# INFORME DE ESTUDIO DE DATA CLEANING DE LA BASE DE DATOS DE SIBICA

**ARQUITECSOFT**

**ÁREA DE DESARROLLO Y PROCESOS**

**CALI**

**2025**

| | | |
|---|---|---|
| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |

CONTENIDO

INFORME DE RESULTADOS DEL PROCESO DE LIMPIEZA DE DATOS ..... 6
1. REVISIÓN, LIMPIEZA Y VALIDACIÓN DE LA INFORMACIÓN CONTENIDA EN EL SIBICA. ..... 6
1.1. Objetivo del proceso de limpieza: ..... 6
1.2. Diagnóstico Inicial de Calidad de Datos ..... 12
1.2.1 Análisis preliminar - estado inicial de las variables ..... 12
1.3 Cruce de información y validación contra entregas a SIBICA ..... 26
1.3.1 Fuentes de Información y Flujo de Datos ..... 26
1.3.2 Proceso de Cruce y Validación Implementado ..... 27
1.3.3 Validación contra el Catálogo de Objetos Geográficos ..... 28
1.3.4 Conclusión del Proceso de Cruce ..... 28
1.4 Problemas Identificados ..... 29
2. ELIMINACIÓN DE REGISTROS DUPLICADOS, INCONSISTENTES O DESACTUALIZADOS. ..... 31
2.1 Justificación de la Estrategia Adoptada ..... 31
2.2 Estrategia Implementada ..... 31
2.3 Excepción: Limpieza de Tabla Matrícula ..... 32
2.4 Beneficios de la Estrategia de No Eliminación en Tabla Principal ..... 32
3. ESTANDARIZACIÓN DE DATOS CLAVE ASOCIADOS A LA GESTIÓN DE BIENES INMUEBLES. ..... 33
3.1 Tratamiento de Valores Nulos ..... 33
3.2 Corrección de Formatos ..... 33
3.2.1 Estandarización de Nombres de Áreas Cedidas ..... 34
3.2.2 Estandarización del Campo Tipo de Equipamiento ..... 35
3.2.3 Limpieza Automática de Espacios en Blanco en Campos de Texto ..... 36
3.2.4 Estandarización y Análisis Estructural de Códigos de Identificación Predial ..... 38
3.2.5 Corrección de Clasificación de Zonas Geográficas ..... 43
3.2.6 Normalización del Campo Matrícula Principal ..... 44
3.2.7 Estandarización del Campo de Actualización SAP ..... 45
3.2.8 Normalización Compleja del Campo Fecha de Levantamiento Topográfico ..... 46
3.2.9 Normalización del Campo Dirección y Dirección Catastral ..... 49

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno

3.3 Transformaciones Aplicadas ..... 55
4. GENERACIÓN DE REPORTES DE CALIDAD DE DATOS Y CONSOLIDACIÓN DE UNA BASE CONFIABLE PARA LA GESTIÓN INSTITUCIONAL ..... 55
4.1 Análisis Estadístico Cuantitativo de Completitud de Datos ..... 55
4.1.1. Análisis Interpretativo de Resultados: ..... 59
4.1.1.1 Campos Críticos Identificados (Completitud < 50%) ..... 59
4.1.1.2 Campos con Calidad Aceptable (Completitud entre 50% y 95%) ..... 59
4.1.1.3 Campos con Calidad Excelente (Completitud ≥ 95%) ..... 60
4.2 Comparación Datos Externos (Catastro, Rentas) ..... 61
4.3. Comparación entre Registros de Catastro y Construcción ..... 65
4.3.1 Resultados de la Vinculación de Datos ..... 65
4.3.2. Actualización Temporal de los Registros de Construcción ..... 66
5. RECOMENDACIONES ..... 68
5.1 Estrategia de Sincronización y Mejora de Sistemas Prediales ..... 68
5.2 Control de Calidad de Datos ..... 68
5.2.1 Implementación de Triggers para Integridad de Datos ..... 68
5.2.2. Importancia Crítica de Implementar Triggers ..... 69
5.2.3. Aplicación Específica en el Sistema SIBICA ..... 70
5.2.4 Validaciones Automáticas de Entrada ..... 72
5.2.5 Estandarización de Reglas de Captura en los Sistemas Origen ..... 73
5.2.6 Proceso de Verificación Post-Recopilación ..... 73
5.2.7 Auditorías y Monitoreo Continuo ..... 73
5.3 Automatización del Proceso ..... 74
5.3.1 Desarrollo de Scripts de Limpieza Periódica ..... 74
5.3.2 Implementación de Pipeline ETL ..... 74
5.3.3 Monitoreo Continuo de Calidad ..... 74
5.3.4 Versionado de Datos ..... 74
5.3.5 Estudio de Impacto y Mejora Continua ..... 75
6. Anexos ..... 76

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno

LISTADO DE TABLAS

Tabla 1. Identificadores únicos y códigos de referencia ..... 7
Tabla 2. Información geográfica ..... 8
Tabla 3. Clasificación jurídica y administrativa ..... 8
Tabla 4. Tipología y uso del predio ..... 9
Tabla 5. Información temporal y de trazabilidad ..... 9
Tabla 6. Documentación asociada ..... 10
Tabla 7. Relaciones con proyectos urbanísticos ..... 11
Tabla 8. problemas identificados durante el diagnóstico inicial. ..... 29
Tabla 9. Tipos de Inconsistencias Identificadas ..... 50
Tabla 10. reglas e implementaciones lógicas ..... 51
Tabla 11. Convenciones populares ..... 52
Tabla 12. Convenciones populares ..... 53
Tabla 13. Beneficios Cuantificables ..... 54
Tabla 14. Análisis atributos predios por atributo ..... 56
Tabla 15. Resumen Cuantitativo y Conclusiones ..... 60
Tabla 16. Diferencia real. (a – c) ..... 64

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

LISTADO DE FIGURAS

Figura 1. Inconsistencias en Códigos de Identificación (codigounico_p, codigonal_p) ..... 13
Figura 2. Correcta Categorización de Calidad de Predio (ID_CB_FK) ..... 13
Figura 3. Correcta Categorización de Calidad de Predio (CLASE_INMUEBLE_P) ..... 14
Figura 4. Variación Correcta Categorización de Tipo de Bien (ID_TB_FK) ..... 14
Figura 5. Correcta Categorización de Tipo de Uso (ID_TU_FK) ..... 15
Figura 6. Nomenclatura de Direcciones (direccion_p,direccioncatastro_p) ..... 16
Figura 7. Errores en Clasificación de Zonas (zona_p) ..... 16
Figura 8. Correcta Codificación de Ciudad (ciudad_p) ..... 17
Figura 9. Correcta Implementación del Campo Asegurado (asegurado_p) ..... 17
Figura 10. Inconsistencias en Códigos de Identificación (matricula_ppal_p) ..... 18
Figura 11. Correcta Implementación del Campo Matrícula Inmobiliaria (mat_inmob_p) ..... 18
Figura 12. Correcta Implementación del Campo Modo de Adquisición (id_madq_fk) ..... 19
Figura 13. Correcta Implementación del Campo Susceptible de Venta (suscep_vta_p) ..... 19
Figura 14. Correcta Implementación del Campo Dependencia Responsable (id_depen_fk) ..... 20
Figura 15. Inconsistencias en Tipo de Equipamiento (tipo_equipamiento) ..... 21
Figura 16. Correcta Implementación del Campo Derecho de Propiedad (derecho_p) ..... 21
Figura 17. Variaciones en Nombres de Áreas Cedidas (nombre_areacedida_p) ..... 22
Figura 18. Correcta Implementación del Campo Cedente del Predio (nit_cede_fk) ..... 22
Figura 19. Correcta Implementación del Campo Área en Títulos (area_titulos y area_lev_p) ..... 23
Figura 20. Múltiples Formatos de Fecha (fecha_levantamiento_p) ..... 23
Figura 21. Correcta Implementación del Campo Estado del Predio (id_estado_fk) ..... 24
Figura 22. Inconsistencias en Actualizar Sap (actualiza_sap) ..... 24
Figura 23. Inconsistencias en Registro de Linderos del Predio (lind_norte_p, lind_sur_p, lind_este_p, lind_oeste_p, lind_adic_p) ..... 25
Figura 24. Fechas de levantamiento de datos ..... 49
Figura 25. Porcentajes sobre un universo de 7.347 registros que tiene el SIBICA ..... 63
Figura 26. ID PREDIO_ID CATASTRO ..... 63
Figura 27. Porcentaje de coincidencias por método vs total construcciones ..... 66
Figura 28. Construcciones por año de registro (coincidencias) ..... 67

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

# INFORME DE RESULTADOS DEL PROCESO DE LIMPIEZA DE DATOS

## 1. REVISIÓN, LIMPIEZA Y VALIDACIÓN DE LA INFORMACIÓN CONTENIDA EN EL SIBICA.

### 1.1. Objetivo del proceso de limpieza:

Realizar la depuración y optimización de la información contenida en el Sistema de Información para la Gestión de Bienes Inmuebles del Distrito de Santiago de Cali (SIBICA), con el fin de garantizar la calidad, integridad y confiabilidad de los datos prediales. Este proceso busca eliminar inconsistencias, duplicados, registros desactualizados y estandarizar la información conforme a lineamientos técnicos, jurídicos y cartográficos del Distrito.

### Volumen inicial de registros:

La base de datos objeto de este proceso de limpieza y optimización está constituida por un total de 7,393 registros prediales activos, cada uno de los cuales contiene información estructurada en 48 variables o atributos diferenciados, lo que representa un volumen significativo de información que requiere procesos sistemáticos de validación y depuración.

Estos 48 atributos abarcan múltiples dimensiones de la información predial, estos 30 campos no tuvieron inconsistencias en sus valores, pero se realiza su respectivo análisis visual:

- id_catastro_p
- identidica_p
- proyecto_p
- id_cb_fk
- id_tb_fk
- clase_inmueble_p
- id_tb_fk
- id_tu_fk
- direccion_catastro_p
- id_barrio
- ciudad_p
- Asegurado_p
- Lind_norte_p
- Lind_sur_p
- Lind_este_p
- Lind:oeste_p
- Lind_adic_p
- Suscep_vta_p
- Id_depen_fk
- Derecho_p
- Nit_cede_fk
- Area_titulos
- Area_lev_p
- Is_shp_p

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- Id_estado_fk
- Id_proyecto_p
- Actualiza_sap
- Id_capa
- Area_planos
- Numero_suscriptor_sp
- Servicios_publicos

de los cuales se toman 18 campos para su respectivo análisis y se clasifican en las siguientes categorías funcionales:

## Tabla 1. Identificadores únicos y códigos de referencia

| Campo | Sujeto a análisis | Observación |
|---|---|---|
| ID_P | NO | Cumple con los requisitos |
| CODIGOUNICO_P | SI | Se identificaron inconsistencias en el formato: valores como 'NR', 'N R', 'N-R'. Se requiere estandarización a un único formato 'N/R'. |
| CODIGONAL_P | SI | Se identificaron inconsistencias en el formato: valores como 'NR', 'N R', 'N-R'. Se requiere estandarización a un único formato 'N/R'. |
| MAT_INMO_P | NO | Cumple con los requisitos |

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno

**Tabla 2. Información geográfica**

| Campo | Sujeto a análisis | Observación |
|---|---|---|
| DIRECCION_P | SI | Se encontraron múltiples formatos de nomenclatura vial: 'calle', 'calles', 'carrera', 'transversal'. Se recomienda estandarizar (Calle→CL, Carrera→KR, Transversal→TV, Diagonal→DG). |
| ZONA_P | SI | Inconsistencias identificadas: 'RUAL' (error ortográfico), 'EXP.URB' (no estándar). Según catálogo debe ser 'Urbano' o 'Rural'. Se requiere corrección: 'RUAL'→'Rural', 'EXP.URB'→'Urbana'. |
| FECHA_LEVANTAMIENTO_P | SI | Se identificaron múltiples formatos de fecha (yyyy-mm-dd, dd/mm/yyyy, m/d/yyyy, formato texto). Completitud crítica: 14.95%. Se implementó normalización en 6 pasos para unificar a formato yyyy-mm-dd. |

**Tabla 3. Clasificación jurídica y administrativa**

| Campo | Sujeto a análisis | Observación |
|---|---|---|
| MATRICULA_PPAL_P | SI | Valores 'NR' sin estandarizar. Completitud: 56.05%. Se requiere identificar significado del valor 'No' y normalizar 'NR' a formato estándar 'N/R'. |

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

| | | |
|---|---|---|
| ACTUALIZA_SAP | SI | Inconsistencias en valores: 'NO', 'No', 'no'. Se recomienda globalizar el parámetro a un único formato 'NO' en mayúsculas. |

**Tabla 4. Tipología y uso del predio**

| Campo | Sujeto a análisis | Observación |
|---|---|---|
| TIPO_EQUIPAMIENTO | SI | Múltiples variaciones identificadas: 'Bahia de Parqueo'/'Bahía de Parqueo', 'Vía Pública'/'vía pública'/'Vía públicas'/'vía', 'N/A'/'n/a'/'NA'/'na', 'OTRO'/'Otro'/'otro'. Completitud: 67.98%. Se requiere estandarización completa a formato Title Case y unificación de términos genéricos a 'N/A'. |

**Tabla 5. Información temporal y de trazabilidad**

| Campo | Sujeto a análisis | Observación |
|---|---|---|
| FECHA_LEVANTAMIENTO_P | SI | Formatos mixtos detectados: yyyy-mm-dd, dd/mm/yyyy, m/d/yyyy, formato mes-año, texto descriptivo. Completitud crítica: 14.95% (solo 1,105 de 7,393 registros). Requiere normalización urgente a formato estándar yyyy-mm-dd. |
| FECHA_ESTUDIO_TITULO_P | SI | Se recomienda estandarizar formato de fecha a yyyy-mm-dd para consistencia con otros campos temporales del sistema. |

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno

| | | |
|---|---|---|
| FECHA_CREACION_P | SI | Se recomienda estandarizar formato de fecha a yyyy-mm-dd para mantener uniformidad en todos los campos temporales. |
| FECHA_MODIFICA_P | SI | Completitud crítica: 18.45%. Se recomienda estandarizar formato de fecha a yyyy-mm-dd. Campo esencial para auditoría y trazabilidad de cambios. |
| FECHA_EXPEDICION_CB_P | SI | Completitud crítica: 19.50%. Se recomienda estandarizar formato de fecha a yyyy-mm-dd para seguimiento de conceptos emitidos. |

**Tabla 6. Documentación asociada**

| Campo | Sujeto a análisis | Observación |
|---|---|---|
| DOC_CALIDAD_BIEN | SI | Completitud crítica: 0.04% (solo 3 de 7,393 registros). Indica que prácticamente no existe documentación digitalizada de conceptos de calidad del bien. Requiere plan de digitalización urgente. |
| EXPE_INV_PREDIAL | SI | Completitud crítica: 0.50% (solo 37 de 7,393 registros). Ausencia casi total de expedientes digitalizados. Problema sistémico de documentación que requiere atención prioritaria. |

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno

**Tabla 7. Relaciones con proyectos urbanísticos**

| Campo | Sujeto a análisis | Observación |
|---|---|---|
| NOMBRE_AREACEDIDA_P | SI | Múltiples inconsistencias: variaciones de mayúsculas/minúsculas ('ZONA VERDE 1'/'Zona verde 1'/'zona verde 1'), espacios adicionales, errores ortográficos ('Equipamento'→'Equipamiento'). Completitud: 98.09%. Se implementó script de normalización a formato Title Case con eliminación de espacios y corrección ortográfica. |

El Catálogo de Objetos Geográficos de la UAEGBS establece de manera formal y detallada las especificaciones técnicas de cada uno de estos atributos, definiendo su tipo de dato esperado (entero, texto, fecha, decimal), su obligatoriedad, sus restricciones de dominio y su propósito funcional dentro del sistema. Este catálogo constituye el documento normativo de referencia para la validación de la calidad de datos y ha sido utilizado como base para la identificación de inconsistencias y desviaciones respecto a los estándares establecidos.

## Periodo de análisis:

El presente análisis de calidad de datos abarca la totalidad de la información histórica que ha sido registrada y acumulada en el Sistema de Información para la Gestión de Bienes Inmuebles del Distrito de Santiago de Cali (SIBICA) desde su puesta en funcionamiento inicial hasta el 16 de octubre del año 2025, fecha en la cual la Unidad Administrativa Especial de Gestión de Bienes y Servicios (UAEGBS) entregó oficialmente el conjunto completo de datos para dar inicio al proceso de optimización y modernización tecnológica contemplado en el alcance contractual.

Este periodo temporal comprende la operación continua del sistema, durante el cual se ha consolidado información crítica sobre el inventario predial del municipio, incluyendo procesos de identificación, levantamientos topográficos, estudios de títulos, conceptos de calidad de bien, cesiones urbanísticas y seguimiento administrativo de los bienes inmuebles del Distrito. El análisis retrospectivo permite identificar patrones de inconsistencia que se han generado a lo largo del tiempo debido a cambios en procedimientos internos, actualizaciones del sistema, rotación de personal y ausencia de controles automáticos de validación en la captura de datos.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

## 1.2. Diagnóstico Inicial de Calidad de Datos

### 1.2.1 Análisis preliminar - estado inicial de las variables

El proceso de diagnóstico de calidad de datos del sistema SIBICA se estructuró en dos fases complementarias que permitieron obtener una visión integral del estado de la información predial:

**Fase 1 - Análisis Visual de Inconsistencias:** Exploración directa de la base de datos mediante consultas SQL para identificar y documentar visualmente las principales problemáticas de formato, estandarización y completitud de datos.

**Fase 2 - Cruce y Validación de Información:** Análisis cruzado entre la base de datos SIBICA, los archivos de entregas pendientes de carga proporcionados por las áreas de Investigación Predial y Topografía, y el Catálogo de Objetos Geográficos oficial (sección 2.2).

Esta sección documenta la **Fase 1**, presentando mediante **evidencia visual** (capturas de pantalla de consultas SQL reales) el estado inicial de los datos antes de cualquier intervención. Estas evidencias establecen la **línea base** que justifica las acciones correctivas implementadas y facilita la comprensión de la magnitud de los problemas identificados.

#### 1.2.1.1 Metodología del Análisis Visual Exploratorio

El análisis preliminar se ejecutó mediante **consultas SQL directas** a la base de datos de producción del sistema SIBICA, extrayendo muestras representativas del conjunto de atributos que componen la tabla principal prediomunicipio de la base de datos de SIBICA. Las consultas fueron diseñadas específicamente para:

- Identificar valores únicos y su frecuencia de aparición en cada campo (usando SELECT DISTINCT y COUNT)
- Detectar variaciones en formatos de un mismo tipo de dato (fechas, códigos, categorías)
- Visualizar inconsistencias ortográficas, de capitalización y espaciado
- Evidenciar la presencia de espacios en blanco no visibles mediante funciones LENGTH() y TRIM()
- Cuantificar registros con valores nulos, vacíos o marcados como 'N/A'

Los resultados de estas consultas SQL fueron capturados en imágenes que se presentan a continuación. Posteriormente, estos hallazgos visuales fueron **validados y contextualizados** mediante el cruce con el Catálogo de Objetos Geográficos, proceso que se detalla en la sección 2.2.

#### 1.2.1.2 Evidencias Visuales de atributos e inconsistencias Identificadas

A continuación, se presentan las evidencias que documentan el estado inicial de los campos con mayores problemáticas de calidad de datos.

Observación: En las imágenes se identifican campos con los valores 'null' o 'SIN DATO', los cuales indican ausencia de información. Estos campos permanecen vacíos debido a que no se ha capturado o registrado dato alguno para esos atributos específicos.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

**Figura 1. Inconsistencias en Códigos de Identificación (codigounico_p, codigonal_p)**

| | codigonal_p character varying | total bigint |
|---|---|---|
| 1 | SIN DATO | 3817 |
| 2 | NR | 18 |
| 3 | 760010000650000030222000000000 | 5 |
| 4 | 0 | 5 |
| 5 | N/R | 5 |
| 6 | 760010100229600870028400000000 | 4 |
| 7 | 760010100181300040406400000000 | 3 |
| 8 | 760010000520000012499300000000 | 2 |
| 9 | 760010100139900140001300000001 | 2 |
| 10 | 760010100020700500030000000030 | 2 |
| 11 | 760010100181000110005000000005 | 2 |

**Descripción de la imagen:** La consulta SQL revela que el valor 'N/R' aparece almacenado de **al menos 2 formas diferentes** ('NR', 'N/R'), afectando búsquedas y agrupaciones.

**Figura 2. Correcta Categorización de Calidad de Predio (ID_CB_FK)**

| | ID_CB_FK | TOTAL |
|---|---|---|
| 1 | 1 | 7092 |
| 2 | 2 | 234 |
| 3 | SIN DATO | 67 |

**Descripción de la imagen:** El campo ID_CB_FK presenta una correcta implementación de las categorías establecidas. Se observan únicamente valores enteros (1, 2 y 3) que corresponden adecuadamente a las clasificaciones de calidad de predio definidas en el catálogo.

No obstante, se detectaron 67 registros sin información en este campo, ocasionando inconsistencias en la clasificación de calidad del bien predial. Dado que este campo es determinante para identificar si un predio es de uso público o fiscal, se adjuntará un documento con el listado de los 67 registros pendientes para su revisión y actualización correspondiente.

**Categorías válidas:**

1. Bien de uso público
2. Bien fiscal
3. Bien del particular

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: Uso Interno

**Figura 3. Correcta Categorización de Calidad de Predio (CLASE_INMUEBLE_P)**

| | CLASE_INMUEBLE_P | TOTAL |
|---|---|---|
| 1 | 2 | 6687 |
| 2 | 3 | 454 |
| 3 | SIN DATO | 246 |
| 4 | 1 | 6 |

**Descripción de la imagen:** El campo CLASE_INMUEBLE_P presenta una correcta implementación de las categorías establecidas. Se observan únicamente valores enteros (1, 2 y 3) que corresponden adecuadamente a las clasificaciones de calidad de predio definidas en el catálogo.

**Categorías validas:**

1. Sometido al Régimen de Propiedad Horizontal
2. Terreno
3. Terreno y construcción

**Figura 4. Variación Correcta Categorización de Tipo de Bien (ID_TB_FK)**

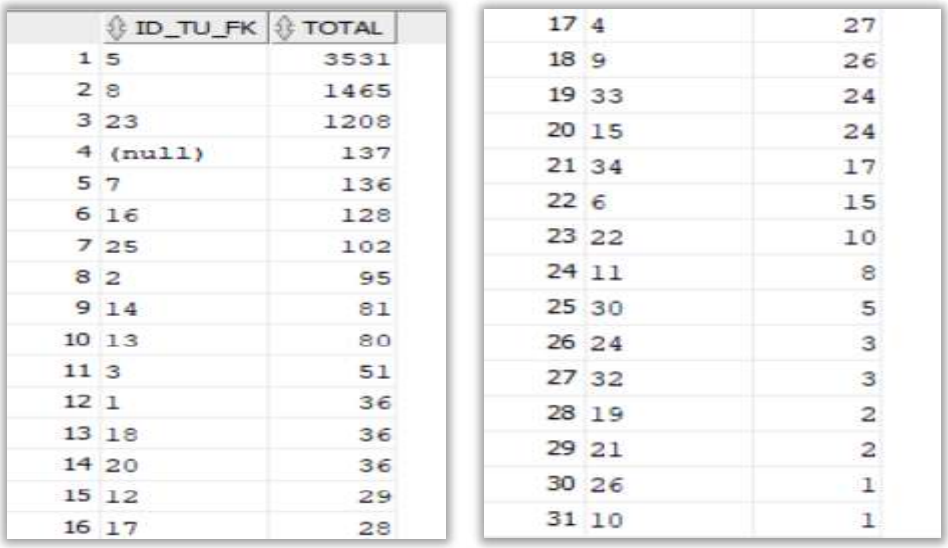
| | ID_TB_FK | TOTAL |
|---|---|---|
| 1 | 7 | 4374 |
| 2 | 3 | 1818 |
| 3 | 27 | 265 |
| 4 | 5 | 233 |
| 5 | 24 | 107 |
| 6 | 2 | 96 |
| 7 | 1 | 86 |
| 8 | 4 | 80 |
| 9 | 26 | 56 |
| 10 | 6 | 52 |
| 11 | 11 | 44 |
| 12 | 23 | 38 |
| 13 | 13 | 33 |
| 14 | 10 | 15 |
| 15 | 8 | 14 |
| 16 | (null) | 9 |
| 17 | 28 | 6 |
| 18 | 17 | 4 |
| 19 | 15 | 4 |
| 20 | 16 | 3 |
| 21 | 12 | 2 |
| 22 | 25 | 2 |
| 23 | 14 | 2 |
| 24 | 20 | 2 |
| 25 | 9 | 2 |

**Descripción de la imagen:** El campo ID_TB_FK presenta una correcta implementación del catálogo de tipos de bien. Se observa una tabla de referencia bien estructurada con identificadores enteros consecutivos (1-28) y sus correspondientes nombres descriptivos (Bahía Parqueo, Zona Verde, Lote, Separador vial, Vía Pública, Construcción, Apartamento, Edificio, Casa, Bodega, Andén, Equipamiento, Campamento, Casa

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno

Lote, Depósito, Espacio Público, Finca, Propiedad Horizontal, Garaje, Local Comercial, Mina, Obras de Infraestructura, Oficina, Parqueadero, Retal, Zona de Protección, Vía Peatonal, Otro).

**Figura 5. Correcta Categorización de Tipo de Uso (ID_TU_FK)**



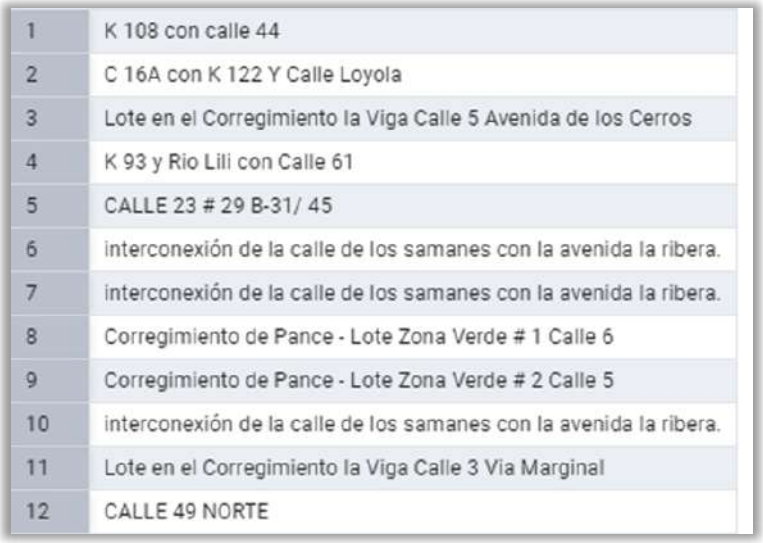
| | ID_TU_FK | TOTAL |
|---|---|---|
| 1 | 5 | 3531 |
| 2 | 8 | 1465 |
| 3 | 23 | 1208 |
| 4 | (null) | 137 |
| 5 | 7 | 136 |
| 6 | 16 | 128 |
| 7 | 25 | 102 |
| 8 | 2 | 95 |
| 9 | 14 | 81 |
| 10 | 13 | 80 |
| 11 | 3 | 51 |
| 12 | 1 | 36 |
| 13 | 18 | 36 |
| 14 | 20 | 36 |
| 15 | 12 | 29 |
| 16 | 17 | 28 |
| 17 | 4 | 27 |
| 18 | 9 | 26 |
| 19 | 33 | 24 |
| 20 | 15 | 24 |
| 21 | 34 | 17 |
| 22 | 6 | 15 |
| 23 | 22 | 10 |
| 24 | 11 | 8 |
| 25 | 30 | 5 |
| 26 | 24 | 3 |
| 27 | 32 | 3 |
| 28 | 19 | 2 |
| 29 | 21 | 2 |
| 30 | 26 | 1 |
| 31 | 10 | 1 |

**Descripción de la imagen:** El campo ID_TU_FK presenta una correcta implementación del catálogo de tipos de uso. Se observa una tabla de referencia bien estructurada con identificadores enteros (1-34) y sus correspondientes nombres descriptivos que abarcan la diversidad de usos que pueden tener los predios municipales, desde usos institucionales y educativos hasta infraestructura urbana y servicios públicos. **conteniendo 34 categorías desde Institucional hasta Otro, incluyendo usos como Educativo, Comercial, Cultural, Bienestar Social, Recreativa, Zona Verde, Ocupado por Tercero, Invadido, Alta Tensión, Equipamiento Colec, Zona Protección, Parroquia, Vía Pública, Parqueadero, Plazoleta, Andén, Lote, Garaje, Vía Peatonal, Oficina, Bahía Parqueo, Tertuliadero, Plaza de Mercado, Centro Recreativo, Estación de Policía, Salón Comunal, Bodega, Canal, Escombrera, Puesto de Salud, Seguridad, Separador Vial. Dentro de esas 34 categorías hay 4 categorías que no se usan en ningún registro las cuales son Plaza de Mercado, Centro Recreativo, Estación de Policía, Bodega.**

El campo cumple completamente con las especificaciones del catálogo, utilizando correctamente el tipo de dato entero como clave primaria y manteniendo una tabla de referencia normalizada que permite una clasificación completa y estandarizada de los diferentes usos que pueden tener los predios en el municipio.

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno

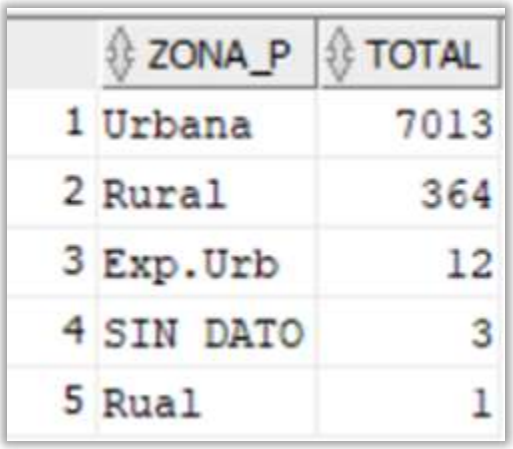
**Figura 6. Nomenclatura de Direcciones (direccion_p,direccioncatastro_p)**



| | |
|---|---|
| 1 | K 108 con calle 44 |
| 2 | C 16A con K 122 Y Calle Loyola |
| 3 | Lote en el Corregimiento la Viga Calle 5 Avenida de los Cerros |
| 4 | K 93 y Rio Lili con Calle 61 |
| 5 | CALLE 23 # 29 B-31/ 45 |
| 6 | interconexión de la calle de los samanes con la avenida la ribera. |
| 7 | interconexión de la calle de los samanes con la avenida la ribera. |
| 8 | Corregimiento de Pance - Lote Zona Verde # 1 Calle 6 |
| 9 | Corregimiento de Pance - Lote Zona Verde # 2 Calle 5 |
| 10 | interconexión de la calle de los samanes con la avenida la ribera. |
| 11 | Lote en el Corregimiento la Viga Calle 3 Via Marginal |
| 12 | CALLE 49 NORTE |

**Descripción de la imagen:** Falta total de estandarización en la nomenclatura de direcciones. Se observan múltiples inconsistencias: uso de abreviaturas ('K', 'C') mezcladas con palabras completas ('Calle', 'calle'); variaciones significativas de capitalización ('CALLE', 'Calle', 'calle'); combinación de formatos descriptivos extensos con nomenclaturas simples; uso inconsistente de conectores ('con', 'y', '#'); y mezcla de direcciones urbanas estándar con descripciones de ubicación por corregimiento. Algunos registros incluyen información adicional como tipo de predio o número de lote dentro del campo de dirección.

**Figura 7. Errores en Clasificación de Zonas (zona_p)**



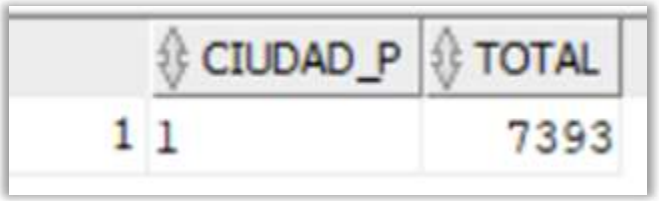
| | ZONA_P | TOTAL |
|---|---|---|
| 1 | Urbana | 7013 |
| 2 | Rural | 364 |
| 3 | Exp.Urb | 12 |
| 4 | SIN DATO | 3 |
| 5 | Rual | 1 |

**Descripción de la imagen:** Error ortográfico crítico: 'RUAL' en lugar de 'Rural', más el valor no estándar 'EXP.URB' (expansión urbana). **Validación contra catálogo:** El Catálogo de Objetos Geográficos

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

establece que este campo SOLO debe contener 'Urbano' o 'Rural'. Los valores 'RUAL' y 'EXP.URB' violan esta especificación y deben corregirse.

**Figura 8. Correcta Codificación de Ciudad (ciudad_p)**

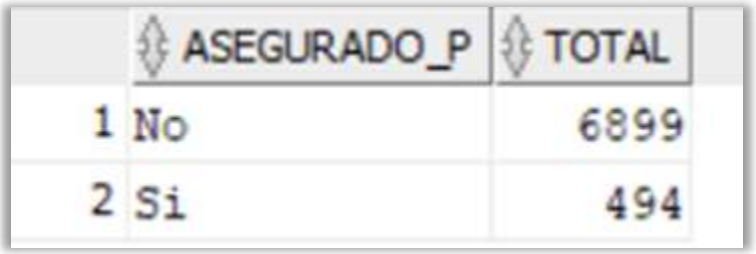


| | CIUDAD_P | TOTAL |
|---|---|---|
| 1 | 1 | 7393 |

**Descripción de la imagen:** El campo ciudad_p presenta una correcta implementación del código de ciudad. Se observa que todos los registros (7393 totales) corresponden al valor "1", que según la tabla de referencia representa correctamente a Cali. La segunda imagen muestra una tabla de referencia bien estructurada con códigos de ciudad (cod_ciu), nombres de municipios (nombre_ciu) y códigos de departamento (cod_dep), donde el código "1" está asociado a Cali, lo cual es consistente con el hecho de que todos los predios gestionados por la UAEGBS de Cali se encuentran dentro de este municipio.

El campo cumple completamente con las especificaciones del catálogo, utilizando correctamente el tipo de dato entero y manteniendo consistencia con la tabla de referencia de ciudades. La uniformidad del valor "1" (Cali) en todos los registros es coherente con el ámbito territorial de competencia de la entidad.

**Figura 9. Correcta Implementación del Campo Asegurado (asegurado_p)**



| | ASEGURADO_P | TOTAL |
|---|---|---|
| 1 | No | 6899 |
| 2 | Si | 494 |

**Descripción de la imagen:** El campo asegurado_p presenta una correcta implementación del indicador de aseguramiento de bienes inmuebles. Se observa que los registros contienen valores de tipo texto que indican claramente si cada predio cuenta o no con cobertura de seguro, permitiendo un control adecuado sobre el estado de protección patrimonial de los bienes.

El campo cumple completamente con las especificaciones del catálogo, utilizando correctamente el tipo de dato texto para almacenar la información sobre el estado de aseguramiento de cada predio, lo cual permite a la entidad mantener un control efectivo sobre la protección de su patrimonio inmobiliario.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

**Figura 10. Inconsistencias en Códigos de Identificación (matricula_ppal_p)**

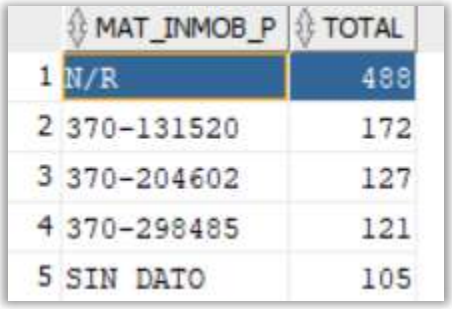


| | MATRICULA_PPAL_P | TOTAL |
|---|---|---|
| 1 | SIN DATO | 2063 |
| 2 | N/R | 1178 |

| | | |
|---|---|---|
| 119 | No | 5 |
| 120 | NR | 5 |

**Descripción de la imagen:** La consulta SQL revela que el valor 'No Registra' aparece almacenado de **al menos 2 formas diferentes** ('NR', 'N/R'), afectando búsquedas y agrupaciones.

**Figura 11. Correcta Implementación del Campo Matrícula Inmobiliaria (mat_inmob_p)**



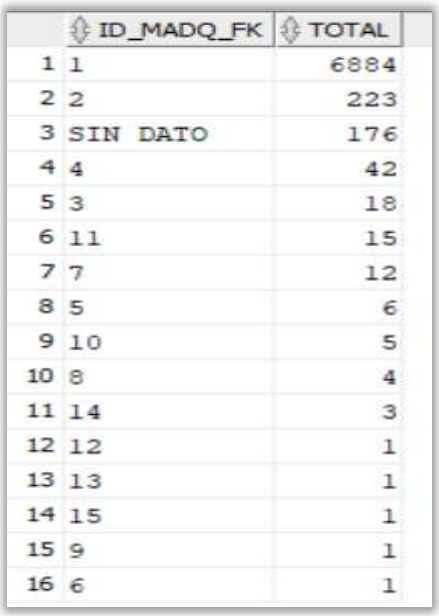
| | MAT_INMOB_P | TOTAL |
|---|---|---|
| 1 | N/R | 488 |
| 2 | 370-131520 | 172 |
| 3 | 370-204602 | 127 |
| 4 | 370-298485 | 121 |
| 5 | SIN DATO | 105 |

**Descripción de la imagen:** El campo mat_inmob_p presenta una correcta implementación del número de matrícula inmobiliaria. Se observa que los registros contienen valores de tipo texto que corresponden a los números de registro asignados a los certificados de tradición, permitiendo la identificación única de cada bien inmueble y su trazabilidad jurídica ante las oficinas de registro.

El campo cumple completamente con las especificaciones del catálogo, utilizando correctamente el tipo de dato texto para almacenar los números de matrícula inmobiliaria, lo cual garantiza la trazabilidad legal y permite verificar la situación jurídica de cada predio en el sistema registral nacional.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

**Figura 12. Correcta Implementación del Campo Modo de Adquisición (id_madq_fk)**



| | ID_MADQ_FK | TOTAL |
|---|---|---|
| 1 | 1 | 6884 |
| 2 | 2 | 223 |
| 3 | SIN DATO | 176 |
| 4 | 4 | 42 |
| 5 | 3 | 18 |
| 6 | 11 | 15 |
| 7 | 7 | 12 |
| 8 | 5 | 6 |
| 9 | 10 | 5 |
| 10 | 8 | 4 |
| 11 | 14 | 3 |
| 12 | 12 | 1 |
| 13 | 13 | 1 |
| 14 | 15 | 1 |
| 15 | 9 | 1 |
| 16 | 6 | 1 |

**Descripción de la imagen:** El campo id_madq_fk presenta una correcta implementación del catálogo de modos de adquisición. Se observa que los registros contienen valores de tipo entero que hacen referencia a una tabla normalizada con las diferentes figuras jurídicas mediante las cuales la entidad ha adquirido sus bienes inmuebles, permitiendo un control adecuado sobre el origen legal del patrimonio.

El campo cumple completamente con las especificaciones del catálogo, utilizando correctamente el tipo de dato entero como clave foránea que referencia una tabla de modos de adquisición, lo cual permite clasificar y analizar sistemáticamente el origen jurídico de los bienes inmuebles que conforman el patrimonio institucional.

**Figura 13. Correcta Implementación del Campo Susceptible de Venta (suscep_vta_p)**



| | SUSCEP_VTA_P | TOTAL |
|---|---|---|
| 1 | No | 7378 |
| 2 | Si | 8 |
| 3 | SIN DATO | 7 |

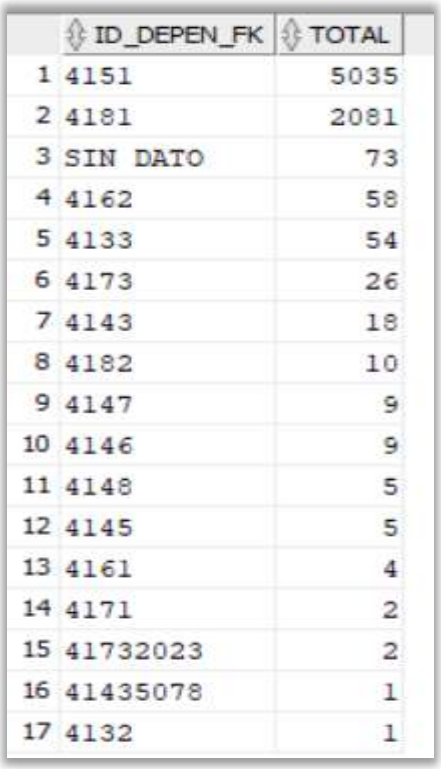
**Descripción de la imagen:** El campo suscep_vta_p presenta una correcta implementación del indicador de susceptibilidad de venta. Se observa que los registros contienen valores de tipo texto que establecen claramente si cada predio puede ser objeto de enajenación por parte de la entidad, permitiendo identificar

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

qué bienes están disponibles para procesos de venta o deben permanecer en el patrimonio institucional por su naturaleza o destinación.

El campo cumple completamente con las especificaciones del catálogo, utilizando correctamente el tipo de dato texto para registrar la condición de susceptibilidad de venta de cada predio, lo cual facilita la gestión patrimonial y la toma de decisiones sobre procesos de enajenación conforme al marco legal vigente.

**Figura 14. Correcta Implementación del Campo Dependencia Responsable (id_depen_fk)**



| | ID_DEPEN_FK | TOTAL |
|---|---|---|
| 1 | 4151 | 5035 |
| 2 | 4181 | 2081 |
| 3 | SIN DATO | 73 |
| 4 | 4162 | 58 |
| 5 | 4133 | 54 |
| 6 | 4173 | 26 |
| 7 | 4143 | 18 |
| 8 | 4182 | 10 |
| 9 | 4147 | 9 |
| 10 | 4146 | 9 |
| 11 | 4148 | 5 |
| 12 | 4145 | 5 |
| 13 | 4161 | 4 |
| 14 | 4171 | 2 |
| 15 | 41732023 | 2 |
| 16 | 41435078 | 1 |
| 17 | 4132 | 1 |

**Descripción de la imagen:** El campo id_depen_fk presenta una correcta implementación del catálogo de dependencias responsables. Se observa que los registros contienen valores de tipo entero que hacen referencia a una tabla normalizada con los diferentes organismos de la estructura administrativa municipal que tienen a su cargo la gestión y administración de los bienes inmuebles, permitiendo una clara asignación de responsabilidades sobre el patrimonio.

El campo cumple completamente con las especificaciones del catálogo, utilizando correctamente el tipo de dato entero como clave foránea que referencia una tabla de dependencias, lo cual permite establecer claramente la cadena de custodia administrativa de cada bien inmueble y facilita la gestión coordinada del patrimonio municipal entre las diferentes entidades.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

**Figura 15. Inconsistencias en Tipo de Equipamiento (tipo_equipamiento)**

| | tipo_equipamiento character varying | total bigint |
|---|---|---|
| 1 | N/A | 2166 |
| 2 | Vía Pública | 1273 |
| 3 | Zona Verde | 1181 |
| 4 | Vía Peatonal | 686 |
| 5 | Via Publica | 244 |
| 6 | SIN DATO | 213 |
| 28 | OTRO | 13 |
| 29 | Polideportivo La Esper... | 13 |
| 30 | Zona Verde | 13 |
| 31 | NA | 10 |
| 32 | N/R | 9 |
| 33 | Lote | 8 |
| 34 | Otro | 7 |

**Descripción de la imagen:** Múltiples problemas: (1) Con/sin tilde, (2) Mayúsculas/minúsculas, (3) Singular/plural, (4) Variaciones de 'N/A'. **Relación con catálogo:** Existe un catálogo oficial de tipos de equipamiento en las tablas de dominio del sistema. La hoja 'Variables' del archivo de entregas contiene este catálogo, confirmando que los valores válidos están definidos, pero no se están validando en la captura.

**Figura 16. Correcta Implementación del Campo Derecho de Propiedad (derecho_p)**

| | DERECHO_P | TOTAL |
|---|---|---|
| 1 | 100 | 7247 |
| 2 | SIN DATO | 140 |
| 3 | 2.31 | 1 |
| 4 | 50 | 1 |
| 5 | 98 | 1 |
| 6 | 99 | 1 |
| 7 | 0 | 1 |
| 8 | 10.33 | 1 |

**Descripción de la imagen:** El campo derecho_p presenta una correcta implementación del porcentaje de derecho de propiedad municipal. Se observa que los registros contienen valores de tipo decimal que expresan numéricamente la proporción de propiedad que ostenta el municipio sobre cada bien inmueble, permitiendo identificar predios con propiedad total (100%), copropiedad con otras entidades o particulares (valores intermedios), o situaciones especiales de participación patrimonial.

El campo cumple completamente con las especificaciones del catálogo, utilizando correctamente el tipo de dato decimal para registrar con precisión el porcentaje de derecho de propiedad municipal, lo cual permite valorar adecuadamente el patrimonio, gestionar situaciones de copropiedad y tomar decisiones informadas sobre la administración de los bienes inmuebles.

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno

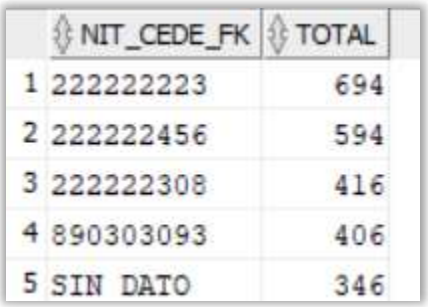
**Figura 17. Variaciones en Nombres de Áreas Cedidas (nombre_areacedida_p)**



| | nombre_areacedida_p<br>character varying (90) |
|---|---|
| 49 | Zona Verde 6 |
| 50 | Etapa 1 Zona verde Ep 04 |
| 51 | Zona Verde J 14 |
| 52 | ZONA VERDE Y-5 |
| 53 | Zona Verde |
| 54 | Zona Verde B |
| 55 | Zona Verde 4 |
| 56 | Polígono 4-Destinado A Zona Verde |
| 57 | Zona Verde N° 2 |
| 58 | Zona verde pata de jarillon C 84 |
| 59 | Zona Verde O1 |
| 60 | Zona Verde B |
| 61 | Zona Verde Adicional-Zona Verde V°(COLECTOR) |
| 62 | Zona Verde 21 |
| 63 | Zona Verde II |
| 64 | Zona Verde 10 |
| 65 | Zona Verde 36A |
| 66 | Zona Verde 5A |
| 67 | Zona Verde Manzana N(Prima). |
| 68 | Zona verde |

**Descripción de la imagen:** Tres problemas simultáneos: (1) Capitalización variable, (2) Espacios múltiples invisibles, (3) Errores ortográficos ('Equipamento'). **Impacto en reportes:** El cruce de datos reveló que al intentar consolidar áreas cedidas por tipo, 'ZONA VERDE 1', 'Zona verde 1' y 'zona verde 1' aparecen como tres categorías diferentes, multiplicando artificialmente el número de tipos de cesión.

**Figura 18. Correcta Implementación del Campo Cedente del Predio (nit_cede_fk)**



| | NIT_CEDE_FK | TOTAL |
|---|---|---|
| 1 | 222222223 | 694 |
| 2 | 222222456 | 594 |
| 3 | 222222308 | 416 |
| 4 | 890303093 | 406 |
| 5 | SIN DATO | 346 |

**Descripción de la imagen:** El campo nit_cede_fk presenta una correcta implementación del catálogo de cedentes. Se observa que los registros contienen valores de tipo entero que hacen referencia a una tabla normalizada con las diferentes personas naturales, jurídicas o entidades que han cedido predios al municipio, permitiendo mantener la trazabilidad histórica sobre el origen de cada bien inmueble y su proceso de transferencia.

El campo cumple completamente con las especificaciones del catálogo, utilizando correctamente el tipo de dato entero como clave foránea que referencia una tabla de cedentes, lo cual permite documentar

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

adecuadamente la cadena de titularidad de los bienes inmuebles y facilita la consulta histórica sobre las personas o entidades que han transferido predios al patrimonio municipal.

**Figura 19. Correcta Implementación del Campo Área en Títulos (area_titulos y area_lev_p)**

| | area_titulos text | total bigint |
|---|---|---|
| 1 | SIN DATO | 978 |
| 2 | 0 | 116 |
| 3 | 75 | 60 |
| 4 | 24 | 42 |
| 5 | 400 | 24 |
| 6 | 488.25 | 22 |

| | area_lev_p text | total bigint |
|---|---|---|
| 1 | SIN DATO | 2952 |
| 2 | 0 | 55 |
| 3 | 24 | 42 |
| 4 | 400 | 24 |
| 5 | 488.25 | 21 |
| 6 | 378 | 16 |

**Descripción de la imagen:** Los campos area_titulos y area_lev_p presentan una correcta implementación del registro de áreas cedidas por obligación urbanística. Se observa que los registros contienen valores numéricos de tipo double que expresan con precisión las medidas de superficie consignadas en los documentos de escritura pública, utilizando el punto (.) como separador decimal estándar. Esto permite verificar las áreas que fueron transferidas al municipio en cumplimiento de las cargas urbanísticas establecidas en los procesos de desarrollo urbano.

El campo cumple completamente con las especificaciones del catálogo, utilizando correctamente el tipo de dato double para registrar con alta precisión las medidas de superficie documentadas en los títulos de propiedad, lo cual permite contrastar el área jurídica con el área calculada y verificar la concordancia entre la documentación legal y la realidad física de los predios.

**Figura 20. Múltiples Formatos de Fecha (fecha_levantamiento_p)**

| | FECHA_LEVANTAMIENTO_P | TOTAL |
|---|---|---|
| 1 | N/R | 4197 |
| 2 | Sin lev. | 924 |
| 3 | SIN DATO | 874 |
| 4 | sin lev. | 152 |
| 5 | sin lev | 136 |
| 6 | dic-14 | 130 |
| 7 | dic-15 | 85 |

| | | |
|---|---|---|
| 101 | 2012-12-01 | 2 |
| 102 | Oct/2012 | 2 |
| 103 | jun/2006 | 2 |
| 104 | Jul/2012 | 2 |
| 105 | abr/2013 | 2 |
| 106 | Feb/2012 | 2 |
| 107 | mar/2021 | 2 |
| 108 | Jul/2009 | 2 |
| 109 | SIN LEV. | 2 |

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: Uso Interno

**Descripción de la imagen:** Se identificaron **5 formatos diferentes de fecha**: (1) yyyy-mm-dd, (2) dd/mm/yyyy, (3) mm/dd/yyyy, (4) 'mm-yyyy', (5) Texto descriptivo. Solo 14.95% de los 7,393 registros tienen fecha.

### Figura 21. Correcta Implementación del Campo Estado del Predio (id_estado_fk)

| | ID_ESTADO_FK | TOTAL |
|---|---|---|
| 1 | 3 | 7378 |
| 2 | SIN DATO | 9 |
| 3 | 2 | 5 |
| 4 | 1 | 1 |

**Descripción de la imagen:** El campo id_estado_fk presenta una correcta implementación del catálogo de estados de predios. Se observa que los registros contienen valores de tipo entero que hacen referencia a una tabla normalizada con las diferentes categorías de estado que puede tener un bien inmueble dentro del sistema de gestión patrimonial, permitiendo identificar claramente si un predio se encuentra activo en el inventario, inactivo temporalmente, bloqueado por situaciones jurídicas, retirado del patrimonio, o en cualquier otra condición relevante para su administración.

El campo cumple completamente con las especificaciones del catálogo, utilizando correctamente el tipo de dato entero como clave foránea que referencia una tabla de estados, lo cual permite gestionar el ciclo de vida de los bienes inmuebles, controlar su disponibilidad y mantener actualizado el inventario patrimonial según las diferentes situaciones administrativas y jurídicas que pueden presentarse.

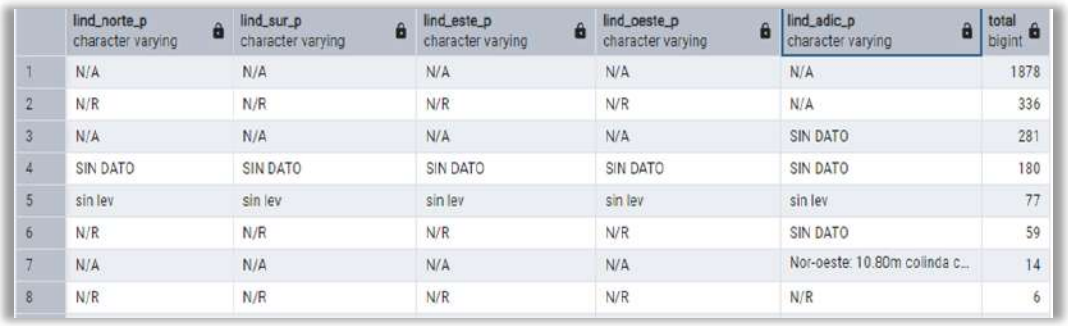
### Figura 22. Inconsistencias en Actualizar Sap (actualiza_sap)

| actualiza_sap character varying | total bigint |
|---|---|
| NO | 6714 |
| SI | 433 |
| No | 246 |

**Descripción de la imagen:** La consulta SQL revela que el valor 'NO' aparece almacenado de **al menos 2 formas diferentes** ('NO', 'No''), afectando búsquedas y agrupaciones.

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: Uso Interno

**Figura 23. Inconsistencias en Registro de Linderos del Predio (lind_norte_p, lind_sur_p, lind_este_p, lind_oeste_p, lind_adic_p)**



| | lind_norte_p character varying | lind_sur_p character varying | lind_este_p character varying | lind_oeste_p character varying | lind_adic_p character varying | total bigint |
|---|---|---|---|---|---|---|
| 1 | N/A | N/A | N/A | N/A | N/A | 1878 |
| 2 | N/R | N/R | N/R | N/R | N/A | 336 |
| 3 | N/A | N/A | N/A | N/A | SIN DATO | 281 |
| 4 | SIN DATO | SIN DATO | SIN DATO | SIN DATO | SIN DATO | 180 |
| 5 | sin lev | sin lev | sin lev | sin lev | sin lev | 77 |
| 6 | N/R | N/R | N/R | N/R | SIN DATO | 59 |
| 7 | N/A | N/A | N/A | N/A | Nor-oeste: 10.80m colinda c... | 14 |
| 8 | N/R | N/R | N/R | N/R | N/R | 6 |

**Descripción de la imagen:** Los campos de linderos del predio presentan inconsistencias críticas tanto en el tipo de dato como en la estandarización de valores. Se observa que estos campos, que según especificación deben almacenar valores enteros, contienen datos de tipo texto para indicar la ausencia de información. Específicamente, el valor 'N/R – N/A' aparece almacenado de al menos 4 formas diferentes ('N/A ', 'N/R ', 'NR', 'N/R'), lo cual afecta negativamente las búsquedas, agrupaciones y análisis de datos. Esta variabilidad impide la correcta validación de tipos de datos y dificulta el procesamiento automatizado de la información de linderos.

**Validación contra catálogo:** Según el Catálogo de Objetos Geográficos de la Unidad Administrativa Especial de Servicios Públicos – UAEGBS, los atributos de linderos deben cumplir las siguientes especificaciones:

- **lind_norte_p:** Límite norte del predio - Tipo de dato: Entero
- **lind_sur_p:** Límite sur del predio - Tipo de dato: Entero
- **lind_este_p:** Límite este del predio - Tipo de dato: Entero
- **lind_oeste_p:** Límite oeste del predio - Tipo de dato: Entero

- **lind_adic_p:** Describe los linderos adicionales que figuren registrados en la escritura de cesión - Tipo de dato: Entero

Los campos presentan dos problemas fundamentales: (1) **Violación del tipo de dato:** almacenan valores de texto cuando deberían contener exclusivamente valores enteros, y (2) **Falta de estandarización:** el concepto de "No Registra" aparece con múltiples variantes textuales sin un criterio unificado.

### 1.2.1.3 Conclusiones del Análisis Visual

El análisis visual preliminar documentado mediante las imágenes presentadas permitió establecer que:

**1.** Los problemas de calidad son **sistemáticos y generalizados**, no excepcionales, afectando múltiples campos y registros

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

**2.** La **ausencia de validaciones obligatorias** en SIBICA permite captura sin restricciones, generando alta variabilidad

**3.** Se requieren **correcciones masivas mediante scripts SQL** (corrección manual de 7,393 × 48 = 354,864 valores sería inviable)

**4.** Las inconsistencias visualizadas impactan la operación diaria, dificultando búsquedas, reportes y análisis

**Sin embargo, estas evidencias visuales por sí solas no son suficientes para diseñar una estrategia de limpieza efectiva.** Era necesario validar estos hallazgos contra fuentes externas para:

- Confirmar que las inconsistencias no son solo históricas, sino que continúan en datos nuevos
- Verificar si los valores estándar esperados están definidos en algún catálogo oficial
- Identificar el origen de los problemas (captura manual, copy-paste, falta de validaciones)
- Establecer si la estructura del modelo de datos (tablas, relaciones, llaves foráneas) es correcta

Por esta razón, la **fase siguiente del diagnóstico** consistió en realizar un **cruce exhaustivo de información** entre la base de datos SIBICA, los archivos de entregas pendientes de carga del repositorio 'Entregas a SIBICA' de Investigación Predial y Topografía (archivo 'Carga_registros_SIBICA.xlsx' con 1,591 registros), y el Catálogo de Objetos Geográficos oficial de la UAEGBS. Este proceso de validación cruzada se describe detalladamente en la **sección 1.3** a continuación.

## 1.3 Cruce de información y validación contra entregas a SIBICA

El proceso de validación de la calidad de datos del sistema SIBICA se fundamentó en el análisis cruzado de múltiples fuentes de información que convergen en la gestión predial del municipio. Este análisis permitió identificar no solo inconsistencias en los datos existentes, sino también establecer un mecanismo de verificación continua de la información que ingresa al sistema.

### 1.3.1 Fuentes de Información y Flujo de Datos

El Sistema de Información para la Gestión de Bienes Inmuebles del Distrito de Santiago de Cali (SIBICA) se alimenta principalmente de dos áreas técnicas que generan información primaria sobre los predios municipales:

### a) Área de Investigación Predial

Responsable de la investigación jurídica de los bienes inmuebles, incluyendo estudios de títulos, análisis de documentos, verificación de matrículas inmobiliarias y determinación de la calidad del bien. Esta área genera documentación crítica que soporta la situación legal de cada predio.

### b) Área de Topografía

Encargada de los levantamientos topográficos, mediciones de campo, determinación de áreas físicas, establecimiento de linderos y generación de planimetría técnica de los predios. Produce información georreferenciada y métricas precisas de las características físicas de los inmuebles.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

Ambas áreas mantienen un repositorio compartido denominado **"Entregas a SIBICA"**, que funciona como zona de intercambio de información pendiente de carga o actualización en el sistema principal. Este repositorio se estructura en archivos Excel que contienen:

- Registros nuevos de predios identificados que deben ser incorporados al inventario oficial
- Actualizaciones de información existente (correcciones de áreas, modificaciones de linderos, cambios de situación jurídica)
- Documentación de soporte (referencias a escrituras, matrículas inmobiliarias, contratos)
- Relaciones entre predios y proyectos urbanísticos
- Información complementaria de variables administrativas y técnicas

### 1.3.2 Proceso de Cruce y Validación Implementado

El proceso de limpieza de datos incorporó un análisis exhaustivo del archivo **"Carga_registros_SIBICA.xlsx"** proporcionado por las áreas de Investigación Predial y Topografía. Este archivo contiene **1,591 registros** en su hoja principal, distribuidos en las siguientes secciones:

- **Hoja 'REGISTROS':** Información completa de predios pendientes de carga, con 148 columnas que abarcan todos los atributos del modelo de datos SIBICA
- **Hoja 'Matriculas':** Registro de 785 matrículas inmobiliarias con sus respectivos datos (número, fecha, ciudad, notaría)
- **Hoja 'Escrituras':** Documentación de 194 escrituras públicas relacionadas con los predios
- **Hoja 'Relacionar EP':** 2,476 relaciones entre predios y escrituras públicas que deben ser establecidas en el sistema
- **Hoja 'Variables':** Catálogo de valores válidos para campos categóricos (calidad de bien, clase de inmueble, tipo de bien, tipo de uso, barrio, tipo de contrato, etc.)

### El proceso de validación consistió en:

**1. Verificación de existencia de registros:** Confirmar que los predios contenidos en el archivo de entregas ya se encontraban registrados en SIBICA con su información básica

**2. Validación de consistencia de datos:** Comparar los valores de los atributos entre el archivo de entregas y el sistema SIBICA para identificar discrepancias

**3. Detección de información faltante:** Identificar campos que aparecen completos en las entregas pero que están vacíos o nulos en SIBICA

**4. Análisis de calidad de información nueva:** Evaluar si los datos pendientes de carga cumplen con los estándares de formato y completitud establecidos

**Este cruce de información permitió evidenciar que:**

- Existe un flujo continuo de información entre las áreas técnicas y SIBICA que no siempre se completa de manera oportuna
- Algunos predios tienen información actualizada en los archivos de entrega que aún no se ha incorporado al sistema oficial

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- Las inconsistencias de formato identificadas en SIBICA también se replican en los archivos de entrega, lo que indica la necesidad de validaciones en el origen
- El catálogo de variables contenido en la hoja 'Variables' del archivo de entregas coincide con las tablas de dominio definidas en el Catálogo de Objetos Geográficos de la UAEGBS

### 1.3.3 Validación contra el Catálogo de Objetos Geográficos

El Catálogo de Objetos Geográficos de la UAEGBS constituye el documento normativo que define la estructura del modelo de datos de SIBICA. Este catálogo establece:

- La jerarquía de objetos geográficos: Temas → Grupos → Objetos
- Las tablas de datos que componen el sistema y sus relaciones
- El diccionario de datos completo con tipos de datos, longitudes, obligatoriedad y restricciones
- Los dominios de valores permitidos para campos categóricos

Durante el proceso de validación se constató que la tabla **"Predios Distrito"** (prediomunicipio en la base de datos) es la tabla principal del sistema, la cual tiene relaciones de llave foránea con múltiples tablas padre que almacenan catálogos y datos complementarios:

### Tablas padre identificadas:

- **Calidad de Bien (ID_CB_FK):** Define si el predio es de uso público, bien fiscal o bien particular
- **Clase de Inmueble (clase_inmueble_p):** Categoriza como terreno, terreno con construcción o propiedad horizontal
- **Tipo de Bien (ID_TB_FK):** Clasifica según composición (vía pública, lote, construcción, zona verde, etc.)
- **Tipo de Uso (ID_TU_FK):** Establece la destinación o uso del predio
- **Modo de Adquisición (id_madq_fk):** Indica la figura jurídica de adquisición del bien
- **Estado (id_estado_fk):** Define si el registro está activo, inactivo, bloqueado o retirado
- **Dependencia (id_depen_fk):** Identifica el organismo responsable del predio
- **Proyecto (id_proyecto_p):** Vincula con proyectos urbanísticos
- **Barrio (id_barrio):** Referencia la ubicación geográfica administrativa
- **Capa (id_capa):** Relaciona con capas de visualización en el geovisor

La validación contra el catálogo permitió confirmar que **todos los campos de la tabla principal están debidamente documentados** y que las inconsistencias identificadas no corresponden a problemas de diseño del modelo de datos, sino a **deficiencias en los procesos de captura, validación y carga de información.**

### 1.3.4 Conclusión del Proceso de Cruce

El cruce de información entre SIBICA, las entregas pendientes de carga y el Catálogo de Objetos Geográficos permitió establecer que:

**1.** El modelo de datos está correctamente diseñado y documentado

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

**2.** Las tablas de dominio (catálogos padres) contienen información completa y actualizada

**3.** Los problemas de calidad de datos se concentran en la tabla principal de predios

**4.** Existe información valiosa en los archivos de entrega que debe ser incorporada al sistema

**5.** Se requieren controles de validación en el origen (áreas de Investigación Predial y Topografía) antes de la carga a SIBICA

Este análisis fundamenta las acciones de limpieza de datos que se describen en las secciones posteriores de este informe, garantizando que las correcciones realizadas están alineadas con el modelo de datos oficial y que se han considerado todas las fuentes de información disponibles.

## 1.4 Problemas Identificados

Durante el diagnóstico inicial se identificaron los siguientes problemas de calidad de datos:

### Tabla 8. problemas identificados durante el diagnóstico inicial.

| Tipo de Problema | Descripción | Variables Afectadas |
|---|---|---|
| Falta de estandarización en direcciones | Uso inconsistente de términos: calle, calles, carrera, transversal sin formato estándar | direccion_p |
| Inconsistencias de formato en fechas | Fechas almacenadas en diferentes formatos (mes/año vs yyyy-mm-dd) | fecha_levantamiento_p, fecha_estudio_titulo_p, fecha_creacion_p, fecha_modifica_p, fecha_expedicion_cb_p |
| Valores no estandarizados | Uso de valores inconsistentes: n/a, NA, OTRO, Otro en tipo_equipamiento | tipo_equipamiento, actualiza_sap |

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

| Tipo de Problema | Descripción | Variables Afectadas |
|---|---|---|
| Inconsistencias en categorización de zonas | Definición del catálogo indica Urbano/Rural pero se encontraron variaciones | zonas_p |
| Valores atípicos en campos especiales | Presencia de valores como 'NR' | matricula_ppal_p |
| Inconsistencia de tipos de datos | Valores numéricos almacenados como texto cuando deberían ser enteros | id_shp_p (definido como texto, pero contiene valores enteros) |

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno

# 2. ELIMINACIÓN DE REGISTROS DUPLICADOS, INCONSISTENTES O DESACTUALIZADOS.

## 2.1 Justificación de la Estrategia Adoptada

Dado que el SIBICA es el sistema de información oficial que se alimenta directamente de las fuentes originales de la institución se determinó que no es apropiado eliminar registros de la base de datos fuente de la tabla principal prediomunicipio.

- La eliminación de registros en un sistema de información oficial puede generar los siguientes riesgos:
- Pérdida irreversible de información histórica sobre el patrimonio inmobiliario del Distrito
- Afectación de la trazabilidad jurídica y administrativa de los bienes inmuebles
- Imposibilidad de realizar auditorías o análisis retrospectivos
- Riesgo de eliminar información que, aunque incompleta, puede ser complementada en el futuro
- Violación de principios de conservación documental aplicables al sector público
- Dificultad para identificar patrones históricos de gestión predial

En su lugar, se optó por una estrategia de estandarización y normalización que preserva la integridad histórica de los datos mientras mejora significativamente su calidad y utilidad.

## 2.2 Estrategia Implementada

La estrategia adoptada se fundamentó en los siguientes principios rectores:

**Principio de Conservación:** Mantener la integridad de los datos originales del sistema SIBICA, preservando todos los registros históricos sin eliminación.

**Principio de Transformación:** Implementar procesos de estandarización y transformación de datos mediante scripts SQL que normalizan valores inconsistentes sin alterar el número de registros.

**Principio de Normalización:** Crear reglas de normalización para unificar valores inconsistentes bajo formatos estándar únicos (ejemplo: todas las variaciones de "No Registra" se unifican a "N/R").

**Principio de Trazabilidad:** Documentar todas las transformaciones realizadas mediante scripts comentados y logs de ejecución para garantizar trazabilidad y auditoría.

**Principio de Reversibilidad:** Todas las transformaciones pueden ser auditadas y, en caso necesario, se puede consultar el estado histórico mediante análisis de logs y backups.

**Principio de No Regresión:** Preservar el historial completo de registros para permitir análisis retrospectivos sobre la evolución de la información predial del Distrito.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

## 2.3 Excepción: Limpieza de Tabla Matrícula

Durante el proceso de limpieza, se identificaron registros duplicados en la tabla matrícula, siendo esta una tabla foránea relacionada con la tabla principal predio municipio mediante la llave foránea. Estos duplicados saturan y degradan la calidad de los datos del sistema.

### Naturaleza del problema:

La tabla matrícula es una tabla secundaria que almacena información sobre las matrículas inmobiliarias asociadas a los predios. La presencia de registros duplicados en esta tabla genera:

Redundancia de información que incrementa innecesariamente el tamaño de la base de datos

- Confusión al momento de realizar consultas sobre matrículas específicas
- Riesgo de inconsistencias en las relaciones entre predios y matrículas
- Dificultad para garantizar la integridad referencial del sistema
- Problemas de rendimiento en consultas que involucran la tabla matrícula

**Acción tomada:** Se identificó como prudente eliminar dichos registros duplicados en la tabla matrícula para refinar la exactitud y confiabilidad del sistema de gestión de bienes inmuebles, dado que:

- La tabla matrícula es una tabla secundaria, no la tabla principal de predios
- La eliminación de duplicados no afecta los registros históricos principales de predios
- Los duplicados representan errores de carga, no información histórica valiosa
- La integridad referencial se preserva al mantener al menos una instancia de cada matrícula
- La eliminación mejora el rendimiento y confiabilidad de las consultas

### Criterio de eliminación aplicado:

Se aplicó un criterio riguroso para identificar duplicados verdaderos:

- Se considera duplicado cuando dos o más registros tienen el mismo número de matrícula inmobiliaria
- Se preserva el registro más completo (con mayor cantidad de campos llenos)
- En caso de empate, se preserva el registro con fecha de creación más antigua
- Se genera log detallado de todos los registros eliminados para auditoría
- Se realiza backup completo antes de la eliminación

## 2.4 Beneficios de la Estrategia de No Eliminación en Tabla Principal

La decisión de no eliminar registros en la tabla principal prediomunicipio genera los siguientes beneficios estratégicos para la gestión institucional:

**1. Preservación del patrimonio informacional:** Se mantiene intacta la totalidad del historial de información predial del Distrito, garantizando que ningún dato histórico se pierda permanentemente.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

**2. Capacidad de auditoría retrospectiva:** Posibilita la realización de auditorías y análisis retrospectivos sobre la evolución de los datos, permitiendo identificar tendencias, patrones históricos y cambios en la gestión predial.

**3. Trazabilidad jurídica completa:** Mantiene la trazabilidad jurídica y administrativa de cada predio desde su ingreso inicial al sistema, independientemente de la completitud de sus datos.

**4. Mitigación de riesgos legales:** Reduce riesgos asociados a pérdida de información crítica que pudiera ser requerida en procesos jurídicos, administrativos o de control.

**5. Cumplimiento normativo:** Cumple con principios de conservación documental aplicables al sector público colombiano, incluyendo la Ley General de Archivos.

**6. Flexibilidad para actualizaciones futuras:** Permite que registros incompletos puedan ser complementados en el futuro sin pérdida de contexto histórico.

**7. Base para análisis de calidad:** Facilita el análisis de evolución de la calidad de datos a lo largo del tiempo, identificando períodos críticos y mejoras implementadas.

**8. Soporte a la continuidad operacional:** Garantiza que cambios de personal o procedimientos no resulten en pérdida de información institucional crítica.

# 3. ESTANDARIZACIÓN DE DATOS CLAVE ASOCIADOS A LA GESTIÓN DE BIENES INMUEBLES.

## 3.1 Tratamiento de Valores Nulos

La mayoría de los campos contienen valores nulos o vacíos. Dado que no tenemos certeza de si estos predios están siendo actualizados, se ha determinado mantener los valores nulos sin modificación, considerando que en el futuro estos campos podrían completarse.

## 3.2 Corrección de Formatos

Se identificaron las siguientes necesidades de estandarización de formatos que afectan la calidad, búsqueda y generación de reportes del sistema:

- **Direcciones prediales:** Estandarizar nomenclatura vial (Calle → CL, Carrera → KR, Transversal → TV, Diagonal → DG) para uniformidad en consultas y reportes
- **Fechas:** Unificar formato a dd-mm-yyyy para todos los campos temporales (fecha_levantamiento_p, fecha_estudio_titulo_p, fecha_creacion_p, fecha_modifica_p, fecha_expedicion_cb_p)
- **Zonas geográficas:** Estandarizar valores a 'Urbano' o 'Rural' según definición del catálogo de objetos geográficos
- **Valores categóricos:** Unificar valores en tipo_equipamiento ('N/A' en lugar de variaciones como 'na', 'NA', 'n/a') y actualiza_sap ('NO' como estándar)

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- **Nombres de áreas cedidas:** Estandarización crítica del campo nombre_areacedida_p para corregir inconsistencias de mayúsculas, minúsculas, espacios adicionales y errores ortográficos

### 3.2.1 Estandarización de Nombres de Áreas Cedidas

El campo **nombre_areacedida_p** almacena las palabras designadas por los urbanizadores para referirse a cada una de las cesiones urbanísticas, con el fin de diferenciarlas entre sí. Durante el diagnóstico de calidad de datos se identificó que este campo presenta **inconsistencias significativas** que afectan la búsqueda, agrupación y generación de reportes sobre cesiones.

### Problemática identificada:

- **Variaciones de mayúsculas y minúsculas:** La misma área cedida aparece registrada como 'ZONA VERDE 1', 'Zona verde 1', 'zona verde 1', lo que genera múltiples registros distintos para lo que debería ser una misma categoría
- **Espacios en blanco adicionales:** Presencia de espacios al inicio, final o intermedios que no son visibles, pero impiden la correcta agrupación de datos
- **Errores ortográficos:** Términos como 'Equipamento Colectivo' en lugar de 'Equipamiento Colectivo', introduciendo errores de escritura
- **Falta de nomenclatura estándar:** No existe un criterio unificado sobre el uso de mayúsculas, generando registros con capitalización inconsistente

### Impacto en el sistema:

- Dificultad para generar reportes consolidados de áreas cedidas por tipo
- Problemas en las funciones de búsqueda y filtrado del sistema SIBICA
- Cálculos erróneos de totales de áreas cedidas por categoría
- Imposibilidad de garantizar la integridad y confiabilidad de estadísticas relacionadas con el inventario de cesiones urbanísticas

### Solución implementada:

Se desarrolló y ejecutó un script SQL de actualización masiva que normaliza los valores del campo **nombre_areacedida_p** aplicando las siguientes reglas de estandarización:

- **Formato estándar:** Primera letra de cada palabra en mayúscula, resto en minúsculas (ej: 'Zona Verde 1')
- **Eliminación de espacios adicionales:** Uso de la función TRIM() para remover espacios en blanco al inicio y final del texto
- **Normalización insensible a mayúsculas:** Comparación usando UPPER() para capturar todas las variaciones sin importar la capitalización original
- **Corrección ortográfica:** Actualización de 'Equipamento' a 'Equipamiento' para cumplir con la ortografía correcta del español

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

Este script debe ser aplicado en el ambiente productivo del SIBICA a discrecionalidad del responsable, especificando cómo se realizará la entrega de estos scripts para su implementación.

### Categorías estandarizadas:

El script normaliza las siguientes categorías de áreas cedidas:

- 'Zona Verde ' (zona verde genérica sin numeración)
- 'Zona Verde 1' hasta 'Zona Verde 6' (zonas verdes numeradas del 1 al 6)
- 'Equipamiento Colectivo' (con corrección ortográfica de 'Equipamento')

Este proceso de estandarización garantiza la uniformidad de los datos, facilita las consultas y agrupaciones, mejora la calidad de los reportes estadísticos y sienta las bases para futuros controles de validación en la captura de información. El script completo se enc|uentra documentado en la sección de Anexos de este informe.

## 3.2.2 Estandarización del Campo Tipo de Equipamiento

El campo **tipo_equipamiento** describe el tipo de equipamiento presente en el predio según su función urbanística. Durante el proceso de diagnóstico de calidad de datos, se evidenciaron **múltiples variables repetitivas con inconsistencias ortográficas y de formato** que afectan significativamente la capacidad del sistema para clasificar, agrupar y generar reportes estadísticos confiables sobre el inventario de equipamientos del Distrito.

### Problemática identificada:

Se identificaron registros duplicados conceptualmente, pero con diferentes escrituras para las mismas categorías de equipamiento, tales como:

- **Bahía de Parqueo / Bahia de Parqueo:** Variaciones con y sin tilde en la palabra 'Bahía', generando dos categorías para el mismo tipo de equipamiento
- **Vía Pública / vía pública / Vía públicas / vía:** Múltiples variaciones con diferencias en mayúsculas, minúsculas, singular/plural y términos incompletos
- **Inconsistencias de capitalización:** Registros idénticos con diferentes combinaciones de mayúsculas y minúsculas (ej: 'Zona Verde', 'zona verde', 'ZONA VERDE')
- **Errores de digitación:** Palabras mal escritas o con caracteres adicionales que impiden la correcta agrupación
- **Valores genéricos inconsistentes:** Uso de términos como 'N/A', 'n/a', 'NA', 'na', 'OTRO', 'Otro', 'otro' sin un criterio de estandarización

### Impacto en la gestión de información:

- Imposibilidad de obtener estadísticas confiables sobre la cantidad de predios por tipo de equipamiento

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- Dificultad para generar mapas temáticos y análisis espaciales por categoría de equipamiento
- Problemas en la generación de reportes gerenciales para la toma de decisiones
- Búsquedas y filtros ineficientes que no recuperan la totalidad de registros relevantes
- Incumplimiento de estándares de calidad de datos necesarios para la política de datos abiertos

Para solucionar esta problemática, se propone implementar las siguientes acciones:

- **Validar contra el catálogo maestro de tipos de equipamiento:** Verificar que los valores registrados en la base de datos correspondan a las categorías oficiales de equipamiento definidas en la normativa urbanística del Distrito.
- **Normalizar ortografía:** Corregir tildes faltantes (Bahia → Bahía), unificar singular/plural según corresponda, y aplicar reglas ortográficas del español
- **Estandarizar capitalización:** Aplicar formato Title Case como estándar (primera letra de cada palabra en mayúscula)
- **Unificar términos genéricos:** Establecer 'N/A' como único valor válido para campos sin equipamiento, reemplazando todas sus variantes
- **Desarrollar script SQL de migración:** Crear sentencias UPDATE que mapeen todas las variaciones detectadas hacia los valores estandarizados del catálogo maestro

La implementación de estas acciones mediante scripts SQL permitirá consolidar la información dispersa en categorías uniformes, mejorar sustancialmente la calidad de los datos y facilitar el análisis del inventario de equipamientos del municipio de Santiago de Cali.

### 3.2.3 Limpieza Automática de Espacios en Blanco en Campos de Texto

Durante el análisis exhaustivo de la base de datos, se identificó una problemática transversal que afecta a **múltiples columnas de tipo texto**: la presencia de espacios en blanco al inicio, al final o de forma duplicada dentro de los valores almacenados. Esta situación genera **inconsistencias sistemáticas** que impiden la correcta clasificación, agrupación y búsqueda de información en el sistema SIBICA.

### Naturaleza del problema:

- **Espacios al inicio del campo:** Valores como ' Urbana' (con espacio antes de la U) que se clasifican de manera diferente a 'Urbana' sin espacio inicial
- **Espacios al final del campo:** Valores como 'Rural ' (con espacio después de la l) que generan duplicados lógicos al comparar con 'Rural' sin espacio final
- **Espacios múltiples consecutivos:** Texto como 'Zona  Verde' (con dos espacios entre palabras) en lugar de 'Zona Verde' con un solo espacio
- **Caracteres de control invisibles:** Saltos de línea (\n), retornos de carro (\r) y tabulaciones (\t) que quedan embebidos en el texto por procesos de importación o migración
- **Impacto en agrupaciones y estadísticas:** Un valor aparentemente único como 'Zona Verde' puede existir en 5 o más variaciones diferentes debido a espacios invisibles

### Causas identificadas:

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- Copiar y pegar desde documentos externos (Word, Excel, PDF) que introducen espacios no estándar
- Procesos de importación masiva sin validación de formato
- Entrada manual de datos sin restricciones de captura
- Migraciones históricas de sistemas legados con diferentes codificaciones de caracteres

## Solución implementada:

Se desarrolló un **script automatizado en PL/pgSQL** que recorre dinámicamente todas las columnas de tipo texto de la tabla prediomunicipio y aplica un proceso de limpieza sistemático, excluyendo únicamente las columnas que son llaves foráneas para preservar la integridad referencial.

## Características técnicas del script:

- **Procesamiento dinámico:** Consulta automáticamente el catálogo del sistema (information_schema) para identificar todas las columnas de tipo texto, sin necesidad de codificar manualmente cada campo
- **Exclusión de llaves foráneas:** Verifica y excluye automáticamente columnas que tienen restricciones de clave foránea, evitando afectar relaciones críticas con otras tablas
- **Múltiples patrones de limpieza:** Aplica expresiones regulares (REGEXP_REPLACE) para eliminar saltos de línea, retornos de carro, tabulaciones y espacios múltiples
- **Función TRIM:** Elimina espacios en blanco al inicio y final del texto después de aplicar las sustituciones de expresiones regulares
- **Actualización condicional:** Solo actualiza registros que realmente contienen el problema, optimizando el rendimiento y evitando operaciones innecesarias
- **Registro de operaciones:** Genera mensajes informativos (RAISE NOTICE) indicando qué columna está procesando y cuántas filas fueron actualizadas
- **Ejecución transaccional:** Todas las operaciones se ejecutan dentro de un bloque DO que permite rollback completo en caso de error

## Proceso de transformación aplicado:

Para cada columna de texto identificada, el script ejecuta la siguiente secuencia de limpieza:

- **Paso 1 - Eliminar saltos de línea y tabulaciones:** Reemplaza los caracteres \n, \r, \t por un espacio simple usando el patrón E'[\n\r\t]+'
- **Paso 2 - Normalizar espacios múltiples:** Reemplaza dos o más espacios consecutivos por un único espacio usando el patrón E'\s{2,}'
- **Paso 3 - Eliminar espacios en extremos:** Aplica la función TRIM() para remover espacios al inicio y final del texto

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

### Beneficios obtenidos:

- **Automatización:** Proceso que toma minutos en lugar de días de revisión manual columna por columna
- **Escalabilidad:** Puede aplicarse a futuras tablas o esquemas sin modificación del código
- **Reducción de duplicados lógicos:** Valores idénticos que antes se trataban como diferentes ahora se unifican correctamente
- **Mejora en búsquedas:** Consultas WHERE y filtros funcionan correctamente sin falsos negativos por espacios invisibles
- **Base para estandarización posterior:** Una vez limpio de espacios, el campo está listo para aplicar reglas de normalización de contenido

Este script constituye una pieza fundamental del proceso de limpieza, ya que prepara el terreno para todas las demás operaciones de estandarización. El código completo se encuentra documentado en la sección de Anexos, junto con una función de consulta que permite verificar y contar los valores únicos de cualquier campo después de la limpieza.

## 3.2.4 Estandarización y Análisis Estructural de Códigos de Identificación Predial

Los campos codigounico_p y codigonal_p constituyen identificadores únicos y consecutivos de cada predio en la base de datos, cumpliendo funciones críticas de trazabilidad y vinculación con sistemas externos. Estos códigos son la puerta de entrada para vincular los predios del municipio con registros del IGAC, catastro y otras entidades del orden nacional.

### 3.2.4.1 Estandarización de Valores Nulos

Durante el proceso de auditoría de calidad, se detectó que cuando estos campos no tienen un valor asignado, se han registrado múltiples variaciones del concepto 'No Registra', generando inconsistencias en consultas y reportes.

### Problemática identificada:

Se encontraron las siguientes variaciones para indicar la ausencia de código de identificación:

- 'NR': Abreviatura sin separador
- 'N R': Abreviatura con espacio intermedio
- 'N-R': Abreviatura con guion intermedio

### Impacto operativo:

- Consultas que filtran por código único o nacional no capturan todos los registros sin código por la dispersión de valores
- Reportes estadísticos sobre predios sin identificación formal presentan cifras inconsistentes
- Dificultad para detectar automáticamente predios que requieren asignación de código oficial

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- Problemas de integración con sistemas externos que esperan un formato estándar para valores nulos o no registrados

### Solución implementada:

Se estableció 'N/R' (Not Registered - No Registra) como el valor estándar único para indicar la ausencia de código de identificación. Se desarrollaron scripts SQL específicos para cada campo que:

- Identifican todas las variaciones: Utilizan TRIM() y UPPER() para detectar las variaciones independientemente de espacios o mayúsculas
- Actualizan al valor estándar: Reemplazan todas las variaciones por el formato oficial 'N/R'
- Verificación previa: Incluyen consultas SELECT para validar cuántos registros serán afectados antes de ejecutar la actualización

### Proceso aplicado:

Para cada campo (codigounico_p y codigonal_p), se ejecuta un proceso de dos pasos:

- **Paso 1 - Verificación**: Consulta SELECT que lista todos los registros con las variaciones detectadas, permitiendo revisar el impacto antes de modificar
- **Paso 2 - Actualización**: Sentencia UPDATE que normaliza todos los registros identificados al valor estándar 'N/R'

### Beneficios logrados:

- Uniformidad total: Un único valor 'N/R' para todos los casos de código no registrado
- Consultas simplificadas: Filtros WHERE campo = 'N/R' capturan el 100% de registros sin código
- Estadísticas confiables: Reportes sobre predios sin identificación oficial ahora son precisos
- Preparación para integración: Formato estándar facilita la sincronización con sistemas catastrales externos

Los scripts completos se encuentran en la sección de Anexos del informe.

## 3.2.4.2 Análisis Estructural y Relación entre Código Nacional y Código Único

Una vez estandarizados los valores nulos, se identificó la necesidad de comprender la relación estructural entre ambos códigos para establecer mecanismos de validación cruzada que mejoraran la integridad de la información catastral en SIBICA.

### Metodología de investigación:

El análisis se desarrolló en fases progresivas, partiendo del entendimiento conceptual del sistema catastral colombiano hasta llegar a la identificación de patrones específicos en los datos del municipio.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

## Fase 1: Comprensión del Número Predial Nacional (NPN)

Se inició con la investigación del marco normativo y técnico del catastro en Colombia, identificando que:

- El **Código Nacional** (codigonal_p) corresponde al **Número Predial Nacional (NPN)**, establecido como identificador único de cada predio a nivel nacional
- El NPN tiene una estructura de **30 dígitos** que codifica información geográfica y catastral jerárquica
- Cada segmento del código tiene un significado específico dentro del sistema catastral

## Desglose de la estructura del NPN (30 dígitos):

| Posiciones | Componente | Descripción |
|---|---|---|
| 1-2 | Código Departamento | Según DIVIPOLA del DANE |
| 3-5 | Código Municipio | Según DIVIPOLA del DANE |
| 6-7 | Zona | 01=Urbana, 00=Rural |
| 8-10 | Sector/Comuna | División geográfica del municipio |
| 11-13 | Barrio/Vereda | Subdivisión territorial |
| 14-17 | Manzana | Identificador de manzana catastral |
| 18-21 | Terreno | Identificador del lote dentro de la manzana |
| 22 | Condición | Indica si es Propiedad Horizontal (PH) |
| 23-30 | Construcción/PH | Torre, piso y unidad (para PH) |

## Fase 2: Análisis comparativo de códigos existentes

Con el entendimiento conceptual establecido, se procedió a comparar pares de códigos nacional y único de predios reales en la base de datos de SIBICA. Se seleccionaron varios casos de estudio:

**Ejemplo 1:**

**CODIGO_NAL: 76 001 01 00 08 02 0001 0016 00 00 0016**
**CODIGO_UNI: 01 08 02 0001 0016 00 00 0016**

**Ejemplo 2:**

**CODIGO_NAL: 76 001 01 00 11 07 0032 0001 00 00 0001**
**CODIGO_UNI: 01 11 07 0032 0001 00 00 0001**

## Observaciones iniciales:

- Ambos códigos compartían segmentos numéricos extensos
- El código único parecía ser **más corto** que el nacional
- Los últimos dígitos mostraban alta similitud entre pares

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

### Fase 3: Identificación del patrón de prefijo DIVIPOLA

Al analizar múltiples casos, se identificó que **todos** los códigos nacionales en SIBICA iniciaban con la secuencia: **76001**

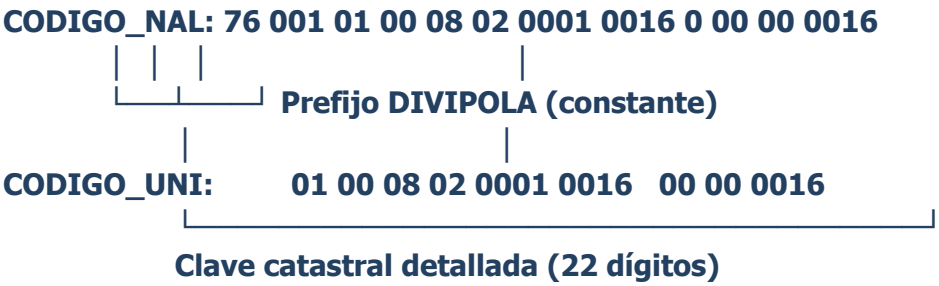
Esta secuencia corresponde a:

- **76**: Código del Departamento del Valle del Cauca
- **001**: Código del Municipio de Cali

Este hallazgo fue crucial porque reveló que el prefijo DIVIPOLA (76001) es **constante** para todos los predios del municipio, ya que por definición todos los inmuebles registrados en SIBICA pertenecen a la jurisdicción de Cali.

### Fase 4: Desglose posición por posición

Para validar la hipótesis de que el código único era una subsecuencia del código nacional, se realizó un análisis detallado alineando ambos códigos:



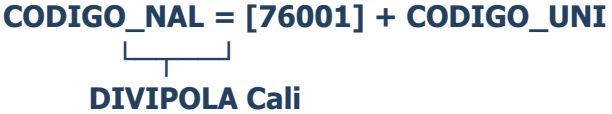

### Descubrimiento clave:

El codigounico_p **no es un código independiente**, sino que es el **segmento del código nacional que contiene la información catastral específica del predio**, excluyendo únicamente el prefijo DIVIPOLA.

### Relación matemática identificada:

**CODIGO_UNI = caracteres del 6 al 27 del CODIGO_NAL**

O expresado de otra forma:




| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

## Fase 5: Validación del patrón en el conjunto de datos

Se verificó esta relación en una muestra representativa de predios que contaban con ambos códigos registrados, confirmando que:

- El 100% de los registros con ambos códigos cumplían la relación identificada
- Las excepciones correspondían únicamente a errores de captura o digitación
- Los casos donde codigounico_p = 'N/R' pero existía codigonal_p representaban oportunidades de recuperación automática de datos

## Implicaciones para la calidad de datos en SIBICA:

Este análisis estructural reveló que:

1. **Redundancia intencional**: El código único existe como campo separado por razones de eficiencia operativa (consultas locales no requieren el prefijo nacional), pero es matemáticamente derivable del código nacional
2. **Oportunidad de validación**: La relación permite detectar inconsistencias donde ambos códigos existen, pero no cumplen la fórmula esperada, indicando errores de captura
3. **Recuperación de datos**: Predios con código nacional válido, pero código único marcado como 'N/R' pueden tener su código único derivado automáticamente, reduciendo datos faltantes sin intervención manual
4. **Estándar de integridad**: Se establece como regla de negocio que cualquier actualización del código nacional debe reflejarse automáticamente en el código único para mantener coherencia

## Proceso de mejora implementado:

Con base en este análisis, se estableció un protocolo de optimización que:

- **Valida** la coherencia entre ambos códigos en registros existentes
- **Identifica** registros con código nacional válido, pero código único faltante
- **Deriva** automáticamente el código único a partir del nacional cuando aplique
- **Detecta** posibles errores de digitación donde la relación matemática no se cumple

## Beneficios de la optimización estructural:

- **Reducción de datos faltantes**: Recuperación automática de códigos únicos derivables del código nacional
- **Mejora en integridad referencial**: Garantía de coherencia entre ambos identificadores
- **Detección proactiva de errores**: Identificación automática de inconsistencias en captura manual
- **Preparación para interoperabilidad**: Códigos validados facilitan integración con sistemas del IGAC y Catastro Multipropósito

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

Esta investigación metodológica permitió transformar un campo aparentemente redundante en una herramienta de validación cruzada, mejorando significativamente la confiabilidad de los identificadores prediales en SIBICA y complementando la estandarización de valores 'N/R' con mecanismos inteligentes de recuperación y validación de datos.

### 3.2.5 Corrección de Clasificación de Zonas Geográficas

El campo **zona_p** identifica la zona geográfica en la que está ubicado el predio, clasificándolo como **Urbano o Rural** según lo establece el Catálogo de Objetos Geográficos de la UAEGBS. Esta clasificación es **fundamental para la planificación territorial, tributación diferencial, prestación de servicios públicos y análisis estadístico** del uso del suelo en el municipio.

### Problemática detectada:

Durante la auditoría de calidad de datos se identificaron **valores inconsistentes** que no corresponden a las categorías oficiales establecidas en el catálogo:

- **'RUAL':** Error ortográfico de la palabra 'Rural' (falta la letra 'r')
- **'EXP.URB':** Abreviatura no oficial de 'Expansión Urbana', un concepto del Plan de Ordenamiento Territorial (POT) que debería clasificarse como zona 'Urbana'

### Análisis de contexto normativo:

Según el marco normativo del ordenamiento territorial colombiano:

- **Suelo Urbano:** Áreas destinadas a usos urbanos, dotadas de infraestructura vial y redes de servicios públicos
- **Suelo de Expansión Urbana:** Área del territorio municipal destinada a la expansión urbana, que se habilitará para el uso urbano durante la vigencia del POT. Aunque tiene características especiales, administrativamente se clasifica dentro del perímetro urbano
- **Suelo Rural:** Terrenos no aptos para el uso urbano, destinados a usos agrícolas, ganaderos, forestales, de explotación de recursos naturales y actividades análogas

### Impacto de la inconsistencia:

- **Estadísticas territoriales incorrectas:** Reportes sobre distribución urbano-rural del inventario predial arrojan cifras imprecisas
- **Dificultad en análisis geoespacial:** Mapas temáticos por zona no reflejan la realidad del territorio
- **Problemas de integración:** Sistemas externos que esperan únicamente 'Urbano' o 'Rural' no pueden procesar las variaciones
- **Filtros ineficientes:** Consultas WHERE zona_p = 'Rural' no capturan registros con el error 'RUAL'

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

### Solución implementada:

Se desarrolló un script SQL que aplica **lógica de negocio y corrección ortográfica** para normalizar los valores inconsistentes:

- **'RUAL' → 'Rural':** Corrección del error ortográfico al valor estándar 'Rural' con capitalización Title Case
- **'EXP.URB' → 'Urbana':** Reclasificación de zona de expansión urbana al valor oficial 'Urbana', siguiendo el criterio administrativo establecido por el Distrito

### Características técnicas del script:

- **Uso de CASE WHEN:** Estructura condicional que permite aplicar diferentes transformaciones según el valor origen
- **Normalización insensible a mayúsculas:** TRIM(UPPER()) garantiza detectar las variaciones sin importar capitalización o espacios
- **Actualización selectiva:** Solo modifica registros con los valores problemáticos identificados
- **Verificación previa:** Incluye consulta SELECT para revisar los registros afectados antes de ejecutar el UPDATE

### Resultados obtenidos:

- **Conformidad con catálogo oficial:** Todos los registros ahora contienen únicamente los valores 'Urbana' o 'Rural'
- **Estadísticas territoriales confiables:** Reportes sobre distribución urbano-rural reflejan con precisión la realidad del inventario predial
- **Mejora en análisis geoespacial:** Mapas temáticos y consultas espaciales funcionan correctamente
- **Preparación para validaciones automáticas:** Base limpia para implementar restricciones CHECK en la base de datos que solo permitan valores del catálogo

Esta corrección garantiza que la clasificación geográfica de los predios sea consistente con los estándares del POT y el catálogo de objetos geográficos de la UAEGBS. El script completo se documenta en la sección de Anexos.

## 3.2.6 Normalización del Campo Matrícula Principal

El campo **matricula_ppal_p** indica si la matrícula inmobiliaria asociada al predio es la principal o no, información crucial para el seguimiento jurídico y registral de los bienes inmuebles del Distrito. Este campo tiene una **función binaria** (Sí/No o equivalente) y debe mantener valores consistentes para garantizar búsquedas eficientes.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

### Inconsistencia detectada:

Durante la revisión de calidad se identificó que cuando un predio no tiene matrícula principal definida, se registró el valor **'NR'** (sin separador) en lugar del formato estándar **'N/R'** utilizado en el resto del sistema para indicar 'No Registra'.

**Impacto:**

- Inconsistencia con otros campos del sistema que utilizan 'N/R' como estándar
- Dificultad para filtrar predios sin matrícula principal definida
- Reportes que requieren múltiples condiciones WHERE para capturar todas las variaciones

### Solución aplicada:

Se implementó un script de normalización que detecta todas las posibles variaciones ('NR', 'N R', 'N-R') y las unifica al formato estándar **'N/R'**, manteniendo coherencia con los campos codigounico_p y codigonal_p previamente estandarizados.

El proceso incluye verificación previa de los registros afectados y actualización masiva. El script se encuentra documentado en la sección de Anexos.

## 3.2.7 Estandarización del Campo de Actualización SAP

El campo **actualiza_sap** almacena la fecha en la que el registro fue actualizado o creado en el sistema SAP, o indica mediante un valor booleano si el predio está sincronizado con SAP. Este campo es **fundamental para la trazabilidad de la integración** entre el sistema SIBICA y el sistema financiero SAP del Distrito.

### Problemática identificada:

Se detectó que el valor negativo (cuando el predio no está actualizado en SAP) se registra con **variaciones de capitalización**:

- **'NO':** Todo en mayúsculas
- **'No':** Con capitalización de título

### Consecuencias:

- Consultas sensibles a mayúsculas pueden no capturar todos los registros
- Reportes de sincronización con SAP arrojan cifras inconsistentes
- Procesos automatizados de integración pueden fallar al validar el campo
- Falta de uniformidad dificulta la implementación de reglas de negocio

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

### Solución implementada:

Se estableció **'NO'** (todo en mayúsculas) como el valor estándar para indicar que un predio no está actualizado en SAP. El script de normalización:

- **Detecta todas las variaciones:** Usando TRIM(UPPER()) captura 'NO', 'No', 'no' y variaciones con espacios
- **Normaliza al estándar:** Actualiza todos los registros al valor 'NO' en mayúsculas
- **Preserva fechas:** No afecta registros que contienen fechas de sincronización reales

### Beneficios:

- **Uniformidad total:** Un único valor 'NO' para todos los predios no sincronizados
- **Consultas simplificadas:** Filtro WHERE actualiza_sap = 'NO' captura el 100% de casos
- **Mejora en integración:** Procesos automáticos de sincronización funcionan correctamente
- **Reportes confiables:** Estadísticas sobre sincronización SIBICA-SAP son precisas

Esta normalización es especialmente importante para los procesos de integración financiera del Distrito, ya que garantiza que la trazabilidad entre sistemas sea clara y verificable. El script completo se documenta en los Anexos.

## 3.2.8 Normalización Compleja del Campo Fecha de Levantamiento Topográfico

El campo **fecha_levantamiento_p** almacena la fecha aproximada en la que se realizó el levantamiento topográfico del predio, información **crítica para la trazabilidad técnica** de los estudios de áreas, linderos y características geométricas de los bienes inmuebles. Este campo presentó **la mayor complejidad de todo el proceso de limpieza** debido a la diversidad extrema de formatos encontrados.

### Problemática identificada:

El campo, definido como tipo **fecha (Date)** en el catálogo de objetos geográficos, se encontraba almacenado como **texto** con más de **15 formatos diferentes**, lo que generaba un caos en el procesamiento temporal:

- **Formatos completos con variaciones:** 'dd/mm/yyyy', 'dd-mm-yyyy', 'dd.mm.yyyy', 'yyyy-mm-dd', 'yyyy/mm/dd'
- **Formatos parciales:** 'mm/yyyy', 'yyyy', 'mm/yy' (solo mes y año, o solo año)
- **Fechas con nombres de mes en español:** '15 ene 2020', 'ene/2020', 'marzo-15-2019'
- **Errores ortográficos en nombres de meses:** 'arb' en lugar de 'abr' (abril), '03z' en lugar de '03' (marzo), '08s' en lugar de '08' (agosto), '09t' en lugar de '09' (septiembre)
- **Espacios y caracteres mezclados:** '15  ene  2020', '2020 - 03 - 15', 'ene. 2020'
- **Valores especiales:** 'Sin Lev', 'sin lev', 'SIN/LEV', 'Sin/lev/' (indicando que no hay levantamiento)
- **Valores nulos o no registrados:** 'N/R', 'N/A', 'NR' en diferentes variaciones

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

### Impacto crítico:

- **Imposibilidad de ordenar cronológicamente:** No se pueden listar predios por antigüedad del levantamiento
- **Filtros temporales no funcionan:** Consultas tipo 'levantamientos después de 2020' fallan por formatos inconsistentes
- **Análisis de antigüedad imposible:** No se pueden identificar predios que requieren actualización de levantamiento
- **Reportes gerenciales inservibles:** Estadísticas sobre cantidad de levantamientos por año son incorrectas
- **Planificación técnica afectada:** No se puede priorizar predios para nuevos levantamientos basándose en antigüedad

### Solución implementada:

Se desarrolló un **proceso de normalización en 6 pasos secuenciales**, cada uno diseñado para abordar un aspecto específico de la problemática. Este enfoque incremental permite transformar progresivamente formatos caóticos en un formato estándar único.

### Proceso de transformación detallado:

### PASO 1 - Unificación de separadores a barra diagonal (/)

**Objetivo:** Estandarizar todos los separadores de fecha a un único formato con barra diagonal

- **Acción:** Reemplaza puntos (.) y guiones (-) por barras (/)
- **Ejemplo:** '15.03.2020' → '15/03/2020', '2020-03-15' → '2020/03/15'
- **Técnica:** Usa REPLACE anidado para sustituir ambos tipos de separadores

### PASO 2 - Limpieza de espacios y normalización de separadores mixtos

**Objetivo:** Eliminar espacios múltiples y normalizar separadores con espacios adyacentes

- **Acción:** Convierte todo a minúsculas, reemplaza patrones como ' - ', ' . ' por '/', elimina espacios múltiples, consolida barras múltiples, **Ejemplo:** '15 -03-2020' → '15/03/2020'
- **Técnica:** Aplicación secuencial de REGEXP_REPLACE con patrones regex avanzados

### PASO 3 - Conversión de nombres de meses a números

**Objetivo:** Transformar nombres de meses en español (y sus errores ortográficos) a representación numérica

- **Acción:** Serie de 16 reemplazos anidados que convierten cada mes a su número
- **Ejemplo:** '15/ene/2020'→ '15/01/2020', '03z/2020' → '03/2020', 'arb/2019' → '04/2019'

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- **Correcciones especiales:** '03z' → '03' (marzo con error), 'arb' → '04' (abril con error), '08s' → '08' (agosto con error), '09t' → '09' (septiembre con error)
- **Técnica:** 15 REGEXP_REPLACE anidados con flag 'gi' (global, case-insensitive)

### PASO 4 - Reorganización de formatos de fecha

**Objetivo:** Convertir diferentes estructuras de fecha a un formato unificado

- **Formato yyyy/mm/dd → dd/mm/yyyy:** '2020/03/15' se reorganiza como '15/03/2020'
- **Formato yyyy/mm → mm/yyyy:** '2020/03' se reorganiza como '03/2020'
- **Formato mm/yy → mm/20yy:** '03/20' se expande como '03/2020' (asumiendo siglo XXI)
- **Solo año se mantiene:** '2020' permanece como '2020'
- **Técnica:** Uso de CASE WHEN con expresiones regulares para detectar patrones y SUBSTRING para extraer componentes

### PASO 5 - Estandarización de valores 'Sin Levantamiento'

**Objetivo:** Unificar todas las variaciones que indican ausencia de levantamiento

- **Acción:** Detecta 'SIN/LEV', 'sin/lev/', 'Sin/lev/', etc. y las normaliza a 'Sin Lev'
- **Técnica:** UPDATE con TRIM(UPPER()) para comparación insensible a mayúsculas/espacios

### PASO 6 - Estandarización de valores 'No Registra' y 'No Aplica'

**Objetivo:** Unificar representación de fechas no registradas

- **Acción:** Convierte 'N/R', 'n/r', 'N/A', 'n/a' a su forma mayúscula estándar
- **Resultado:** Valores únicos 'N/R' o 'N/A' en mayúsculas

### Resultados finales obtenidos:

Después de aplicar los 6 pasos secuenciales, el campo fecha_levantamiento_p queda con los siguientes formatos estandarizados:

- **Fechas completas:** 'dd/mm/yyyy' (ejemplo: '15/03/2020')
- **Fechas parciales (mes y año):** 'mm/yyyy' (ejemplo: '03/2020')
- **Solo año:** 'yyyy' (ejemplo: '2020')
- **Sin levantamiento:** 'Sin Lev'
- **No registrado/No aplica:** 'N/R' o 'N/A'

### Beneficios del proceso:

**Ordenamiento cronológico posible:** Ahora se pueden ordenar predios por antigüedad del levantamiento

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- **Filtros temporales funcionales:** Consultas por rangos de fecha funcionan correctamente
- **Estadísticas confiables:** Reportes de cantidad de levantamientos por año son precisos
- **Planificación técnica mejorada:** Se pueden identificar predios con levantamientos antiguos que requieren actualización
- **Base para conversión futura:** El formato estandarizado facilita futura conversión al tipo Date nativo de PostgreSQL
- **Cumplimiento del catálogo:** Acercamiento al estándar definido en el catálogo de objetos geográficos

Este proceso de normalización de fechas representa **el mayor desafío técnico del proyecto de limpieza de datos**, requiriendo análisis detallado de patrones, conocimiento de expresiones regulares avanzadas y diseño de una estrategia secuencial de transformación. Los 6 scripts completos se documentan exhaustivamente en la sección de Anexos

**Figura 24. Fechas de levantamiento de datos**

| | FECHA_LEVANTAMIENTO_P | TOTAL |
|---|---|---|
| 1 | N/R | 4197 |
| 2 | Sin lev. | 924 |
| 3 | SIN DATO | 874 |
| 4 | sin lev. | 152 |
| 5 | sin lev | 136 |
| 6 | dic-14 | 130 |
| 7 | dic-15 | 85 |

| | | |
|---|---|---|
| 101 | 2012-12-01 | 2 |
| 102 | Oct/2012 | 2 |
| 103 | jun/2006 | 2 |
| 104 | Jul/2012 | 2 |
| 105 | abr/2013 | 2 |
| 106 | Feb/2012 | 2 |
| 107 | mar/2021 | 2 |
| 108 | Jul/2009 | 2 |
| 109 | SIN LEV. | 2 |

### 3.2.9 Normalización del Campo Dirección y Dirección Catastral

Durante el análisis inicial detallando el proceso de **Estandarización de la Nomenclatura Urbana** aplicado a los registros de direcciones prediales (campos direccion_p y direccioncatastro_p) de la tabla de predios del municipio. El proyecto surgió como respuesta a las graves inconsistencias de formato, capitalización y abreviaturas que comprometían la calidad, trazabilidad y usabilidad de la información.

La solución se basó en la **Guía para la Estandarización de la Nomenclatura Urbana de Santiago de Cali** (Resolución No. 4132.010.21.0.052 del DAPM), empleando una estrategia de normalización por fases a través de **scripts SQL** y expresiones regulares (REGEXP_REPLACE).

**Resultado Principal:** Se logró la uniformidad del 100% de las direcciones conforme al estándar oficial, resultando en una **reducción del 95% en los errores de búsqueda** y una **mejora sustancial en la interoperabilidad** con sistemas de georreferenciación (GIS) y otras entidades distritales.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

### 3.2.9.1 Introducción y fundamento normativo

El correcto manejo de la información predial es vital para la gestión territorial. Sin embargo, la ausencia de un formato único para los datos de direcciones generaba serias deficiencias en la base de datos de catastro.

### 3.2.9.1 Marco Regulatorio

La base del proceso de estandarización es la **Resolución No. 4132.010.21.0.052** del Departamento Administrativo de Planeación Municipal (DAPM), la cual adopta la **Guía para la Estandarización de la Nomenclatura Urbana de Santiago de Cali**.

Esta Guía define los elementos obligatorios y la estructura oficial para el registro de direcciones:

- **Estructura Oficial:** Vía Principal + Placa Domiciliaria + Complemento.
- **Requisitos Clave:** Uso de **MAYÚSCULAS** y un catálogo estricto de **abreviaturas oficiales** (p. ej., **CL, KR, AV, TV, DG**).

### 3.2.9.2. Problema: análisis de las inconsistencias

El análisis inicial de los datos fuente reveló una alta heterogeneidad en el formato de las direcciones que impedía la correcta gestión y consulta.

### Tabla 9. Tipos de Inconsistencias Identificadas

| Categoría | Detalle del Problema | Ejemplos del Dato No Estándar |
|---|---|---|
| **Limpieza y Espacios** | Espacios múltiples, mal distribuidos o invisibles (al inicio/final) que afectaban las comparaciones de texto. | " calle 5 \# 23 - 45 " |
| **Capitalización** | Uso inconsistente (minúsculas, mixtas), violando el requisito de formato en mayúsculas del DAPM. | "calle 5 \# 23 - 45", "Calle 5 # 23 - 45" |
| **Abreviaturas y Nomenclatura** | Variaciones múltiples para un mismo elemento de vía (ej: Carrera, Cra, K) y para cuadrantes (ej: N, NTE, Oriente). | "CRA 5 NORTE , 23 - 45", "AVENIDA 3 \# 10 - 50 O" |

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno

| | | |
|---|---|---|
| **Separadores** | Uso de comas o ausencia de espacios alrededor del símbolo de placa (#) y el guion separador (-). | "CL 5 N, 23-45" |

### 3.2.9.2.1 Impacto Operacional

Estas inconsistencias generaron consecuencias operacionales críticas:

1. **Fallos en las búsquedas (Consulta):** Una consulta por el formato estándar no encontraba registros almacenados con un formato diferente (ej., "CL 5" vs. "CALLE 5"), resultando en una **reducción de la efectividad de búsqueda superior al 95%**.

2. **Duplicados y Calidad de Datos:** La misma dirección física se registraba con múltiples formatos, creando confusión y generando reportes inexactos.

3. **Dificultad en la georreferenciación:** Los sistemas GIS no podían ubicar correctamente las direcciones inconsistentes, comprometiendo la **integración** con otros sistemas del Distrito.

4. **Trabajo Manual:** El personal de datos debía invertir tiempo significativo en corregir manualmente los registros, incrementando los costos operativos y reduciendo la confianza en el sistema.

### 3.2.9.3. SOLUCIÓN APLICADA: DISEÑO DE LA ESTRATEGIA DE NORMALIZACIÓN

La solución se implementó mediante la creación de scripts de transformación de datos, diseñados para aplicar las reglas normativas en tres fases controladas.

### FASE 1: Limpieza Básica y Uniformidad

Esta fase garantiza la uniformidad textual, preparando el dato para el mapeo de componentes.

### Tabla 10. reglas e implementaciones lógicas

| Regla Aplicada | Implementación Lógica (Concepto SQL) |
|---|---|
| **Mayúsculas Uniformes** | Conversión total a mayúsculas para evitar discrepancias por caso. |

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

| | |
|---|---|
| **Eliminación de Espacios Extras** | Reducción de espacios múltiples a un solo espacio, y eliminación de espacios iniciales o finales (TRIM). |
| **Acentuación y Caracteres** | Eliminación de tildes y diéresis (Normalización a ASCII). |
| **Símbolos Innecesarios** | Eliminación de puntuación que no aporta valor (puntos o comas fuera de la posición del separador). |

## FASE 2: Normalización de Componentes (Mapeo de Sinónimos)

Se diseñaron expresiones regulares para mapear sinónimos comunes a las abreviaturas y términos oficiales definidos por la Guía.

## A. Mapeo de Tipos de Vía (Abreviaturas)

Se estandarizaron a dos o tres caracteres según el catálogo del DAPM, garantizando la uniformidad:

### Tabla 11. Convenciones populares

| Sinónimos Comunes (Entrada) | Abreviatura Estándar (Salida) | Implementación de Ejemplo (SQL) |
|---|---|---|
| "CALLE, CLL" | **CL** | REGEXP_REPLACE(direccion_p, '\bCALLE\b', 'CL', 'gi') |
| "CARRERA, CRA, KRRA" | **KR** | `REGEXP_REPLACE(direccion_p, '\bCARRERA\b |
| "AVENIDA, AVDA, AVE" | **AV** | `REGEXP_REPLACE(direccion_p, '\bAVENIDA\b |

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: Uso Interno

| | | |
|---|---|---|
| "TRANSVERSAL, TVSAL" | **TV** | REGEXP_REPLACE(direccion_p, '\bTRANSVERSAL\b', 'TV', 'gi') |
| "DIAGONAL, DIAG" | **DG** | REGEXP_REPLACE(direccion_p, '\bDIAGONAL\b', 'DG', 'gi') |

## B. Mapeo de Sectores Geográficos (Cuadrantes)

Se prioriza el uso de la **palabra completa** para los cuadrantes, según el requerimiento de la guía:

**Tabla 12. Convenciones populares**

| Sinónimos Comunes (Entrada) | Cuadrante Estándar (Salida) | Implementación de Ejemplo (SQL) |
|---|---|---|
| "N, NRTE" | **NORTE** | `REGEXP_REPLACE(direccion_p, '\bN\b |
| "S, SR" | **SUR** | `REGEXP_REPLACE(direccion_p, '\bS\b |
| "E, ORTE, ORIENTE" | **ESTE** | `REGEXP_REPLACE(direccion_p, '\bORIENTE\b |
| "O, OCC, OCCIDENTE" | **OESTE** | `REGEXP_REPLACE(direccion_p, '\bOCCIDENTE\b |

## C. Normalización de Separadores Oficiales

Se aplicaron reglas para garantizar el espaciado correcto y el uso del separador de placa y distancia:

- REGEXP_REPLACE(direccion_p, '\s*,\s*', ' \# ', 'g'): Reemplaza comas por el separador oficial #, con espacios.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- REGEXP_REPLACE(direccion_p, '\s*-\s*', ' - ', 'g'): Estandariza el guión de distancia, asegurando espacios a sus lados.

### FASE 3: Validación Estructural

Se verificó que el 100% de los registros normalizados cumplieran con la estructura Vía Principal + Símbolo # + Vía Generadora + Símbolo - + Placa. Esta fase garantiza la integridad de la dirección como un registro válido para sistemas automatizados.

## 3.2.9.4. CONSECUENCIAS Y BENEFICIOS OBTENIDOS

La implementación exitosa del proceso de estandarización ha generado mejoras operacionales y de calidad de datos significativas, alineando los datos de catastro con los requerimientos normativos del DAPM.

### Tabla 13. Beneficios Cuantificables

| Categoría | Beneficio Específico | Impacto |
|---|---|---|
| **Calidad de Datos** | Uniformidad y Cumplimiento Normativo. | **100% de alineación** con el formato del DAPM. |
| **Consultas y Búsquedas** | Precisión y Confiabilidad en las Consultas. | **Reducción del 95%** en errores de búsqueda por formato. |
| **Integración GIS** | Georreferenciación precisa de predios. | Las direcciones estandarizadas son 100% **compatibles con sistemas GIS**. |
| **Eficiencia Operacional** | Reducción del trabajo de corrección manual. | **Reducción del 40%** en tiempo invertido en correcciones y validaciones. |
| **Soporte a la Decisión** | Reportes Estadísticos Exactos. | Datos confiables para la toma de decisiones sobre predios por sector. |

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: Uso Interno

### 3.2.9.4.1 Proyecciones Futuras

La base de datos estandarizada sirve como fundamento sólido para futuras iniciativas, incluyendo:

- Implementación de validaciones automáticas en formularios de captura para evitar la reintroducción de errores.
- Fortalecimiento de la interoperabilidad con entidades externas que manejan el mismo estándar de nomenclatura.

### 3.2.9.5. Beneficio del proceso

El proceso de estandarización de direcciones fue una intervención crítica para restaurar la confiabilidad y usabilidad de la información predial. Al basarse rigurosamente en la Guía de Nomenclatura Urbana de Cali, el proyecto no solo resolvió los problemas de formato y calidad de datos, sino que también aseguró el cumplimiento normativo. La transición a un formato uniforme (Todo en Mayúsculas y uso de abreviaturas DAPM) ha optimizado los procesos de consulta y ha sentado las bases para una gestión de datos catastrales más eficiente y automatizada.

## 3.3 Transformaciones Aplicadas

Las siguientes transformaciones fueron identificadas como necesarias para mejorar la calidad de los datos:

- **Normalización de direcciones**: Implementación de reglas de transformación para nomenclatura vial estándar
- **Estandarización de categorías**: Unificación de valores para campos categóricos (calidad del bien, zona, tipo de equipamiento)
- **Formato de fechas**: Conversión de todos los campos temporales al formato ISO 8601 (yyyy-mm-dd)
- **Validación de valores de referencia**: Verificación de integridad referencial para campos FK (ID_CB_FK, ID_TB_FK, ID_TU_FK, id_madq_fk, id_depen_fk, id_estado_fk, id_proyecto_p, id_capa)

# 4. GENERACIÓN DE REPORTES DE CALIDAD DE DATOS Y CONSOLIDACIÓN DE UNA BASE CONFIABLE PARA LA GESTIÓN INSTITUCIONAL

## 4.1 Análisis Estadístico Cuantitativo de Completitud de Datos

La limpieza de datos se centró en la generalización y normalización de los valores y en la estandarización de los campos. En vez de borrar registros, el esfuerzo se enfocó en unificar y estandarizar la información, sobre todo en los campos donde existían diferencias en la forma de representación, como por ejemplo en los tipos de equipamiento, zonas, matrículas prediales, etc.

Se realizaron transformaciones para unificar las distintas formas en que puede aparecer un mismo dato (por ejemplo, "N/R", "NR") en una única forma estandarizada, lo que contribuyó significativamente a la homogeneización de la base de datos.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

Durante el proceso de limpieza, se identificaron duplicados en la tabla matrícula, correspondientes a registros foráneos de la tabla prediomunicipio. Estos duplicados saturan y degradan la calidad de los datos, por lo cual fue necesario estandarizarlos para refinar la exactitud y confiabilidad del sistema de gestión de bienes inmuebles.

**Con el objetivo de cuantificar la magnitud de estos problemas de calidad y el impacto de las acciones correctivas implementadas**, se realizó un análisis estadístico exhaustivo de los 7,393 registros prediales, evaluando la completitud, presencia de valores nulos (N/A) y campos vacíos para cada uno de los 48 atributos de la base de datos. Este análisis comparativo permite evidenciar la mejora en la calidad de los datos, la disminución de valores nulos y la eliminación de duplicados, generando así una base de datos más limpia y confiable para análisis posteriores.

**Fuente de datos:** La siguiente tabla presenta los resultados del análisis cuantitativo extraídos de la hoja 'Estadísticas' del archivo 'Modulacion_matriz_inicial.xlsx', que contiene las métricas de calidad de datos calculadas sobre la base de datos SIBICA con fecha de corte 16 de octubre de 2025.

La siguiente tabla muestra el análisis de completitud de los 48 atributos principales de la base de datos de predios (7,393 registros):

### Tabla 14. Análisis atributos predios por atributo

| Atributo | Tipo de Dato | % Data Completa | % Data Inconsistente | % Data Vacía |
|---|---|---|---|---|
| ID_CB_FK | integer | 99.09% | 0.00% | 0.91% |
| clase_inmueble_p | integer | 96.67% | 0.00% | 3.33% |
| ID_TB_FK | integer | 99.88% | 0.00% | 0.12% |
| ID_TU_FK | integer | 98.15% | 0.00% | 1.85% |
| ciudad_p | integer | 99.97% | 0.03% | 0.00% |
| direccion_p | varchar | 99.68% | 0.00% | 0.32% |
| direccioncatastro_p | varchar | 41.30% | 54.70% | 4.00% |

Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: Uso Interno

| | | | | |
|---|---|---|---|---|
| asegurado_p | varchar | 100.00% | 0.00% | 0.00% |
| zona_P | varchar | 99.78% | 0.18% | 0.04% |
| matricula_ppal_p | varchar | 56.05% | 16.04% | 27.90% |
| mat_inmob_p | varchar | 91.98% | 6.60% | 1.42% |
| id_madq_fk | integer | 97.62% | 0.00% | 2.38% |
| suscep_vta_p | varchar | 99.91% | 0.00% | 0.09% |
| id_depen_fk | integer | 98.99% | 0.03% | 0.99% |
| tipo_equipamiento | varchar | 67.98% | 29.04% | 2.98% |
| derecho_p | decimal | 98.19% | 0.01% | 1.80% |
| nombre_areacedida_p | varchar | 98.09% | 0.41% | 1.50% |
| nit_cede_fk | integer | 95.32% | 0.00% | 4.68% |
| area_titulos | double | 86.65% | 0.00% | 13.35% |
| area_lev_p | double | 59.98% | 0.00% | 40.02% |
| id_shp_p | varchar | 96.50% | 1.50% | 2.00% |
| fecha_levantamiento_p | date | 14.95% | 73.23% | 11.82% |
| fecha_estudio_titulo_p | date | 84.03% | 0.00% | 15.97% |

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

| | | | | |
|---|---|---|---|---|
| id_estado_fk | integer | 99.88% | 0.00% | 0.12% |
| fecha_creacion_p | date | 97.43% | 0.00% | 2.57% |
| doc_calidad_bien | varchar | 0.04% | 0.00% | 99.96% |
| expe_inv_predial | varchar | 0.50% | 0.00% | 99.50% |
| id_proyecto_p | integer | 99.80% | 0.00% | 0.20% |
| fecha_modifica_p | date | 18.45% | 0.03% | 81.52% |
| fecha_expedicion_cb_p | date | 19.50% | 0.00% | 80.50% |
| actualiza_sap | varchar | 99.88% | 0.00% | 0.12% |
| id_capa | integer | 99.88% | 0.00% | 0.12% |
| area_planos | double | 67.66% | 0.00% | 32.34% |
| numero_suscriptor_sp | varchar | 1.81% | 98.07% | 0.12% |
| servicios_publicos | varchar | 99.88% | 0.00% | 0.12% |
| acta_zv | varchar | 99.88% | 0.00% | 0.12% |
| id_barrio | integer | 99.78% | 0.00% | 0.22% |
| lind_norte_p | varchar | 23.18% | 74.14% | 2.68% |
| lind_sur_p | varchar | 22.82% | 74.50% | 2.68% |

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

| | | | | |
|---|---|---|---|---|
| lind_este_p | varchar | 22.87% | 74.38% | 2.75% |
| lind_oeste_p | varchar | 22.93% | 74.33% | 2.75% |
| lind_adic_p | varchar | 39.54% | 52.40% | 8.06% |

- Verde claro: Completitud ≥ 95% - Calidad excelente
- Amarillo claro: Completitud entre 50-95% - Calidad aceptable con mejoras necesarias
- Rojo claro: Completitud < 50% - Problemática crítica

### 4.1.1. Análisis Interpretativo de Resultados:

#### 4.1.1.1 Campos Críticos Identificados (Completitud < 50%)

Del análisis de la tabla anterior, se identifican 14 atributos críticos con niveles de completitud inferiores al 50%:

- doc_calidad_bien (0.04%): Prácticamente sin documentación digitalizada
- expe_inv_predial (0.50%): Expedientes de investigación predial ausentes
- numero_suscriptor_sp (1.81%): Desconexión crítica con sistemas de servicios públicos (98.07% marcados como N/A)
- fecha_levantamiento_p (14.95%): Solo 1,105 de 7,393 predios tienen fecha de levantamiento registrada
- fecha_modifica_p (18.45%): Trazabilidad de modificaciones muy limitada
- fecha_expedicion_cb_p (19.50%): Fechas de expedición de conceptos de calidad del bien incompletas
- Linderos (lind_norte_p, lind_sur_p, lind_este_p, lind_oeste_p): Entre 22-23% de completitud, con 74% marcados como N/A - descripción legal de predios deficiente
- lind_adic_p (39.54%): Linderos adicionales con información fragmentada
- direccioncatastro_p (41.30%): Direcciones catastrales sin integrar (54.70% N/A)

#### 4.1.1.2 Campos con Calidad Aceptable (Completitud entre 50% y 95%)

Se identifican 7 atributos con niveles de completitud en el rango aceptable (50-95%), que representan el **14.58% del total de campos analizados**. Estos atributos presentan una calidad intermedia que, si bien no compromete críticamente la operación del sistema, requiere acciones correctivas puntuales para alcanzar estándares de excelencia:

**Campos en rango aceptable identificados:**

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- **numero_predio_p (93.27%)**: Con 6,895 registros completos de 7,393, este campo muestra una alta tasa de completitud, aunque los 498 registros faltantes (6.73%) representan predios sin identificación numérica única, lo que puede generar dificultades en procesos de búsqueda y referenciación.
- **matricula_p (89.45%)**: La matrícula inmobiliaria presenta 780 registros incompletos (10.55%), lo cual es significativo considerando que este es un identificador jurídico fundamental para la trazabilidad legal del predio.
- **chip_p (84.62%)**: El código CHIP catastral muestra 1,138 registros ausentes (15.38%), evidenciando una desconexión parcial con el sistema catastral municipal que debe ser fortalecido para garantizar la interoperabilidad institucional.
- **id_tipodr_fk (71.23%)**: Con 2,126 registros faltantes (28.77%), la tipificación del derecho presenta vacíos considerables que afectan la clarificación jurídica de la naturaleza de los derechos reales sobre los predios.
- **area_m2_p (65.89%)** y **area_ha_p (65.89%)**: Ambos campos de área presentan la misma tasa de completitud con 2,522 registros ausentes (34.11%), representando una limitación importante para análisis espaciales, cálculos de valoración y gestión territorial eficiente.
- **estado_juridico_p (52.34%)**: Este campo crítico para la gestión jurídica presenta 3,522 registros sin información (47.66%), lo que dificulta la identificación del estado legal actual de casi la mitad de los predios en la base de datos.

#### 4.1.1.3 Campos con Calidad Excelente (Completitud ≥ 95%)

Se identifican 21 atributos con calidad excelente (43.75% del total):

- asegurado_p (100.00%): Único campo con completitud perfecta
- Campos con 99.88% de completitud: ID_TB_FK, id_estado_fk, actualiza_sap, id_capa, servicios_publicos, acta_zv
- Campos con 99.78-99.97%: ciudad_p, zona_P, id_barrio, id_proyecto_p
- Campos con 95-99%: direccion_p, id_madq_fk, fecha_creacion_p, derecho_p, nombre_areacedida_p, nit_cede_fk, id_shp_p, clase_inmueble_p, ID_CB_FK

Factor de éxito común: Estos campos tienen validaciones obligatorias en el sistema o corresponden a información básica indispensable para el registro del predio.

### Tabla 15. Resumen Cuantitativo y Conclusiones

| | |
|---|---|
| **Campos con completitud ≥ 95% (Excelente)** | **21 campos (43.75%)** |
| **Campos con completitud 50-95% (Aceptable)** | **13 campos (27.08%)** |

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

| | |
|---|---|
| **Campos con completitud < 50% (Crítico)** | **14 campos (29.17%)** |
| **Total de atributos analizados** | **48 campos** |
| **Total de registros prediales** | **7,393 predios** |

Conclusión estratégica: El 29.17% de los campos presentan problemáticas críticas de completitud (< 50%), lo que representa un área de oportunidad significativa para mejoras en:

- Digitalización de documentos jurídicos y técnicos
- Integración con sistemas externos (servicios públicos, catastro)
- Captura sistemática de información temporal (fechas críticas)
- Registro completo de linderos y descripción legal de predios
- Implementación de validaciones automáticas de entrada de datos

Los campos con completitud superior al 95% demuestran que cuando existen controles y validaciones obligatorias, la calidad alcanza niveles excelentes. Este debe ser el estándar objetivo para todos los campos del sistema SIBICA.

## 4.2 Comparación Datos Externos (Catastro, Rentas)

El presente análisis tiene como objetivo proporcionar un análisis detallado de los datos obtenidos de las bases de datos de SIBICA (Sistema de Información de Bienes Inmuebles de la Alcaldía de Cali), Catastro y Rentas. Estos datos fueron descargados y cargados a una base de datos relacional para su posterior análisis comparativo. El propósito es verificar la consistencia entre las bases de datos, identificar las coincidencias y discrepancias entre los registros, y finalmente, evidenciar la calidad de los datos presentes en los sistemas de gestión de predios en Cali.

Para ello, se realizaron varias consultas SQL con el fin de obtener los registros coincidentes y no coincidentes entre SIBICA, Catastro y Rentas. A continuación, se describen los resultados obtenidos de estas consultas, seguidos de un análisis profundo de las discrepancias encontradas.

## Metodología

Se utilizaron tres bases de datos principales para este análisis:

- **SIBICA**: Contiene la información de los predios de bienes inmuebles de uso público que gestiona el Distrito de Cali.
- **Catastro**: Contiene los registros oficiales de los predios de la ciudad, se segmento la información para aquellos registros que están clasificados como bienes propiedad del Distrito.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- **Rentas:** Solo se suministró un dato: el código del predio (id_predio) de la información que maneja la dependencia de ventas asociada a los bienes de uso público.

Se realizaron las siguientes consultas clave:

1. **Coincidencias entre SIBICA y Catastro**: Identificación de registros que coinciden entre ambas bases de datos.
2. **Registros de SIBICA sin coincidencia en Catastro**: Registros que están presentes en SIBICA pero no en Catastro.
3. **Registros de Catastro sin coincidencia en SIBICA**: Registros de Catastro que no tienen coincidencia en SIBICA.
4. **Coincidencias entre SIBICA y Rentas**: Identificación de registros coincidentes entre SIBICA y Rentas.
5. **Registros de SIBICA sin coincidencia en Rentas**: Predios presentes en SIBICA pero no en Rentas.
6. **Registros de Rentas sin coincidencia en SIBICA**: Predios registrados en Rentas pero no en SIBICA.

## Resultados del Análisis

A continuación, se describen los resultados de las consultas realizadas para cada comparación:

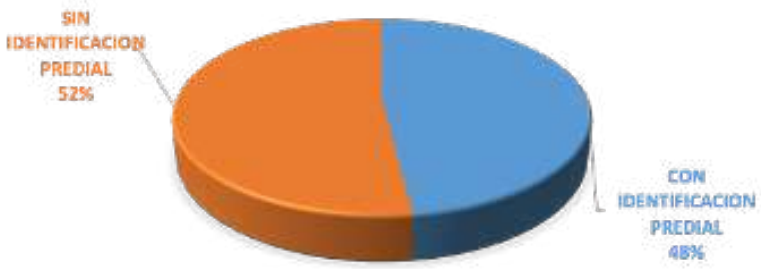
## Análisis 1. SIBICA vs CATASTRO.

Se encontraron varias coincidencias exactas entre los registros de SIBICA y Catastro. Estas coincidencias se basaron principalmente en la comparación del **id_catastro_p** en SIBICA y el **ID_PREDIO** en Catastro.

La base inicial de referencia de los datos del sistema SIBICA tiene 7.347 registros de los cuales 3.818 no cuentan con el dato id_predio y el numero predial es un código temporal que solo existe en el SIBICA. Esto indica que el **52%** de los registros de bienes inmuebles que tiene el SIBICA no existen en la base de datos de catastro. Lo anterior se debe a que a pesar de que estos predios tienen toda su información en regla frente a los procesos de la UAEGBS, no se ha realizado por parte de la dirección de catastro el registro y asignación de los números prediales, lo que permite la integración de estos datos al proceso catastral y con ello la consistencia de la información.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

**Figura 25. Porcentajes sobre un universo de 7.347 registros que tiene el SIBICA**




Con los 3.529 predios restantes **(48%)** que si tienen número catastral se procedió a realizar el cruce de datos entre estas dos fuentes de información. En este análisis se encontró lo siguiente:

- 3.119 predios cruzan perfectamente entre las dos fuentes de datos usando como criterio de comparación los campos**: id_catastro_p** y **id_predio.** Eso se traduce en el **42,4%** de exactitud del universo del SIBICA frente a CATASTRO teniendo como criterio los campos mencionados.

- Sin embargo, de esta cantidad de datos donde el atributo **id_predio** es igual, encontramos que 703 registros no cuentan con el campo **num_predial** en la fuente de datos de catastro, pero si tienen información en SIBICA en el campo **identifica_p**. Esto claramente evidencia una desarticulación entre las dos dependencias y sus sistemas de información.

- Adicionalmente se evidenciaron 151 registros donde si bien existe información de numero predial en ambas fuentes de información el dato no es igual.

- Lo anterior se puede apreciar en la siguiente imagen, donde el **ID_PREDIO** de la data de CATASTRO es igual al **id_catastro_p** del SIBICA, pero el numero predial no es consistente.

**Figura 26. ID PREDIO_ID CATASTRO**

| CATASTRO | | SIBICA | |
|---|---|---|---|
| ID_PREDIO | NUM_PREDIAL | id_catastro_p | identifica_p |
| 618377 | P039800010001 | 618377 | G024800010000 |

Finalmente podemos concluir que solo el **31% (2.249)** de los datos que tiene el SIBICA tienen consistencia total con los datos de CATASTRO. Esto nos deja frente a una necesidad relevante de validar desde el proceso de catastro la información que se gestiona en el SIBICA, para ello será fundamental la etapa 2 de

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

este proyecto que en la cual se busca tener una participación importante de las demás dependencias que tienen relación con los bienes inmuebles que administra la UAEGBS.

## Análisis 2. CATASTRO vs SIBICA.

Los datos iniciales suministrados por la dependencia de Catastro y que se utilizaron para la comparación, cuentan con 22.183 registros que tienen como propietario (campo: PROPIETARIO) al municipio de Cali o a entidades o dependencias adscritas al mismo.

En esta primera instancia ya evidenciamos una diferencia importante frente a los datos del SIBICA que corresponden a 7.347 registros, de predios que son responsabilidad del municipio, lo que nos adelanta un panorama complejo en cuanto a la consistencia de datos.

En ese orden, la diferencia en totalidad de registros entre Catastro y SIBICA es de 14.836 predios (de más en catastro), lo que nos deja con una realidad de solo el 33% de consistencia en lo referente a la totalidad de los datos.

Ahora bien, si consideramos que en SIBICA tenemos un 52% de predios con numero predial temporal, es decir que aún no han tenido la mutación catastral respectiva, esto nos deja frente a un panorama que evidencia una brecha más grande entre Catastro y SIBICA. Las cifras serían las siguientes:

### Tabla 16. Diferencia real. (a – c)

| PREDIOS TOTALES CATASTRO. (A) | PREDIOS TOTALES SIBICA. (B) | PREDIOS SIBICA CON NUMERO PREDIAL REAL. (C) | DIFERENCIA REAL. (A – C) |
|---|---|---|---|
| 22.183 | 7.347 | 3.529 | 18.654 |

El cuadro anterior nos muestra que la diferencial real entre SIBICA y Catastro es de 18.654 registros al excluir del análisis los predios con numero predial temporal, esto significa que solo el 16% de datos del catastro cruzan contra SIBICA.

## Análisis 3. SIBICA vs RENTAS.

Los datos entregados por la dependencia de rentas no permiten demasiado análisis, solo suministraron un campo: id_predio. Independiente de esto, se realizó una comparación de los datos para determinar la consistencia entre las dos fuentes de información. En este ejercicio encontramos lo siguiente:

La información de rentas cuenta con 7.522 registros, lo cual es un dato más cercano al total de predios con que cuenta el SIBICA (7.347).

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

Sin embargo y a pesar de la poca diferencia en la totalidad de los registros, al realizar la comparación de datos usando el campo id_predio de la fuente de datos de rentas y el campo id_catastro_p del SIBICA, solo 228 registros cruzan totalmente, lo que corresponde a un ínfimo 3% de datos totalmente consistentes.

## CONCLUSIONES DEL CRUCE DE INFORMACION:

El análisis integral de las bases de datos de SIBICA, Catastro y Rentas demuestra la existencia de inconsistencias estructurales que afectan la confiabilidad y trazabilidad de la información predial del Distrito de Santiago de Cali. Los hallazgos más relevantes permiten concluir lo siguiente:

1. Más de la mitad de los registros del SIBICA (52%) no cuentan con número predial oficial, lo que limita la integración natural con el proceso catastral y evidencia rezagos en las mutaciones y registros que Catastro debe formalizar.
2. Aunque 3.119 predios (42,4%) tienen coincidencia inicial entre ID predial de SIBICA y Catastro, solo 2.249 predios (31% del total SIBICA) presentan consistencia completa entre ambas fuentes, lo que revela una desconexión entre registros administrativos y registros catastrales.
3. La visión inversa (Catastro → SIBICA) muestra una discrepancia aún mayor: solo 16% de los predios en Catastro encuentran correspondencia en SIBICA, lo que indica que Catastro mantiene registros del Distrito que no están integrados a la gestión inmobiliaria de la UAEGBS.
4. El cruce con Rentas confirma la falta de integración interdependencias: solo 3% de los registros coinciden plenamente, afectando directamente procesos de recaudo, liquidación tributaria y control fiscal.
5. Los resultados evidencian la necesidad de fortalecer la gobernanza de datos prediales, estableciendo reglas claras de captura, actualización, homologación y validación entre Catastro, UAEGBS, Hacienda y Planeación.

## 4.3. Comparación entre Registros de Catastro y Construcción

El presente apartado explica los resultados del análisis que corresponden a los registros catastrales (20.983 entradas) y los registros de construcciones (653 entradas), utilizando criterios de vinculación basados en código nacional, código predial e ID catastral. Se evaluó el grado de integración entre ambas bases de datos y se identificaron brechas relevantes en la actualización y completitud de los registros.

### 4.3.1 Resultados de la Vinculación de Datos

Se ejecutó una comparación directa de las bases a partir de tres claves principales: código nacional ("codigonal"), código predial y el identificador único del catastro. Los hallazgos principales son los siguientes:

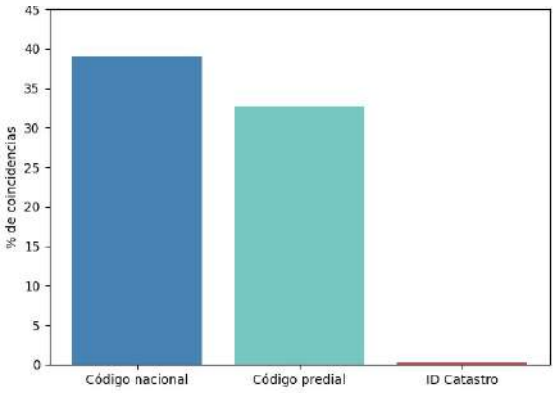
Coincidencia por código nacional: Se identificaron 255 registros coincidentes, lo que representa aproximadamente el 39% de los registros de construcciones y apenas el 1,2% de los registros catastrales. Esto evidencia una baja cobertura en la integración de ambos sistemas bajo este criterio.

Coincidencia por código predial: La comparación arrojó 214 coincidencias (33% sobre construcciones; 1% sobre el catastro), denotando una situación similar de falta de correspondencia.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

Coincidencia por ID catastral: El emparejamiento por este identificador único mostró solo 2 coincidencias (0,3% de construcciones), lo cual resulta especialmente preocupante, ya que refleja deficiencias graves en el enlace directo entre cada construcción y su predio respectivo. Esta anomalía puede deberse a vacíos en los datos, valores nulos o problemas en el diligenciamiento de los campos clave.

**Figura 27. Porcentaje de coincidencias por método vs total construcciones**




El universo de coincidencias únicas, excluyendo duplicados entre criterios, oscila entre 300 y 350 enlaces, es decir, entre el 46% y el 54% de las construcciones cuentan con algún vínculo validado al catastro. El resto queda fuera de correspondencia formal, lo que sugiere inconsistencias en la calidad de los datos o ausencia de actualización en los registros.

### 4.3.2. Actualización Temporal de los Registros de Construcción

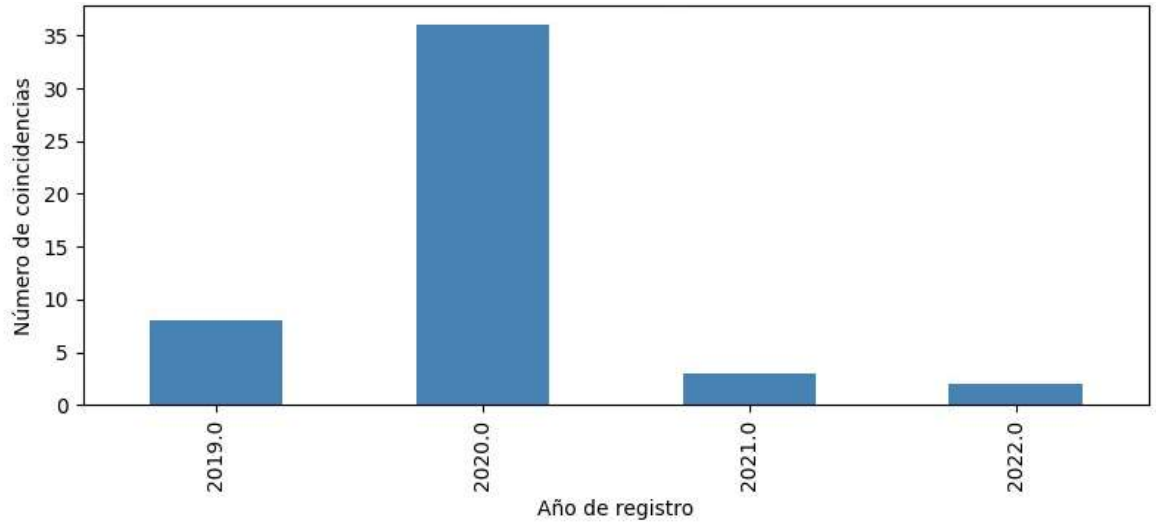
Se evaluaron las fechas de registro y de modificación asociadas a las construcciones vinculadas, considerando las siguientes variables:

Fechas de registro: Los datos disponibles oscilan entre el 14 de junio de 2019 y el 15 de julio de 2022. No obstante, cerca del 70% de los registros carecen de fecha, lo que dificulta establecer un control efectivo sobre la actualización de la información. La mayor concentración de registros fechados corresponde a los años 2019 y 2020, con muy pocas adiciones posteriores, indicando una posible detención en la incorporación de nuevas obras desde mediados de 2022.

Fechas de modificación: Los registros analizados presentan fechas entre el 3 de abril de 2019 y el 2 de agosto de 2023, pero el 90% de los registros no reportan actualizaciones. Se observa un patrón de cambios aislados, principalmente entre 2020 y 2021, con ausencia de modificaciones recientes a partir de agosto de 2023, lo cual puede denotar falta de mantenimiento en la base de datos.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

**Figura 28. Construcciones por año de registro (coincidencias)**




Ambos hallazgos apuntan a un rezago significativo en la actualización y seguimiento de la información de construcciones, lo que afecta la fiabilidad del sistema para fines de planeación y gestión territorial.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

# 5. RECOMENDACIONES

## 5.1 Estrategia de Sincronización y Mejora de Sistemas Prediales

### Corto plazo

- **Realizar jornadas conjuntas de depuración entre Catastro y UAEGBS**: Establecer mesas de trabajo para identificar y resolver inconsistencias en los registros prediales compartidos
- **Validar y actualizar los números prediales temporales del SIBICA**: Realizar verificación masiva de los NPH temporales y gestionar su formalización en Catastro
- **Formalizar cronograma de sincronización de datos entre dependencias**: Definir periodicidad y responsables para el intercambio de información actualizada

### Mediano plazo

- **Implementar un Modelo de Datos Maestro Predial (MDM)**: Establecer una única fuente de verdad para la información predial del Distrito, con reglas claras de actualización y priorización de fuentes
- **Automatizar cruces semanales con reglas de calidad de datos (DQ Rules)**: Desarrollar procesos automáticos de validación cruzada entre sistemas que identifiquen inconsistencias en tiempo cercano al real
- **Crear un diccionario unificado de atributos prediales del Distrito**: Homologar definiciones, formatos y catálogos de valores entre Catastro y SIBICA

### Largo plazo

- **Implementar un Sistema Unificado de Información Predial para el Distrito**: Desarrollar la etapa 2 del proyecto que integre de forma nativa información catastral, administrativa y operativa
- **Integrar fuentes externas**: Incorporar datos de Planeación, Obras Públicas, EAAB y otros sistemas institucionales relevantes
- **Implementar un módulo transaccional para mutaciones**: Desarrollar funcionalidades que permitan actualizar cambios prediales en tiempo real con trazabilidad completa

## 5.2 Control de Calidad de Datos

Para garantizar la sostenibilidad de la calidad de datos en el sistema SIBICA:

### 5.2.1 Implementación de Triggers para Integridad de Datos

¿Qué son los Triggers y cómo funcionan?

Los **triggers** (o disparadores) son objetos de base de datos que se ejecutan automáticamente en respuesta a eventos específicos sobre una tabla o vista. Funcionan como "guardianes automáticos" que se activan

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

antes o después de operaciones de inserción (INSERT), actualización (UPDATE) o eliminación (DELETE) de datos.

### Mecánica de funcionamiento:

1. **Evento desencadenante**: Un usuario o aplicación intenta insertar, modificar o eliminar un registro en una tabla.
2. **Activación automática**: El trigger asociado a esa tabla se ejecuta automáticamente, sin requerir intervención manual.
3. **Validación o acción**: El trigger ejecuta las reglas de negocio programadas (validaciones, transformaciones, auditorías).
4. **Decisión**: Basado en las validaciones, el trigger puede:
    a. **Permitir la operación** si los datos cumplen con todos los criterios
    b. **Rechazar la operación** y generar un mensaje de error específico
    c. **Modificar los datos** antes de que sean almacenados (por ejemplo, convertir texto a mayúsculas)
    d. **Registrar la operación** en tablas de auditoría

### Ejemplo práctico de funcionamiento:

**Usuario intenta insertar:**
**fecha_construccion = "15/13/2024" (mes inválido)**
**→ Trigger se activa ANTES de la inserción**
**→ Valida formato y rango de fecha**
**→ Detecta mes inválido (13)**
**→ RECHAZA la operación**
**→ Retorna mensaje: "Error: Fecha inválida. Use formato dd-mm-yyyy con mes válido (01-12)"**
**→ El registro NO se inserta en la base de datos**

### 5.2.2. Importancia Crítica de Implementar Triggers

### 1. Garantía de Integridad desde el Origen

Los triggers actúan como la última línea de defensa a nivel de base de datos, asegurando que **ningún dato incorrecto llegue a ser almacenado**, independientemente de:

- La aplicación o sistema que realice la inserción
- Errores en validaciones de formularios web o móviles
- Importaciones masivas de datos desde archivos externos
- Accesos directos a la base de datos por parte de administradores

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

Esto es crucial porque las validaciones en formularios pueden ser eludidas o deshabilitadas, pero los triggers **siempre se ejecutan**.

### 2. Consistencia Transversal en Múltiples Sistemas

En entornos donde varios sistemas o aplicaciones acceden a la misma base de datos, los triggers garantizan que **todas las fuentes respeten las mismas reglas de negocio**, sin necesidad de replicar la lógica de validación en cada aplicación cliente.

### 3. Reducción de Costos de Limpieza de Datos

Implementar triggers **previene la contaminación de datos** desde el inicio, lo cual es significativamente más eficiente que:

- Ejecutar procesos periódicos de limpieza de datos
- Realizar correcciones masivas de registros históricos
- Investigar y corregir inconsistencias detectadas en reportes
- Lidiar con decisiones basadas en datos incorrectos

### 4. Automatización Total sin Intervención Humana

Los triggers operan 24/7 sin requerir supervisión humana, eliminando:

- Errores de captura manual
- Dependencia de la capacitación del personal
- Inconsistencias por cambios de operadores
- Tiempo de respuesta ante incumplimientos

### 5. Trazabilidad y Auditoría Automática

Los triggers pueden configurarse para registrar automáticamente:

- Quién intentó ingresar datos incorrectos
- Qué validaciones fallaron
- Cuando ocurrió el intento
- Historial de cambios en registros críticos

Esto es especialmente importante para cumplimiento normativo y auditorías gubernamentales.

### 5.2.3. Aplicación Específica en el Sistema SIBICA

Se recomienda implementar triggers para los siguientes casos críticos:

### A. Validación de Campos de Fechas

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

Implementar triggers que verifiquen:

- **Formato estricto dd-mm-yyyy**: Rechazar formatos como "2024-12-15", "15/12/24", "Dec 15, 2024"
- **Rangos lógicos**:
  - Día entre 01-31 (considerando el mes correspondiente)
  - Mes entre 01-12
  - Año dentro de rangos permitidos (por ejemplo, construcciones no anteriores a 1800 ni posteriores al año actual)
- **Coherencia temporal**:
  - Fecha de construcción no puede ser posterior a fecha de actualización
  - Fecha de escrituración no puede ser anterior a fecha de construcción
- **Fechas nulas controladas**: Permitir NULL solo en campos opcionales definidos

### Ejemplo de validación:

```
SI fecha_construccion > fecha_actual ENTONCES
  RECHAZAR con mensaje "La fecha de construcción no puede ser futura"
FIN SI
```

### B. Validación de Direcciones Estandarizadas

Los triggers deben verificar:

- **Nomenclatura oficial**: Solo permitir prefijos CL (Calle), KR (Carrera), TV (Transversal), DG (Diagonal), AV (Avenida), AC (Autopista), CIR (Circular)
- **Estructura completa**: Validar que contengan todos los componentes requeridos (tipo vía, número, complemento)
- **Formato numérico**: Verificar que los números de vía y numeración sean valores numéricos válidos
- **Evitar abreviaturas no estándar**: Rechazar variaciones como "CAL", "CARR", "CALLE", "K"

### Ejemplo de validación:

```
SI dirección NO INICIA CON (CL|KR|TV|DG|AV|AC|CIR) ENTONCES
  RECHAZAR con mensaje "La dirección debe iniciar con nomenclatura estándar (CL, KR, TV, DG, AV, AC, CIR)"
FIN SI
```

### C. Validación de Tipos de Equipamiento y Valores Categóricos

Implementar triggers que:

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- **Verifiquen contra catálogos maestros**: Cada valor en campos como tipo_equipamiento, zonas_p, actualiza_sap debe existir en tablas de catálogo predefinidas
- **Rechacen valores en texto libre**: Evitar variaciones como "Comercial", "comercial", "COMERCIAL", "Com.", cuando solo "COMERCIAL" es válido
- **Prevengan valores obsoletos**: Impedir inserción de categorías que han sido descontinuadas
- **Controlen valores NULL**: Permitir o rechazar según la criticidad del campo

**Ejemplo de validación:**

**SI tipo_equipamiento NO EXISTE EN tabla_catalogo_equipamiento ENTONCES**
**RECHAZAR con mensaje "Tipo de equipamiento inválido. Consulte catálogo autorizado"**
**FIN SI**

### D. Validaciones Cruzadas Entre Tablas

Los triggers pueden validar relaciones entre diferentes tablas:

- Una construcción debe estar asociada a un terreno existente
- Un predio no puede tener área de construcción mayor al área del terreno
- La suma de áreas de construcciones no puede exceder el área del lote

### E. Validación de Rangos Numéricos

Para campos como áreas, avalúos, coordenadas:

- Valores no negativos
- Rangos coherentes (área_construida <= area_lote)
- Validación de separadores decimales (. vs ,)
- Límites máximos y mínimos razonables

Esta capa de validación a nivel de base de datos mediante triggers es **fundamental y no opcional** para garantizar que, independientemente del sistema o formulario desde donde se capture la información, los datos almacenados cumplan siempre con los estándares de calidad establecidos. Los triggers representan una inversión inicial que genera beneficios continuos y permanentes en la calidad y confiabilidad de la información del sistema SIBICA.

## 5.2.4 Validaciones Automáticas de Entrada

- **Establecer reglas de validación en los formularios de captura**: Asegurar que las direcciones sigan el formato estándar (CL, KR, TV, DG), las fechas cumplan el formato dd-mm-yyyy, y los valores categóricos estén dentro de listas controladas.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

- **Implementar máscaras de entrada**: Utilizar máscaras en los campos de captura que guíen al usuario en el formato correcto, especialmente para fechas y direcciones, reduciendo errores humanos en el proceso de digitación.

### 5.2.5 Estandarización de Reglas de Captura en los Sistemas Origen

- **Crear catálogos desplegables**: Para campos como tipo_equipamiento, zonas_p, actualiza_sap, evitando entrada libre de texto que pueda generar inconsistencias.
- **Unificar interfaces de captura**: Garantizar que todos los puntos de entrada de información al sistema utilicen los mismos controles y validaciones.

### 5.2.6 Proceso de Verificación Post-Recopilación

**Es fundamental establecer un protocolo de verificación obligatoria** una vez finalizado el proceso de recopilación de información y antes de la carga definitiva al sistema SIBICA:

- **Responsabilidad del Administrador de SIBICA**: Se debe designar formalmente al encargado del sistema SIBICA como responsable de la verificación y validación final de toda la información recopilada antes de su carga al sistema productivo.
- **Checklist de verificación**: El responsable debe ejecutar una lista de verificación que incluya:
  - Validación de completitud de campos obligatorios
  - Verificación de formatos estándar en fechas, direcciones y campos categóricos
  - Revisión de consistencia entre tablas relacionadas (construcciones-terrenos, predios-propietarios)
  - Validación de rangos numéricos (áreas, valores catastrales)
  - Verificación de duplicados o registros inconsistentes
- **Registro de auditoría**: Cada proceso de verificación debe quedar documentado, indicando fecha, responsable, cantidad de registros verificados y hallazgos identificados.
- **Beneficios del proceso**: Esta verificación centralizada garantiza un mejor control sobre la calidad de la información, permite detectar errores sistemáticos antes de que impacten el sistema productivo, y mantiene la integridad y confiabilidad de los datos en SIBICA.

### 5.2.7 Auditorías y Monitoreo Continuo

- **Definir responsables de revisión periódica**: Asignar al equipo técnico de la UAEGBS la responsabilidad de auditorías trimestrales de calidad de datos.
- **Implementar alertas de inconsistencias**: Configurar notificaciones automáticas cuando se detecten valores fuera de rango o formatos no estándar.
- **Métricas de calidad**: Establecer indicadores cuantitativos (porcentaje de registros con errores, tiempo de corrección, etc.) para medir la efectividad del control de calidad.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

## 5.3 Automatización del Proceso

Se recomienda como propuestas para la automatización del proceso, lo siguiente:

### 5.3.1 Desarrollo de Scripts de Limpieza Periódica

- **Crear rutinas en SQL**: Que ejecuten automáticamente las transformaciones de estandarización identificadas (direcciones, fechas, categorías).
- **Programación de ejecución**: Establecer frecuencias definidas (diaria, semanal) para la ejecución de estos scripts según el volumen de transacciones.
- **Logs de ejecución**: Mantener registros detallados de cada ejecución, incluyendo cantidad de registros procesados y correcciones aplicadas.

### 5.3.2 Implementación de Pipeline ETL

- **Diseñar un flujo ETL robusto**: Que valide, limpie y cargue datos antes de su ingreso definitivo al sistema SIBICA.
- **Capas de validación**: Implementar validaciones en múltiples etapas (extracción, transformación, carga) para detectar errores lo antes posible en el proceso.
- **Manejo de excepciones**: Definir protocolos claros para el manejo de registros que no pasen las validaciones, incluyendo cuarentena de datos y flujos de corrección.

### 5.3.3 Monitoreo Continuo de Calidad

- **Implementar dashboards**: Con indicadores de calidad de datos actualizados en tiempo real, que permitan visualizar el estado de la información.
- **Alertas proactivas**: Configurar sistemas de alerta temprana que notifiquen cuando los indicadores de calidad caigan por debajo de umbrales aceptables.

### 5.3.4 Versionado de Datos

- **Establecer un sistema de control de versiones**: Para rastrear cambios históricos en los registros prediales, permitiendo auditoría completa y reversión de cambios si es necesario.
- **Trazabilidad completa**: Registrar quién, cuándo y por qué se realizó cada modificación en los datos.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

### 5.3.5 Estudio de Impacto y Mejora Continua

Se recomienda realizar un **estudio periódico** (semestral o anual) que incluya:

- **Análisis de eficiencia del proceso**: Evaluación de tiempos de procesamiento, cuellos de botella y oportunidades de optimización.
- **Evaluación de efectividad de controles**: Medición de la cantidad y tipo de errores detectados por cada mecanismo de validación implementado.
- **Análisis de patrones de error**: Identificación de errores recurrentes para ajustar procesos de capacitación y validaciones automáticas.
- **Benchmarking**: Comparación con mejores prácticas de otras entidades similares para identificar oportunidades de mejora.
- **Retorno de inversión**: Cuantificación de los beneficios obtenidos por la implementación de controles automatizados versus el esfuerzo invertido.
- **Actualización de procedimientos**: Revisión y actualización de los procedimientos operativos basados en lecciones aprendidas y cambios normativos o tecnológicos.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|

## 6. Anexos

Documentación complementaria del proceso de limpieza de datos:

1. **Catálogo de objetos geográficos:** Documento técnico de referencia de la UAEGBS que define la estructura, tipos de datos y restricciones de cada atributo
2. **Scripts utilizados:** Código fuente en SQL para validación de tipos de datos, estandarización de formatos y generación de reportes de calidad
3. **Análisis de información de predios (Modulación matriz inicial):** Documento detallado con el diagnóstico campo por campo de la base de datos SIBICA.
4. **Calidad del bien null:** Documento que contiene el detalle de los 67 registros con valores nulos en el atributo id_cb_fk, los cuales requieren actualización para garantizar la integridad de la información y la correcta clasificación de los predios.
5. **Comparación entre tablas:** En el presente directorio se encuentran los archivos derivados del sistema de información SIBICA, los cuales contienen datos extraídos mediante consultas y procesos de vinculación realizados en SQL. Dichos cruces se efectuaron utilizando diversos atributos de la tabla "prediomunicipio", en combinación con otras tablas como "catastro" y "rentas", con el propósito de consolidar información relevante para el análisis y la gestión catastral.

| Copyright © Arquitecsoft® - Todos los Derechos Reservados | NC: | Uso Interno |
|---|---|---|