos registros los cuales estén “Cerrados” dado que en la práctica se considera que el bien inmueble, jurídicamente no existe.
* **Caracteres especiales:** Se debe validar que los atributos de tipo texto no tengan caracteres especiales, los cuales van desde saltos de línea como signos de puntuación, la presencia de estos no permite generar de manera correcta del archivo XTF.

### Insumos geográficos

Los insumos con una componente espacial, presentan una serie de error de carácter topológico, los cuales no permitirán garantizar la generación del archivo XTF de manera exitosa, estos problemas son:

* **Se debe eliminar los nodos duplicados.**
* **Superposición de nodos de la misma capa.**

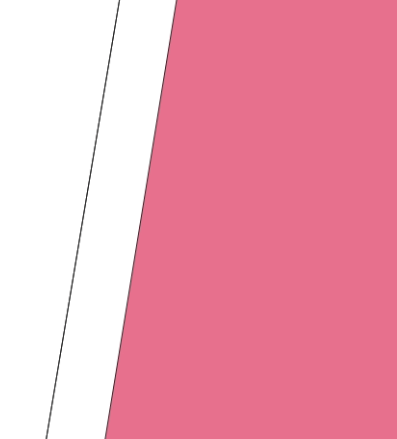


Figura 2 Problema topológico

## Normalización

Posterior a realizar el proceso de limpieza y depuración se debe iniciar con el proceso de normalización de los insumos entregados, en este se debe adaptar la información a la estructura definida por el modelo LADM, más específicamente en el modelo de registro información catastral, en este proceso se debe realizar las siguientes adaptaciones de la información:

* **Adaptación de dominios:** El modelo LADM en su estructura define una serie de dominios para algunos atributos, esto con el fin de generar un garantizar un control de la información que se ingresa, claramente muchos de los insumos bases entregados, no van a coincidir en primera instancia con los valores definidos por cada dominio, para esto es vital realizar la homologación de la información, para esto es necesario iniciar un proceso de exploración de los valores disponibles en los insumos y realizar el proceso de normalización, ejemplo de esto fue necesario adaptar atributos como tipo de documento de identificación, derecho asociado al predio, grupo étnico, tipo de fuente administrativa y estado civil
* **Sistema de proyección cartográfico:** Con todos los insumos los cuales sean de tipo geográfico, requiere realizar el proceso para que el sistema de proyección cartográfico, sea el adoptado en la resolución 471 de 2020 expedida por el Instituto Geográfico Agustín Codazzi, en el cual se establece un origen único el cual es denominado **Origen Nacional** con EPSG:9377.

## Implementación ETL

Finalmente, luego de realizar cada una de las anteriores etapas y preparar la información de cada uno de los insumos, los cuales se han estandarizado bajo la estructura del modelo LADM-COL, este proceso de migración se realiza bajo implementación del ETL diseñado, para esto se debe realizar la creación de una base de datos con ayuda del software libre PostgreSQL, esta base será administrada bajo el gestor Dbeaver el cual brindara un entorno más eficaz para la realización del proceso de migración, por último se emplea el software QGIS en el cual se manejara toda la información espacial y alfanumérica de la estructura del modelo RIC, por último se contempla un orden para la ejecución del ETL, el cual es el siguiente:

1. Creación del gestor y operador catastral.
2. Registro de la información asociada a la tabla predios y terrenos.
3. Registro de las direcciones asociadas a cada predio.
4. Registro de la información de interesados, además de la creación de las agrupaciones asociadas.
5. Registro de la información asociada de la tabla de derechos.
6. Registro de las fuentes administrativas.
7. Registro de los tramites catastrales asociadas a cada predio.

Por último, se procede a generar el archivo XTF el cual se debe exportar de manera exitosa.

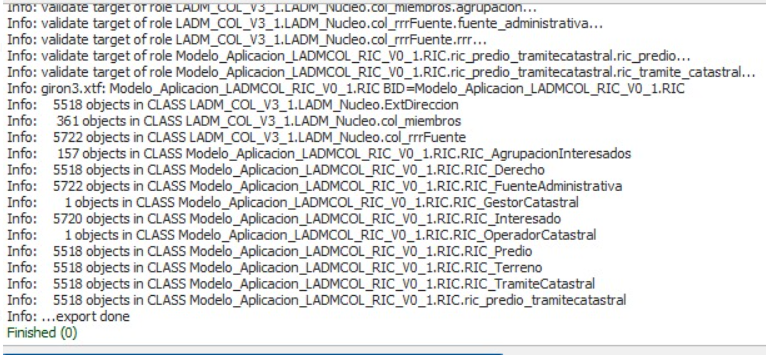


Figura 3 Generación de archivo XTF