9/15/2026

Modelo de Datos

Complemento del documento técnico del modelo de datos + propuesta de arquitectura escalable

Un-named
[COMPANY NAME]

DOCUMENTO TÉCNICO

Modelo de Datos

Geoportal Verde — Secretaría de Bienestar Verde, Gobernación de Cundinamarca

Contrato No. SBV-MC-PS-192-2026

Elaborado por: UNBIT LTDA · Tulis Estela Llorente, Director de Proyecto

Fecha: 13 de agosto de 2026

Basado en el Informe de Control de Calidad de Datos Geoespaciales v1.7 (94 capas revisadas) y en el Diagnóstico Técnico del Geovisor Ambiental v2.2, elaborados previamente dentro de este mismo contrato.

CONTROL DE VERSIONES

Versión	Fecha	Descripción	Elaboró
1.0	13/08/2026	Análisis y definición del modelo de datos	Claudia Berenicce Rojas Rincón
2.0	26/08/2026	Modificación del documento de acuerdo a las observaciones enviadas	Claudia Berenicce Rojas Rincón

TABLA DE CONTENIDO

1. Introducción y propósito 3
2. Metodología 3
3. Modelo de datos actual (AS-IS) 3
3.1 Sistemas de referencia en uso 4
3.2 Patrones de identificador único encontrados 4
3.3 Esquemas recurrentes por familia 5
4. Diagnóstico de problemas estructurales 5
4.1 Ausencia de un estándar de identificador único 5
4.2 Sistema de referencia no estandarizado 6
4.3 Corrupción de codificación de texto 6
4.4 Ausencia total de metadatos 6
4.5 Geometría sin control de calidad antes de publicar 6
5. Modelo de datos estandarizado propuesto (TO-BE) 6
5.1 Estándar de sistema de referencia 6
5.2 Estándar de identificador único 7
5.3 Estándar de nomenclatura 7
5.4 Estándar de metadatos obligatorios 7
5.5 Diccionario de datos y control de duplicados 7
6. Hoja de ruta de implementación 8
7. Conclusión 8

1. Introducción y propósito

Este documento consolida, en un modelo de datos formal, todo lo aprendido durante el control de calidad geoespacial (QA/QC) de las 94 capas del Geoportal Verde y el diagnóstico técnico del Geovisor Ambiental actual, ambos elaborados dentro de este mismo contrato. No parte de cero: cada patrón, cada estándar y cada **recomendación** que aparece aquí está respaldado por un hallazgo verificado sobre datos reales, no por buenas prácticas genéricas de literatura SIG.

El objetivo es doble: (1) documentar formalmente el modelo de datos tal como existe hoy (AS-IS) — sus esquemas, convenciones de identificador, sistemas de referencia y problemas estructurales recurrentes — y (2) proponer un modelo de datos estandarizado y una arquitectura técnica escalable (TO-BE) que resuelva esos problemas de raíz antes de que el Geoportal Verde crezca más allá de las 94 capas actuales.

2. Metodología

- Se partió del inventario completo y ya corregido de las 94 capas del QA/QC — nombres de campo, tipos de dato, sistemas de referencia, identificadores designados y hallazgos documentados capa por capa.
- Los 11 hallazgos transversales del informe QA/QC (numeral 3, versión 1.7) se reclasificaron por tipo de problema estructural (identificador, geometría, codificación, metadatos, rendimiento) para identificar patrones de causa raíz, no solo síntomas.
- El diagnóstico del Geovisor Ambiental (estructura del Web Map, servicios publicados, duplicados de catálogo en ArcGIS Online) se usó para entender cómo el modelo de datos actual se traduce — o se pierde — en la capa de publicación.

3. Modelo de datos actual (AS-IS)

Las 94 capas revisadas se agrupan en 15 familias temáticas, cada una con su propio esquema de campos, pero sin un estándar común entre familias. La siguiente tabla resume el inventario:

Familia temática	Capas	Fuente / origen
Negocios Verdes (economía sostenible por municipio)	10	Shapefiles sueltos, esquema MpCodigo/CATEGORIA/SUBCATEGOR
Serie Nodo (economía circular)	9	Shapefiles sueltos, un polígono por nodo
Deforestación (Hansen / IDEAM)	2	Shapefile único con 2 capas
Bosques y cobertura de la tierra (Corine Land Cover)	3	Shapefile único con 3 capas
RUNAP — áreas protegidas (DRMI, DCS, PNN, PNR, RNSC, RFPR, RFPN)	7	Shapefiles sueltos, esquema IDPNN/AP_ID/APID

POMCA — planes de ordenación de cuencas (10 cuencas)	10	Shapefiles sueltos, uno por cuenca
Geodatabase climática GDB_PIGCC	33	Única geodatabase (.gdb) del lote
Recursos hídricos (PUEAA, bocatomas, hidrografía)	7	Shapefiles sueltos y un zip con 3 capas
Geología	2	Shapefile único con 2 capas
Límites administrativos (departamento, provincias, municipios)	3	Shapefiles sueltos
Individuales (Biomas, Ecosistemas, Páramos, Predios, Curvas de nivel, Títulos mineros, Territorios CARs)	8	Shapefiles sueltos, un tema cada uno

En total, 61 capas llegaron como shapefiles individuales (uno por archivo .zip), 33 como parte de una única geodatabase, y varias 'familias' comparten un solo archivo fuente con múltiples capas internas — una mezcla de formatos y granularidades que ya es, en sí misma, el primer síntoma de la ausencia de un modelo de datos unificado.

3.1 Sistemas de referencia en uso

Se identificaron cinco sistemas de referencia distintos coexistiendo en el lote, sin un estándar declarado:

EPSG	Descripción	Uso observado
4686	MAGNA-SIRGAS geográfico (grados)	La mayoría de shapefiles sueltos de origen departamental
9377	MAGNA-SIRGAS 2018 / Origen-Nacional (proyectado, metros)	Familia POMCA, familia Nodo, capas de la geodatabase PIGCC
3857	Web Mercator (proyectado, metros, distorsión fuera del ecuador)	Varias capas reproyectadas para publicación web; el propio Web Map del Geovisor está en este sistema
3116	MAGNA-SIRGAS / Bogotá zone Gauss-Krüger (proyectado, metros)	Al menos una capa (Predios adquiridos)
6247	MAGNA-SIRGAS / Bogotá urban grid (proyectado, metros)	Geologico_Permeabilidad_Cund — una malla local urbana usada para datos de todo el departamento

Esta dispersión no es cosmética: fue la causa raíz del hallazgo transversal 3.6 del informe QA/QC (campos de área desactualizados tras reproyección), presente en al menos tres capas donde el campo Shape_Area quedó calculado en un sistema de referencia distinto al que la geometría tiene hoy.

3.2 Patrones de identificador único encontrados

Patrón	Ejemplo real	Capas que lo usan

Código DANE compuesto (departamento + municipio)	COD_DEPART + COD_MUNICI → 25120	Territorios CARs, 4 capas municipales de la geodatabase, munPUEAA, Municipios del Departamento
OBJECTID / FID nativo del formato	OBJECTID, FID_	~20 capas, cuando el campo autogenerado no tenía duplicados
Identificador coalescido de tres campos	IDPNN, si es nulo AP_ID, si es nulo APID	Las 7 capas RUNAP
Identificador secuencial agregado en el QA/QC	ID_CAPA (1, 2, 3…)	~15 capas sin ningún campo confiable de origen
Código de clasificación (NO es identificador, aunque parece uno)	bioma, codigo (Corine Land Cover), COD_AH/COD_ZH (subzona hidrográfica)	Biomas, Bosques/Cobertura, ISH_CORTE, ENA — se repite por diseño

El último patrón es el más importante de entender: en al menos 6 capas, un campo con apariencia de identificador (nombre corto, valores numéricos o alfanuméricos) en realidad codifica una categoría temática que se repite legítimamente en decenas o cientos de registros.

3.3 Esquemas recurrentes por familia

Familia	Campos comunes	Patrón de nombres
Negocios Verdes / Economía Circular	MpCodigo, MpNombre, CATEGORIA, SUBCATEGOR, Proyeccion (siempre vacío)	CamelCase mixto, truncado a 10 caracteres (límite de shapefile/DBF)
RUNAP (áreas protegidas)	IDPNN, AP_ID, APID, NOMBRE, CATEGORIA, URL, HARES, ORG, FECHAACT	MAYÚSCULAS con guion bajo
POMCA	POMCA, CAT_ORD, ZON_USO_MA/SZON_USO_M, AREA_ha, PERIMETRO	MAYÚSCULAS abreviadas, inconsistentes entre cuencas (ZON_USO_MA vs SZON_USO_M)
Corine Land Cover / taxonomías jerárquicas	codigo, leyenda, nivel_1…nivel_6	snake_case, profundidad variable por categoría

4. Diagnóstico de problemas estructurales

Los 11 hallazgos transversales del informe QA/QC se agrupan en cinco categorías de causa raíz. Ninguno es un problema de una capa aislada — todos se repiten en múltiples familias, lo que confirma que son fallas del proceso de producción de datos, no errores puntuales de captura.

4.1 Ausencia de un estándar de identificador único

87% de los hallazgos críticos del QA/QC (numeral 3.2 de este documento) se debieron a la falta de un campo de identificador único diseñado a propósito en el momento de crear cada capa. El equipo que

produce los datos no tiene una convención — cada capa resuelve el problema de forma distinta (o no lo resuelve), dejando que quien la consume tenga que inferir cuál campo sirve como llave.

4.2 Sistema de referencia no estandarizado

Cinco sistemas de referencia en 94 capas (numeral 3.1) generan dos problemas concretos y verificados: (a) campos de área calculados antes de una reproyección quedan matemáticamente desactualizados sin que nada lo señale, y (b) al publicar todo junto en un único Web Map en Web Mercator (como hace el Geovisor actual), la mezcla de orígenes aumenta el riesgo de desalineación visual entre capas.

4.3 Corrupción de codificación de texto

Se documentaron dos mecanismos distintos de corrupción de tildes y eñes en nombres de categorías y lugares: uno reversible matemáticamente (cadena CP850 → CP1252 → UTF-8, numeral 3.5 del informe QA/QC), presente en al menos dos capas, y uno irreversible con reemplazo de carácter Unicode, presente en al menos dos capas más. Ambos apuntan a un proceso de captura o migración de datos que no fijó una codificación de caracteres consistente — no a errores de digitación.

4.4 Ausencia total de metadatos

Ninguna de las capas de origen shapefile trae metadatos formales (.xml) con origen, fecha, responsable o método de captura. Solo una capa de las 94 (Nodo del Agua) los tenía. Sin metadatos, cada nueva persona que hereda el dato tiene que reconstruir su procedencia por inferencia — exactamente el trabajo que este mismo QA/QC tuvo que hacer capa por capa.

4.5 Geometría sin control de calidad antes de publicar

Las 10 capas POMCA (numeral 3.8) llegaron con entre 31% y 100% de sus polígonos topológicamente inválidos por autointersección — un defecto que bloquea la publicación en ArcGIS Online y que ninguna herramienta de digitalización debería dejar pasar sin corregir. Esto confirma que no existe, hoy, un paso de validación geométrica antes de que una capa se considere lista para publicarse.

5. Modelo de datos estandarizado propuesto (TO-BE)

Cada uno de los cinco problemas del numeral 4 tiene un estándar concreto que lo resuelve. Se proponen como reglas obligatorias para toda capa nueva, y como criterio de aceptación para migrar una capa existente al nuevo Geoportal Verde.

5.1 Estándar de sistema de referencia

- Dato maestro (fuente de verdad): EPSG:9377 (MAGNA-SIRGAS 2018 / Origen-Nacional) para todo el departamento — es el estándar nacional vigente para Colombia

- Capa de publicación web: reproyección a EPSG:3857 (Web Mercator) se hace en el momento de publicar el servicio, nunca se guarda como si fuera el dato maestro. Esto elimina de raíz el problema de campos de área desactualizados (numeral 4.2).
- Toda capa nueva debe declarar su EPSG en la documentación de entrega; no se acepta un .prj sin código EPSG reconocible.

5.2 Estándar de identificador único

- Todo feature class debe tener un campo GlobalID o UUID generado en el momento de la creación, garantizado único por el propio software SIG, independiente del contenido temático.
- Adicionalmente, cuando exista un identificador de negocio oficial (código DANE, código RUNAP, código de resolución), se documenta explícitamente cuál es y se valida su unicidad antes de la entrega — no se asume.
- Los campos que codifican categoría (como 'bioma' o 'codigo' de Corine Land Cover) deben documentarse explícitamente como campos de clasificación, no de identificador, en el diccionario de datos (numeral 5.5).

5.3 Estándar de nomenclatura

- Nombres de capa y de campo en snake_case, sin tildes ni caracteres especiales, máximo 10 caracteres solo si el formato de entrega es shapefile — de lo contrario, nombres completos y descriptivos.
- Codificación de caracteres UTF-8 declarada y verificada en el archivo .cpg de cada shapefile antes de la entrega.
- Un mismo campo temático (ej. categoría, código de municipio) debe llamarse igual en todas las capas de una misma familia — hoy varía entre ZON_USO_MA y SZON_USO_M dentro de la misma familia POMCA.

5.4 Estándar de metadatos obligatorios

Toda capa se entrega con metadato FGDC o ISO 19115 embebido (origen, fecha de creación, fecha de última actualización, responsable, método de captura, escala de digitalización, sistema de referencia).

Se propone una plantilla mínima de 8 campos, generada automáticamente a partir de un formulario de entrega estandarizado, para no depender de que cada persona la diligencie manualmente.

5.5 Diccionario de datos y control de duplicados

Se recomienda mantener un diccionario de datos vivo (una tabla maestra, no un documento estático) con: nombre de capa, familia temática, esquema de campos, identificador designado, sistema de referencia, y estado de vigencia. Este diccionario es también el mecanismo de control de duplicados: antes de cargar una capa nueva a la geodatabase corporativa o a ArcGIS Online, se verifica contra el diccionario si ya existe una versión — evitando el patrón de duplicación por doble carga (numeral 3.1 del informe QA/QC, presente en 8 de las primeras 10 capas revisadas) y los duplicados de catálogo

encontrados dentro de la propia geodatabase PIGCC y en ArcGIS Online (Bosques_Cundinamarca_1, Acuiferos_1, COpPIA_MATRIZ COMPONENTE EDUCACIÓN AMBIENTAL, entre otros).

6. Hoja de ruta de implementación

1. Formalizar el diccionario de datos (numeral 5.5) con las 94 capas ya revisadas como primer contenido — no partir de cero, partir del trabajo ya hecho.

2. Definir y documentar los 5 estándares del numeral 5 como criterio de aceptación oficial, y aplicarlos retroactivamente a las 94 capas ya corregidas antes de migrarlas a la geodatabase corporativa.

3. Consolidar las 94 capas dispersas (shapefiles sueltos + geodatabase parcial) en la geodatabase corporativa única (numeral 6.1), resolviendo primero los duplicados de catálogo ya identificados.

4. Definir y aplicar el umbral de simplificación de geometría (numeral 6.2) a la familia POMCA antes de publicarla en su forma definitiva.

7. Conclusión

El modelo de datos actual del Geoportal Verde funciona, pero funciona por el esfuerzo manual invertido en este control de calidad, no por diseño: 94 capas en 5 sistemas de referencia distintos, con al menos 5 patrones de identificador diferentes y ningún metadato formal, es un sistema que no escala sin repetir el mismo trabajo de investigación capa por capa cada vez que crece.