

**Desarrollo de una base de datos hidrogeológica para el grupo de aguas subterráneas del
Servicio Geológico Colombiano**

Dayana Isabel Manotas Ávila

Asesor

Julio Eduardo Mejía Manzano

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI
Especialización en Ciencias de Datos y Analítica

2026

Resumen

El presente trabajo de grado tuvo como objetivo desarrollar una base de datos relacional sobre agua subterránea para el Servicio Geológico Colombiano (SGC), con el fin de facilitar la organización, gestión, el acceso y la calidad de los datos hidrogeológicos. La propuesta surge ante la identificación de problemas asociados a la dispersión, heterogeneidad y falta de estandarización de los datos provenientes de inventarios y campañas de caracterización hidrogeológica, lo cual dificulta su aprovechamiento técnico, institucional y analítico.

El proyecto se desarrolló mediante la metodología en cascada, estructurando el proceso en tres fases principales: diseño conceptual y lógico, consolidación de datos e implementación y verificación. En la fase de diseño se construyó un modelo entidad-relación compuesto por 22 tablas organizadas en seis esquemas temáticos, aplicando formas normales, restricciones de integridad y dominios controlados para asegurar la coherencia y calidad de la información. En la fase de consolidación se integraron datos provenientes de diferentes formatos, aplicando procesos de extracción, transformación y depuración que permitieron corregir inconsistencias y estandarizar categorías. Finalmente, la base de datos se implementó en un Sistema de Gestión de la Bases de Datos (SGBD) relacional con capacidades espaciales, incorporando reglas de negocio y validaciones técnicas orientadas al contexto hidrogeológico.

Como aplicación de la base de datos desarrollada, se realizó un Análisis Exploratorio de Datos Espaciales que permitió describir la distribución espacial de los puntos de agua subterránea y analizar parámetros fisicoquímicos como la conductividad eléctrica, el pH y la temperatura, evidenciando patrones espaciales y diferencias entre tipos de captación.

Los resultados evidencian que la base de datos es una herramienta de apoyo para la gestión del conocimiento hidrogeológico, fortaleciendo la capacidad del SGC, aportando una solución práctica y escalable acorde con las necesidades institucionales.

Palabras clave: aguas subterráneas, bases de datos, hidrogeología, análisis exploratorio de datos.

Abstract

The objective of this undergraduate thesis was to develop a relational groundwater database for the Colombian Geological Survey (SGC), in order to facilitate the organization, management, access to, and quality of hydrogeological data. The proposal arose from the identification of problems associated with data dispersion, heterogeneity, and lack of standardization in information derived from inventories and hydrogeological characterization campaigns, which hinder their technical, institutional, and analytical use.

The project was developed using the waterfall methodology, which allowed the process to be structured into three main phases: conceptual and logical database design, data consolidation, implementation and verification. During the design phase, an entity–relationship model composed of 22 tables organized into six thematic schemas was developed, applying normal forms, integrity constraints, and controlled domains to ensure data consistency and quality. In the consolidation phase, data from different formats were integrated through extraction, transformation, and cleansing processes, which made it possible to correct inconsistencies and standardize categories. Finally, the database was implemented in a relational database management system with spatial capabilities, incorporating business rules and validations oriented to the hydrogeological context.

As an application of the developed database, an Exploratory Spatial Data Analysis was conducted to describe the spatial distribution of groundwater points and to analyze physicochemical parameters such as electrical conductivity, pH, and temperature, revealing spatial patterns and differences among types of abstraction.

The results demonstrate that the database is a support tool for hydrogeological knowledge management, strengthening the SGC's capacity and providing a practical and scalable solution aligned with institutional needs.

Keywords: groundwater, database, hydrogeology, spatial analysis.

Tabla de Contenido

| | |
|--|----|
| Introducción | 11 |
| Justificación | 12 |
| Objetivos..... | 15 |
| Objetivo General..... | 15 |
| Objetivos Específicos | 15 |
| Marco de Referencia | 16 |
| Datos Hidrogeológicos | 16 |
| Servicio Geológico Colombiano (SGC) | 16 |
| Bases de Datos Relacional..... | 16 |
| Modelo Entidad-Relación..... | 17 |
| Sistemas de Gestión de Bases de Datos (SGBD) | 18 |
| Metodología en Cascada..... | 19 |
| Análisis Exploratorio de Datos Espaciales | 20 |
| Antecedentes en la Estandarización de Información Hidrogeológica | 21 |
| Desarrollo Metodológico | 24 |
| Diseño de la Base de Datos | 24 |
| Consolidación de Datos | 24 |
| Implementación de la Base de Datos..... | 25 |
| Análisis Exploratorio de Datos Espaciales | 25 |
| Resultados y Discusión | 27 |
| Diseño de la Base de Datos | 27 |
| Resultados Generales del Diseño..... | 28 |

| | |
|---|----|
| Diseño por Esquemas: Organización Temática | 32 |
| Consolidación de los Datos Hidrogeológicos..... | 33 |
| Fuentes y Características Iniciales de los Datos | 33 |
| Proceso de Extracción y Transformación | 34 |
| Identificación y Tratamiento de Errores e Inconsistencias | 35 |
| Resultado del Proceso de Depuración | 38 |
| Implementación de la Base de Datos..... | 40 |
| Elección del SGBD y Configuración del Entorno | 41 |
| Creación de Tablas, Dominios y Restricciones | 42 |
| Pruebas de Integridad y Consistencia | 43 |
| Análisis de Datos Espaciales | 48 |
| Distribución Espacial de los Puntos de Agua Subterránea en Colombia | 48 |
| Análisis Espacial Detallado del Departamento del Cesar..... | 50 |
| Análisis Estadístico Espacial de Parámetros Fisicoquímicos..... | 54 |
| Clasificación Espacial de la Calidad del Agua Según Conductividad Eléctrica | 58 |
| Conclusiones | 61 |
| Recomendaciones | 63 |
| Referencias Bibliográficas | 64 |
| Apéndices..... | 66 |

Lista de Tablas

| | |
|--|----|
| Tabla 1 <i>Principales Bases de Datos Hidrogeológicas en el Mundo</i> | 21 |
| Tabla 2 <i>Esquemas y Número de Tablas por Esquema</i> | 29 |
| Tabla 3 <i>Resumen de Errores e Inconsistencias Detectadas y Acciones Aplicadas Durante la Consolidación de los Datos</i> | 39 |
| Tabla 4 <i>Estadística Descriptiva de los Valores de Conductividad Eléctrica ($\mu\text{S/cm}$) en el Departamento del Cesar</i> | 54 |
| Tabla 5 <i>Estadística Descriptiva de los Valores de pH en el Departamento del Cesar</i> | 55 |
| Tabla 6 <i>Estadística Descriptiva de los Valores de Temperatura ($^{\circ}\text{C}$) en el Departamento del Cesar</i> | 55 |

Lista de Figuras

| | |
|--|----|
| Figura 1 <i>Modelo de Entidad Relación (ER)</i> | 28 |
| Figura 2 <i>Ejemplo Inserción Correcta</i> | 44 |
| Figura 3 <i>Prueba de Dominio</i> | 44 |
| Figura 4 <i>Prueba Restricción Rango Geográfico</i> | 45 |
| Figura 5 <i>Prueba Restricción Llave Foránea y Relación Municipio-Departamento</i> | 46 |
| Figura 6 <i>Prueba Llave Foránea cod_sgc</i> | 46 |
| Figura 7 <i>Prueba Unicidad Espacial</i> | 47 |
| Figura 8 <i>Prueba Unicidad de Punto</i> | 47 |
| Figura 9 <i>Distribución Espacial de los Puntos de Agua Registrados en la Base De Datos</i> | 49 |
| Figura 10 <i>Departamento con Mayor Cantidad de Registros</i> | 50 |
| Figura 11 <i>Cantidad de Registros por Municipio</i> | 51 |
| Figura 12 <i>Distribución de los Puntos en el Departamento del Cesar</i> | 53 |
| Figura 13 <i>Distribución de los Parámetros en Campo por Tipo de Punto</i> | 57 |
| Figura 14 <i>Clasificación del Agua por su Conductividad Eléctrica</i> | 59 |

Lista de Apéndices

Apéndice A *Modelo de Entidad-Relación* 66

Apéndice B *Diccionario de la Base de Datos*..... 66

Introducción

El Servicio Geológico Colombiano (SGC), como entidad responsable de generar y administrar el conocimiento geocientífico del país, produce un vasto volumen de información hidrogeológica. No obsta, esta información se ha almacenado históricamente en diversos formatos y estructuras, lo que dificulta su consulta, análisis y aprovechamiento.

En respuesta a esta problemática, el presente trabajo de grado propuso el desarrollo de una base de datos relacional sobre agua subterránea, empleando la metodología en cascada, orientada a la consolidación, estandarización y aseguramiento la calidad de la información hidrogeológica generada por el SGC. El diseño se fundamentó en principios de normalización, integridad referencial y control de calidad, incorporando además componentes espaciales que facilitan el análisis exploratorio de los datos.

Adicionalmente, se llevó a cabo un Análisis Exploratorio de Datos Espaciales (AEDE) a modo de validación funcional del sistema desarrollado, con el fin de evidenciar su potencial para identificar patrones espaciales y apoyar la interpretación hidrogeológica. De esta manera, el trabajo no solo aborda un reto técnico de gestión de datos, sino que, además contribuye al fortalecimiento institucional y a la generación de conocimiento útil para la gestión sostenible del agua subterránea en Colombia.

Justificación

El agua subterránea es un recurso estratégico para garantizar la seguridad hídrica, especialmente en un país como Colombia, donde las dinámicas climáticas, la variabilidad geográfica y la presión antropogénica comprometen su disponibilidad y calidad. Sin embargo, la ausencia de un sistema para gestionar información hidrogeológica restringe las capacidades de monitoreo, análisis y planificación del recurso.

El Servicio Geológico Colombiano (SGC) enfrenta el desafío de consolidar la información hidrogeológica dispersa y no estandarizada en un sistema accesible y confiable. La metodología en cascada permite el desarrollo ordenado y estructurado del proyecto, asegurando que cada fase se ejecute con precisión y alineada a los objetivos del SGC.

El desarrollo de esta base de datos se justifica desde los puntos de vista técnico, social y ambiental:

- Fortalecimiento de la gestión del conocimiento geocientífico. El diseño e implementación de la base de datos permitirá al SGC consolidar información sobre agua subterránea, incluyendo parámetros hidrogeológicos, fisicoquímicos e isotópicos. Esto no solo facilitará la consulta de registros históricos y datos actuales, sino que también impulsará la generación de conocimiento científico de alta calidad para atender necesidades locales y nacionales.
- Optimización de la toma de decisiones. Con una base de datos organizada y centralizada, los responsables de la política pública y gestión del recurso dispondrán de información oportuna y precisa para diseñar estrategias de manejo, conservación y aprovechamiento sostenible del agua subterránea. Esto incluye la identificación de áreas

vulnerables, la evaluación de impactos por actividades humanas y el monitoreo de la calidad y cantidad del recurso.

- Impacto social y prevención de conflictos. Al mejorar la disponibilidad y transparencia de la información hidrogeológica, se fomentará la transparencia y la confianza diversos grupos de interés (comunidades, autoridades ambientales e industrias). Esto contribuirá a la prevención de conflictos por el acceso y uso del agua subterránea, especialmente en regiones con alta dependencia del recurso.

- Contribución a la sostenibilidad ambiental. El proyecto permitirá identificar riesgos asociados a la sobreexplotación y contaminación de acuíferos, promoviendo la gestión sostenible del recurso hídrico. Asimismo, se alinea con los compromisos internacionales de Colombia en el marco de los Objetivos de Desarrollo Sostenible (ODS), especialmente el ODS 6 (Agua limpia y saneamiento).

- Adaptación al cambio climático. En un contexto de creciente incertidumbre climática, contar con una herramienta que integre y actualice información hidrogeológica es fundamental para fortalecer la resiliencia del país. La base de datos apoyará la implementación de estrategias de adaptación, como la gestión eficiente del agua en periodos de sequía y la protección de zonas estratégicas para la recarga de acuíferos.

- Transferencia tecnológica y fortalecimiento institucional. La aplicación de la metodología en cascada asegura un desarrollo estructurado del proyecto, que incluye fases críticas como la recolección de requisitos, el diseño de modelos de datos y la implementación tecnológica. Este enfoque fomenta la transferencia de conocimiento y la capacitación del personal del SGC, posicionándolo como un referente regional en la gestión de información hidrogeológica.

- Valor estratégico para el desarrollo nacional. Una base de datos hidrogeológica no solo beneficiará al SGC, sino que también fortalecerá las capacidades de otros actores, como investigadores, autoridades ambientales y entidades gubernamentales, promoviendo la planificación integral del recurso hídrico y el desarrollo sostenible del país.

En conclusión, el desarrollo de una base de datos sobre agua subterránea utilizando la metodología en cascada es una iniciativa estratégica que responde a una necesidad crítica de fortalecer la gestión de la información y el conocimiento geocientífico en el SGC. Esta iniciativa no solo asegurará una gestión eficiente del recurso hídrico, sino que también impulsará el desarrollo social, económico y ambiental del país, alineándose con los objetivos de sostenibilidad y resiliencia a nivel nacional e internacional.

Objetivos

Objetivo General

Desarrollar una base de datos relacional sobre agua subterránea, utilizando la metodología en cascada, con el propósito de facilitar el análisis exploratorio y descriptivo de la información hidrogeológica en el Servicio Geológico Colombiano (SGC).

Objetivos Específicos

Diseñar una base de datos relacional que incluya el modelo entidad-relación, diccionario y relaciones, aplicando las formas normales existentes, para garantizar la organización y estructura adecuada de los datos hidrogeológicos.

Integrar datos hidrogeológicos del Servicio Geológico Colombiano (SGC), incluyendo inventarios de puntos de agua subterránea y sus características fisicoquímicas e isotópicas, a la base de datos, automatizando procesos de ingreso, consulta y validación, para optimizar la gestión y disponibilidad.

Realizar un análisis exploratorio de datos espaciales sobre la información estructurada en el sistema, para identificar patrones y tendencias espaciales a través de la generación de mapas temáticos y el cálculo de medidas estadísticas.

Marco de Referencia

Datos Hidrogeológicos

La hidrogeología es la rama de la geología que estudia la distribución y movimiento del agua subterránea en el suelo y las rocas de la corteza terrestre (Custodio y Llamas, 1996).

Desempeña un papel importante en el diseño de proyectos sostenibles así como en la gestión de recursos hídricos, al facilitar la identificación de fuentes de agua subterránea, evaluar su calidad y cantidad, y prever los impactos ambientales derivados de su explotación (Agencia Nacional de Licencias Ambientales (ANLA), 2021; Ministerio de Ambiente y Desarrollo Sostenible, 2014)

Los datos hidrogeológicos son la información relacionada con las aguas subterráneas, que incluye niveles piezométricos, caudales, composición química, firmas isotópicas, permeabilidad, porosidad y ubicación geográfica de los acuíferos. Estos datos también comprenden aspectos relacionados con la interacción agua-roca, las condiciones hidrogeológicas locales y regionales, y los parámetros climáticos que afectan la recarga de los acuíferos (Custodio y Llamas, 1996; Ministerio de Ambiente y Desarrollo Sostenible, 2014).

Servicio Geológico Colombiano (SGC)

El Servicio Geológico Colombiano (SGC) es la entidad responsable de generar, gestionar y divulgar la información geocientífica en Colombia. Su misión incluye la recopilación, análisis y difusión de datos relacionados con los recursos naturales. En el área de agua subterránea, el SGC trabaja para desarrollar, promover y divulgar el conocimiento hidrogeológico del país, con el objetivo de fortalecer la gestión sostenible de este recurso (Servicio Geológico Colombiano, s. f.).

Bases de Datos Relacional

Una base de datos relacional es una colección de datos organizados según el modelo

relacional, gestionada por un Sistema de Gestión de Bases de Datos (SGBD). Este tipo de bases de datos permite almacenar, acceder y gestionar la información de manera eficiente y estructurada (Elmasri y Navathe, 2015). Se fundamenta en el modelo de datos relacionales propuesto por E. F. Codd (1970), en el cual la información se estructura lógicamente mediante relaciones, representadas físicamente como tablas. Cada relación tiene un nombre y se compone de atributos (columnas), en tanto que cada tupla (fila) contiene un valor específico para cada atributo (Connolly y Begg, 2015).

Una entidad es un objeto o concepto del mundo real que se almacena en la base de datos, y cada entidad se representa como una tabla. Los atributos son las propiedades o características de una entidad, representadas por las columnas de la tabla. Las relaciones definen las conexiones entre entidades, mostrando cómo las tablas se relacionan entre sí (Connolly y Begg, 2015; Elmasri y Navathe, 2015).

Las formas normales son reglas de diseño que organizan los datos dentro de una base de datos para reducir la redundancia y mejorar su integridad (Connolly y Begg, 2015; Elmasri y Navathe, 2015).

- Primera Forma Normal (1FN): Garantiza que cada columna contenga valores atómicos (indivisibles) y que cada entrada en la tabla sea única.
- Segunda Forma Normal (2FN): Asegura que todos los atributos no clave dependan completamente de la clave primaria.
- Tercera Forma Normal (3FN): Evita que los atributos no clave dependan transitivamente de la clave primaria.

Modelo Entidad-Relación

El modelo Entidad-Relación (E-R) es una técnica utilizada para diseñar la estructura de

una base de datos. Dicho modelo formaliza mediante Diagrama Entidad Relación (DER), los cuales proporcionan una representación gráfica de las entidades, sus atributos y las relaciones entre ellas. Los DER facilitan la visualización y definición de cómo los diferentes componentes de una base de datos interactúan entre sí (Connolly y Begg, 2015; Silberschatz et al., 2002).

Las relaciones representan cómo las entidades están conectadas entre sí, se ilustran mediante rombos que se colocan entre las entidades participantes de la relación. La cardinalidad especifica el número de instancias de una entidad que pueden estar asociadas con una instancia de otra entidad.

Las cardinalidades comunes son:

- Uno a uno (1:1): Cada instancia de una entidad está asociada con exactamente una instancia de otra entidad.
- Uno a muchos (1:M): Una instancia de una entidad está asociada con múltiples instancias de otra entidad.
- Muchos a muchos (M:N): Varias instancias de una entidad están asociadas con múltiples instancias de otra entidad.

Sistemas de Gestión de Bases de Datos (SGBD)

Los sistemas de gestión de bases de datos (SGBD), como MySQL, PostgreSQL y otros, son herramientas para la implementación y administración de bases de datos. Estos sistemas simplifican la creación, consulta, actualización y validación de datos a través de lenguajes como SQL (Elmasri y Navathe, 2015). Además, estos sistemas facilitan la implementación de lógica de negocio mediante disparadores (*triggers*) y procedimientos almacenados que aseguran la integridad, consistencia y disponibilidad de la información.

Metodología en Cascada

La metodología en cascada es un enfoque secuencial para el desarrollo de proyectos, donde cada fase debe completarse antes de que la siguiente comience. Este modelo lineal sigue un flujo descendente, similar al comportamiento de una cascada, la cual comprende etapas como la recolección y análisis de requisitos, diseño del sistema, implementación, pruebas, despliegue y mantenimiento.

En el desarrollo de bases de datos, la metodología en cascada puede aplicarse siguiendo el enfoque descrito por Nina y Gissela (2017) de la siguiente manera:

- **Recolección y Análisis de Requisitos:** se identifican y documentan los requisitos de la base de datos, incluyendo los tipos de datos a almacenar, las relaciones entre ellos y las necesidades específicas de los usuarios.
- **Diseño del Sistema:**
 - Se elabora un modelo entidad-relación (ERD) que representa gráficamente las entidades, sus atributos y las relaciones entre ellas.
 - Se diseña la estructura de la base de datos, definiendo tablas, índices y restricciones.
 - Se establecen las formas normales necesarias para garantizar la integridad y la eficiencia de los datos.
- **Implementación:** se desarrolla la base de datos en un sistema de gestión de bases de datos (SGBD) utilizando SQL u otro lenguaje de definición de datos. También se insertan datos iniciales necesarios para la operación.
- **Pruebas:** se ejecutan pruebas funcionales para verificar que la base de datos cumple con los requisitos establecidos. Esto incluye la validación de consultas, operaciones de

manipulación de datos y pruebas de casos extremos, asegurando que las restricciones y reglas de integridad se respeten.

- **Despliegue:** la base de datos se implementa en el entorno de producción y se pone a disposición de los usuarios finales.
- **Mantenimiento:** se monitorea el rendimiento de la base de datos y se realizan ajustes u optimizaciones según sea necesario. Asimismo, se implementan actualizaciones o cambios para corregir problemas y mejorar la funcionalidad.

Análisis Exploratorio de Datos Espaciales

El Análisis Exploratorio de Datos Espaciales (AEDE) es un conjunto de técnicas estadísticas y visuales utilizadas para examinar y describir patrones espaciales en datos geográficos.

Este tipo de análisis parte del principio de que la ubicación geográfica influye en los atributos de los objetos o fenómenos observados. En el caso de estudios hidrogeológicos, esto implica que variables hidrogeológicas como el nivel del agua, la profundidad de captación o la conductividad eléctrica de los acuíferos pueden presentar autocorrelación espacial, es decir, que los valores de estas variables tienden a ser similares entre puntos cercanos. Este concepto es desarrollado por Anselin (1995), quien introduce herramientas estadísticas como el índice de Moran, el estadístico de Geary y los mapas de clúster, útiles para identificar agrupamientos espaciales significativos.

En el contexto específico del análisis de datos hidrogeológicos, aplicar AEDE permite explorar y visualizar la distribución espacial de variables fisicoquímicas, facilitando la comprensión de sus patrones y características hidrogeológicas.

Antecedentes en la Estandarización de Información Hidrogeológica

La gestión de los recursos hídricos subterráneos requiere de herramientas avanzadas que centralicen información clave sobre acuíferos, pozos, niveles freáticos y calidad del agua.

Diversos países han desarrollado bases de datos hidrogeológicas que sirven como referencia en la organización, análisis y difusión de información para la toma de decisiones y la investigación científica. En la tabla 1 se describen algunas de las bases de datos más destacadas a nivel global.

Tabla 1

Principales Bases de Datos Hidrogeológicas en el Mundo

| Nombre | Características e Información |
|---|--|
| <p data-bbox="284 1255 714 1291"><i>Sistema de Informações de Águas</i></p> <p data-bbox="341 1329 657 1365"><i>Subterrâneas (SIAGAS)</i></p> | <p data-bbox="824 856 1404 1764">El SIAGAS, desarrollado por el Servicio Geológico de Brasil (CPRM), es una de las bases de datos hidrogeológica más completas de América Latina. Este sistema centraliza información detallada sobre aguas subterráneas en Brasil, facilitando su acceso y uso tanto para entidades gubernamentales como para investigadores y empresas privadas. Entre sus principales característica destaca los más de 25.000 registros almacenados, georreferenciación, características y diseño de pozos profundos, entre otros tipos de puntos.</p> |

| | |
|--|---|
| <i>National Groundwater Information System</i> (NGIS) | El NGIS, desarrollado por <i>Geoscience</i> Australia, es una base de datos centralizada que reúne información sobre pozos de agua subterránea en todo el país. Cuenta con información de más de 870.000 puntos de agua, incluye información de contexto geológico, caracterización de puntos y variación de niveles. |
| <i>United States Geological Survey (USGS) –</i> Base de Datos Nacional de Aguas Subterráneas | El USGS gestiona una de las bases de datos más extensas en los Estados Unidos. Agrupa información sobre características del agua. variación de niveles, disponibilidad de agua subterránea para explotación. Destaca su sinergia con redes de monitoreo telemétrico, permitiendo así inspeccionar de forma detallada variaciones en el nivel estático de unidades acuíferas estratégicas. |
| <i>British Geological Survey (BGS) – Reino</i> Unido | El BGS administra bases de datos hidrogeológicas que proporcionan información detallada sobre acuíferos en el Reino Unido. Incluye información sobre calidad de agua, y variaciones en |

almacenamiento y recarga de agua
subterránea.

En Colombia, la centralización de la información relacionada con la disposición, uso, almacenamiento y aprovechamiento del agua recae directamente sobre el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), entidad que tiene entre sus objetivos principales la consolidación y disponibilidad de esta información.

En particular, el monitoreo del agua subterránea se realiza en tres escalas diferentes:

- Nacional: a través de la Red Básica Nacional de Aguas Subterráneas
- Regional: mediante el Programa Institucional Regional de Monitoreo del Agua
- Local o específica: por medio de redes de monitoreo empresariales o asociadas a proyectos particulares.

Diversas entidades contribuyen con información recolectada a nivel local o regional, producto de concesiones o actividades de exploración en el marco de proyectos específicos. Este es el caso de las Corporaciones Autónomas Regionales (CAR) o de institutos científicos cuyos objetivos requieren el conocimiento y caracterización de fuentes de agua subterránea.

Si bien existen programas y protocolos para la recopilación y disposición de esta información, aún persisten brechas en la interoperabilidad de los datos y, fundamentalmente, en su accesibilidad pública mediante plataformas digitales que garanticen una disponibilidad integral y estandarizada. Esta situación representa un reto para el conocimiento detallado de las fuentes de agua subterránea, su calidad y dinámica temporal. De hecho, según el IDEAM (2023), en Colombia se encuentran inventariadas 68.397 captaciones de agua subterránea, entre pozos, piezómetros, aljibes, y manantiales.

Desarrollo Metodológico

El desarrollo de la base de datos sobre agua subterránea se realizará bajo los lineamientos de la metodología en cascada, los cuales incluyen el diseño conceptual y lógico, consolidación de datos y la implementación y validación de la base de datos.

Diseño de la Base de Datos

En esta etapa se diseñará la base de datos mediante la elaboración de un modelo Entidad-Relación (E-R) con el fin de representar las entidades, sus atributos y las relaciones que conforman los datos hidrogeológicos, como el inventario de puntos de agua, las características fisicoquímicas y los parámetros isotópicos. Este modelo facilitará la estructuración de la información antes de su implementación en un sistema de base de datos (Elmasri y Navathe, 2015). Durante la fase de diseño lógico, se aplicará el proceso de normalización, un conjunto de reglas definidas por Edgar F. Codd en 1970, con el fin de estructurar las tablas y sus relaciones, para mitigar la redundancia y evitar inconsistencias. Estas reglas contribuyen a obtener un diseño eficiente y coherente de las bases de datos relacionales (Mendoza y López, 2018).

Además, se desarrollará un diccionario de datos que describirá la estructura, el significado y las reglas asociadas a los datos presentes en la base de datos, con el fin de estandarizar las definiciones y facilitar el acceso a la información para los usuarios finales (Date, 2004; Elmasri y Navathe, 2015; Silberschatz et al., 2002).

Consolidación de Datos

En esta fase se integrarán los datos hidrogeológicos provenientes del Servicio Geológico Colombiano (SGC), incluyendo información relacionada con el inventario de puntos de agua subterránea, sus características fisicoquímicas e isotópicas. Este proceso implicará una consolidación de los datos, durante la cual se identificarán y corregirán posibles errores, valores

inconsistentes o duplicados. Adicionalmente, se estandarizarán las unidades de medida para garantizar la coherencia en la información.

Este proceso es esencial para la integración de datos de diferentes fuentes y asegurar su compatibilidad, siguiendo las buenas prácticas recomendadas en el diseño de bases de datos y manejo de bases de datos (Elmasri y Navathe, 2015; Silberschatz et al., 2002; Velasco et al., 2013).

Implementación de la Base de Datos

Se seleccionará un Sistema de Gestión de Bases de Datos (SGBD) para gestionar la base de datos, y se implementará el diseño previamente definido. Dicha etapa comprenderá el uso del lenguaje SQL para la creación tablas, relaciones y restricciones (Elmasri y Navathe, 2015), y el despliegue en motores de bases de datos relacionales como PostgreSQL.

Adicionalmente, se desarrollarán scripts y procedimientos para la carga, consulta y validación de la información, con el objetivo de optimizar la eficiencia en la gestión y disponibilidad de la información, así como facilitar su actualización continua (Date, 2004). Finalmente, se llevarán a cabo pruebas para verificar la integridad, coherencia y exactitud de la base de datos.

Análisis Exploratorio de Datos Espaciales

Se llevará a cabo un Análisis Exploratorio de Datos Espaciales (AEDE), el cual permite identificar patrones espaciales, tendencias y relaciones en la distribución geográfica de las variables registradas en la base de datos construida. Este enfoque metodológico se apoya en los procesos propuestos por Buzai y Baxendale (2006), así como en los métodos gráficos descritos por De Corso Sicilia y Pinilla Rivera (2017).

Inicialmente, se realizará la consulta de los registros almacenados en la base de datos, con

el fin de seleccionar las variables de interés para el análisis, tales como conductividad eléctrica, el pH y temperatura. Luego, se calcularán medidas de tendencia central (media, mediana) y dispersión (rango, desviación estándar), para describir el comportamiento estadístico de las variables en el área.

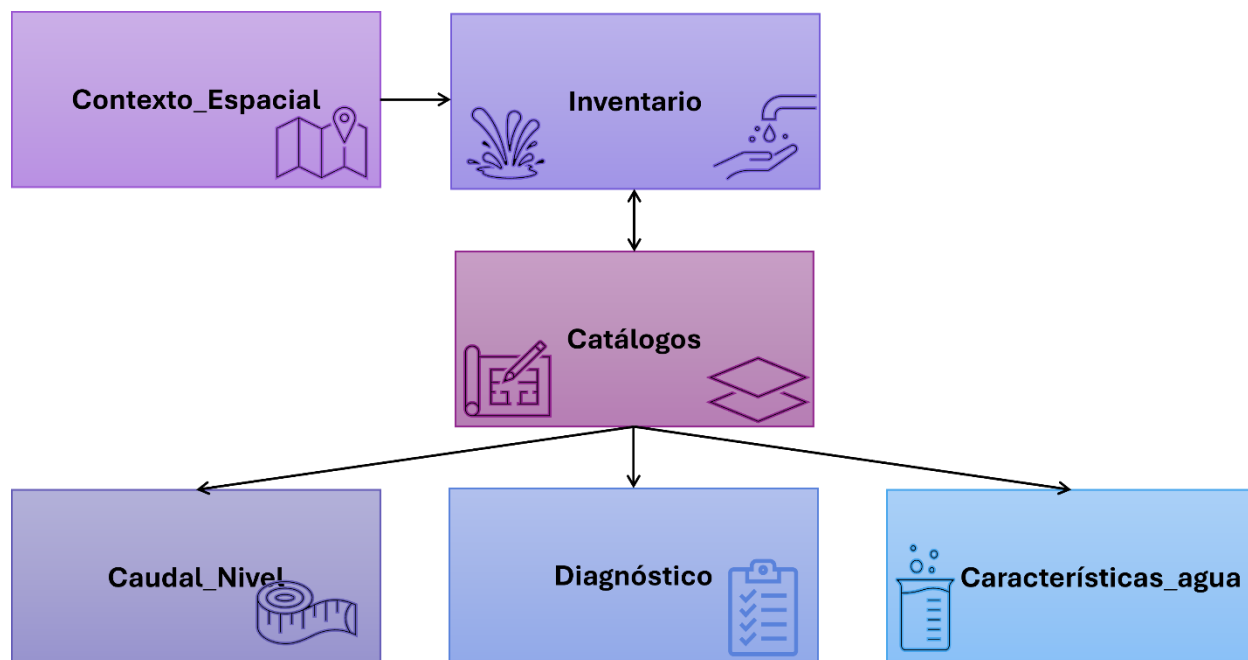
Asimismo, se generará cartografía temática, tales como mapas de coropletas, mapas de símbolos proporcionales y mapas de calor, que permitirán visualizar la distribución espacial de dichas variables. Estas representaciones gráficas facilitarán la identificación de concentraciones, gradientes o zonas atípicas, orientada a una comprensión preliminar de la distribución espacial de los datos.

Resultados y Discusión

Diseño de la Base de Datos

El diseño de la base de datos se desarrolló con el objetivo de consolidar, organizar y asegurar la calidad de la información hidrogeológica producida por el Servicio Geológico Colombiano (SGC). Para ello, se construyó un modelo entidad–relación (ER) que integra 22 tablas distribuidas en seis esquemas temáticos. Estos esquemas agrupan la información según su naturaleza (espacial, administrativa, descriptiva) y su función dentro del sistema.

La estructura propuesta contribuye a la consistencia de los datos mediante la implementación de claves primarias y foráneas, reglas de unicidad y políticas de actualización en cascada. Estas características aseguran que las modificaciones realizadas en cualquier tabla se reflejen de manera coherente en el resto del modelo, independientemente de su distribución en diferentes esquemas. La figura 1 presenta una representación gráfica de los esquemas que componen el modelo entidad–relación obtenido en esta fase, el cual se presenta de manera completa en el apéndice a.

Figura 1*Modelo de Entidad Relación (ER)*

Nota. La figura corresponde a una representación esquemática de los componentes de la base de datos, dado que el modelo entidad–relación completo no es legible en el cuerpo del documento. El modelo detallado se presenta en el apéndice a.

Resultados Generales del Diseño

El modelo entidad–relación agrupa las tablas según seis bloques temáticos, resumidos en la tabla 2.

Tabla 2*Esquemas y Número de Tablas por Esquema*

| Esquema | Función | Tablas Principales |
|----------------------|--|---|
| contexto_espacial | Marco territorial y geoespacial | departamento, municipio, zonificación rural/urbana, zonificación hidrográfica |
| Catalogos | Información de los proyectos, áreas de estudio, parámetros, y visitas | visita, parámetros, proyectos, área de estudio |
| Inventario | Objetos principales (puntos de agua subterráneas) y sus características. | punto_agua, aljibe, pozo, manantial, condiciones_dinamicas |
| Caudal_nivel | Mediciones hidrodinámicas. | caudal, nivel |
| Caracteristicas_agua | Datos fisicoquímicos, de campo y laboratorio. | parametro_campo, propiedad_organoleptica, resultados_laboratorio |
| Diagnóstico | Información sobre diagnóstico sanitario y posibles fuentes de contaminación. | diagnostico_sanitario |

Los esquemas contexto_espacial y catalogos constituyen el marco territorial y administrativo sobre el cual se organiza la información. El esquema inventario define la entidad central del sistema: el punto de agua subterránea, representado mediante un tipo de datos geométrico puntual bajo el identificador de referencia espacial 4686 (MAGNA-SIRGAS) y

controlado mediante restricciones que aseguran su unicidad espacial. Los esquemas caudal_nivel, características_agua y diagnóstico_sanitario complementan la descripción del recurso mediante la integración de datos hidrodinámicos, fisicoquímicos, isotópicos y sanitarios.

Cada tabla y cada atributo fueron documentados mediante comentarios descriptivos dentro del archivo maestro SQL. Este archivo sirvió de base para la elaboración del diccionario de datos, herramienta que facilita la interpretación de las estructuras, atributos y relaciones del sistema. El diccionario se presenta en el apéndice b.

Restricciones y Aseguramiento de la Calidad de la Información. El diseño incorpora un conjunto de restricciones orientadas a asegurar la validez, confiabilidad y coherencia interna de la información. Estas restricciones actúan como un filtro inicial que previene inconsistencias y disminuye la necesidad de depuración manuales en etapas analíticas subsecuentes.

Restricciones de Integridad Estructural. Se implementaron claves primarias, claves foráneas, reglas de unicidad y restricciones de obligatoriedad (NOT NULL) en atributos críticos, entre los controles más relevantes se destacan:

- Integridad referencial con actualización en cascada: permite que los cambios realizados en los catálogos o en el inventario se actualicen automáticamente en las tablas dependientes.
- Reglas de unicidad: aplicadas a códigos de puntos, nombres de proyectos, identificadores de visitas y geometrías espaciales.
- Unicidad espacial del punto de agua: evita duplicados en la ubicación exacta del inventario.
- Coherencia territorial: el municipio debe corresponder al departamento según el código DANE.

- Evitar visitas duplicadas.
- Cohesión entre tablas especializadas: permite que solo existan registros en tablas

de tipo de captación si el punto está clasificado como tal.

Lógicas de Negocio Orientadas al Contexto Hidrogeológico. Estas reglas incorporan conocimiento sobre el comportamiento del recurso hídrico subterráneo y optimizan la analítica de datos al minimizar la incidencia de valores atípicos o errores de captura.

- Cohesión entre tablas especializadas: permite que solo existan registros en tablas de tipo de captación si el punto está clasificado como tal.
- Rango de valores físicamente razonables: incluye controles como el rango de pH (0 a 14).
- Coherencia espacial y altitudinal: evita el registro de puntos fuera del territorio colombiano o con altitudes imposibles.
- Coherencia hidrodinámica: restricción para niveles freáticos en caso de nivel saltante el valor registrado debe ser negativo.
- Lógica de succión en pozos: la profundidad de succión debe ser menor a la profundidad del pozo.

Validaciones para Registros Completos y Consistentes. Se implementaron reglas que evitan registros incompletos o inconsistentes, por ejemplo:

- Propiedades organolépticas: al menos uno de los tres campos de la tabla (olor, color y turbiedad) no debe ser nulo.
- Diagnóstico sanitario: evita indicar contaminación sin describirla.

Uso de Dominios para Estandarizar Categorías. Se definieron listas controladas para categorías como: morfología, color, olor y turbiedad, tipo de uso del agua, unidades de medida, estados de proyecto, métodos de medición, tipos de zona.

El uso de dominios favorece la estandarización, interoperabilidad y calidad en las consultas, al reducir la ambigüedad y los errores de digitación.

Diseño por Esquemas: Organización Temática

La organización en seis esquemas permite administrar de manera diferenciada los componentes espaciales, administrativos y descriptivos, así facilitar la trazabilidad de la información desde la ubicación del punto hasta sus características fisicoquímicas, isotópicas e hidrodinámicas. El modelo conceptual se estructura en tres niveles: contexto, objeto y temática.

Contexto. Incluye los esquemas contexto_espacial y catalogo_proyectos, que conforman el marco estadístico geoespacial y administrativo. Contiene información de departamentos y municipios (DANE), zonas hidrográficas (IDEAM), zonificación rural/urbana, proyectos, áreas de estudio y metadatos del SGC.

Objeto. Corresponde al eje central del diseño: el punto de agua subterránea. Aquí se incluye la identificación, coordenadas y geometrías validadas, morfología y tipo de captación, fecha de inventario. Las tablas aljibe, pozo y manantial complementan las características específicas según el tipo de captación y la tabla visita introduce la dimensión temporal, permitiendo registrar visitas y condiciones a lo largo del tiempo.

Temática. Agrupa la información observada, medida o evaluada: parámetros fisicoquímicos, propiedades organolépticas, datos isotópicos, niveles freáticos y caudales, diagnóstico sanitario.

La integración entre estos esquemas permite recorrer el conjunto de datos desde su localización espacial hasta los atributos que definen su comportamiento hidrogeológico, manteniendo la coherencia entre tablas.

Consolidación de los Datos Hidrogeológicos

La consolidación de los datos hidrogeológicos correspondió a la fase en la cual se integró la información proveniente de múltiples tablas y formularios históricos del Servicio Geológico Colombiano (SGC), los cuales presentaban diferencias estructurales. Esta etapa tuvo como propósito obtener un conjunto de datos homogéneo consistente, trazable y compatible con la estructura de la Base de Datos Hidrogeológica (BDH) diseñada.

El proceso permitió identificar inconsistencias, errores de digitación y vacíos de información, los cuales fueron corregidos o documentados mediante reglas definidas durante el diseño de la base de datos.

Fuentes y Características Iniciales de los Datos

Los datos consolidados provienen de inventarios de puntos de agua subterránea y campañas de caracterización hidrogeológica realizadas por el grupo de Aguas Subterráneas del Servicio Geológico Colombiano y consultada a través del Motor de Integración de Información Geocientífica (MIIG). La información se encontraba distribuida en distintos formatos, principalmente hojas de cálculo y base de datos en Access, con diferencias en la estructura, la nomenclatura de los campos, las unidades de medida y el nivel de completitud. Esta heterogeneidad evidenció la necesidad de aplicar un proceso de estandarización y validación previo a la carga en la base de datos.

Se integraron registros correspondientes a condiciones dinámicas del punto, mediciones de caudal, niveles piezométricos, análisis organolépticos y parámetros fisicoquímicos de campo.

Cada conjunto de datos se vinculó mediante el código único del punto (`cod_sgc`) y la fecha de inventario o medición, lo que permitió establecer relaciones temporales entre las observaciones realizadas en un mismo punto.

Proceso de Extracción y Transformación

La integración de los datos se desarrolló mediante un proceso de extracción y transformación, alineado con el diseño lógico de la base de datos.

Extracción. Los datos fueron extraídos de los archivos fuente, extrayendo únicamente los campos relevantes para cada componente hidrogeológico y organizados en estructuras temporales que permitieron su revisión, identificación de campos equivalentes y detección de registros incompletos.

Transformación. Durante esta etapa se aplicaron procedimientos de estandarización y control de calidad que incluyeron la normalización de nombres de campos, la conversión de unidades de medida, la homogenización de categorías mediante dominios, la corrección de errores tipográficos, la validación de rangos físicos y químicos, y la revisión de la coherencia espacial y temporal de los registros.

Las restricciones definidas en el modelo de datos actuaron como un mecanismo de control automático, al impedir el ingreso de registros inconsistentes y facilitar la identificación de errores recurrentes durante el proceso de carga y consolidación.

En la etapa de integración se identificaron diversos tipos de inconsistencias, entre las que se destacan la presencia de registros duplicados para un mismo punto y fecha, diferencias en el código del punto (`cod_sgc`) asociadas a coordenadas coincidentes, valores faltantes en variables clave, fechas improbables o inconsistentes, y mediciones sin asociación a una visita válida o a un punto activo dentro de un proyecto.

De manera complementaria, se aplicaron procesos adicionales de estandarización que incluyeron la homologación de estructuras y nombres de campos, la conversión de formatos de fecha a un estándar común, la corrección de fechas inconsistentes, en particular aquellas interpretadas erróneamente como posteriores a la fecha actual, el reemplazo de códigos internos por categorías descriptivas a partir de tablas de dominio, y la normalización de unidades junto con la reclasificación de categorías, como los métodos de medición y los tipos de uso del agua.

Identificación y Tratamiento de Errores e Inconsistencias

El proceso de consolidación permitió identificar distintos tipos de inconsistencias, entre las que se destacan:

- Registros duplicados de puntos de agua
- Visitas repetidas en la misma fecha
- Valores fuera de rangos físicos plausibles
- Inconsistencias entre parámetros relacionados
- Errores en la asignación territorial (municipio–departamento)
- Uso de categorías no estandarizadas

Estos casos fueron tratados mediante corrección directa, estandarización o exclusión del registro, según el tipo de error y su impacto en la calidad de la información. La aplicación de restricciones y dominios redujo significativamente la posibilidad de que estos errores persistieran en la base consolidada.

Una de las principales inconsistencias detectadas correspondió a variables categóricas almacenadas originalmente como códigos numéricos, cuya interpretación dependía de tablas intermedias en la base de datos de origen. Entre estas variables se incluyen la morfología, la

permanencia de los manantiales, el tipo de surgencia, el tipo de punto de agua y el origen de las coordenadas.

En el sistema original, estas categorías se representaban mediante valores enteros (por ejemplo, 1 = aljibe, 2 = manantial, 3 = pozo), lo que implicaba la existencia de múltiples tablas auxiliares para su interpretación. Durante la consolidación se realizó la traducción explícita de estos códigos a categorías textuales, utilizando las tablas de referencia disponibles, con el fin de hacer la información legible, interoperable y coherente con el uso de dominios definidos en la base de datos relacional.

Como resultado de este proceso, se transformaron 14350 registros asociados a códigos categóricos, eliminando ambigüedades y reduciendo la dependencia de tablas intermedias. Esta decisión permitió simplificar el modelo lógico, disminuir la complejidad estructural y garantizar la estandarización de los valores mediante dominios controlados.

Adicionalmente, los valores clasificados como “desconocido” o equivalentes fueron recodificados como “sin información”, evitando categorías ambiguas que no aportan información analítica.

Registros Incompletos y Valores Nulos. El análisis inicial evidenció la presencia de campos con altos porcentajes de valores nulos, particularmente en atributos que no aplican a todos los tipos de puntos de agua. Para abordar esta situación, se implementó un procedimiento de normalización estructural que consistió en:

- Unificar la estructura de las tablas de aljibes, pozos y manantiales
- Incorporar columnas faltantes con valores nulos cuando el atributo no era

aplicable

- Reordenar los campos de acuerdo con la estructura definida para el esquema inventario

Duplicidad Administrativa y Espacial. Durante la consolidación se detectaron duplicados asociados tanto al código del punto de agua como a su ubicación espacial. En una primera etapa se eliminaron los registros duplicados asociados al mismo código SGC. Posteriormente, se identificaron 268 registros con coordenadas planas coincidentes (Este–Norte) y fecha de inventario, correspondientes a posibles duplicados espaciales.

Dado que estos registros presentan códigos distintos y pueden contener información complementaria asociada a diferentes campañas o momentos de levantamiento, se optó por no eliminarlos de manera inmediata. En su lugar, se incorporó un marcador de control que permite identificar estos casos como registros a revisar por duplicidad espacial, con el fin de postergar la decisión de consolidación definitiva hasta contar con criterios objetivos basados en la cantidad y calidad de la información disponible por punto.

Esta estrategia permitió conservar la totalidad de los datos originales, mantener la trazabilidad de la información y evitar decisiones arbitrarias que pudieran afectar la representatividad del inventario.

Inconsistencias Geométricas y Espaciales. Se identificaron registros sin información de coordenadas o con geometrías inválidas, los cuales fueron excluidos del conjunto consolidado debido a la imposibilidad de asociarlos a un contexto espacial confiable. Posteriormente, se realizó un proceso de reproyección y estandarización de coordenadas, garantizando la coherencia entre coordenadas geográficas y planas dentro del sistema de referencia MAGNA–SIRGAS.

Adicionalmente, se completaron automáticamente los campos de departamento, municipio, zonificación rural–urbana y zonificación hidrográfica mediante cruces espaciales con

capas oficiales, lo que permitió corregir inconsistencias territoriales y asegurar la correcta localización administrativa de los puntos de agua.

La asignación de los atributos de departamento, municipio y zonificación rural–urbana se realizó a partir de la División Político-Administrativa oficial de Colombia (DIVIPOLA), administrada por el Departamento Administrativo Nacional de Estadística (DANE) y soportada en la cartografía base oficial del Instituto Geográfico Agustín Codazzi (IGAC).

La zonificación hidrográfica se determinó mediante cruces espaciales con las unidades hidrográficas oficiales definidas por el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), las cuales constituyen el marco de referencia nacional para la gestión del recurso hídrico.

Resultado del Proceso de Depuración

En conjunto, el proceso de identificación y tratamiento de errores permitió transformar un conjunto de datos heterogéneo y con inconsistencias en una base consolidada, estructurada y validada. Las decisiones adoptadas priorizaron la conservación de la información original, la trazabilidad de los registros y la coherencia con el diseño de la base de datos relacional, dejando los datos en condiciones adecuadas para su almacenamiento y posterior uso.

Como resultado del proceso de consolidación, se obtuvo un conjunto de tablas depuradas, normalizadas y estructuradas conforme al modelo de la BDH, las cuales conservan la trazabilidad hacia los registros originales e incorporan controles de calidad básicos.

Tabla 3

Resumen de Errores e Inconsistencias Detectadas y Acciones Aplicadas Durante la Consolidación de los Datos

| Tipo de Inconsistencia | Descripción del Problema Identificado | Registros Afectados (n) | Acción Aplicada | Resultado |
|--|--|-------------------------|--|---|
| Codificación categórica no estandarizada | Variables categóricas (morfología, permanencia, surgencia, tipo de punto, origen de coordenadas) | 11125 | Traducción de códigos a categorías textuales y estandarización mediante dominios | Eliminación de ambigüedad y reducción de dependencias estructurales |
| Categorías ambiguas o no informativas | Presencia de valores como “desconocido” o códigos nulos en variables categóricas | 522 | Recodificación a la categoría “otro” | Mejora de la interpretabilidad y consistencia semántica |
| Registros sin información espacial | Ausencia de coordenadas o geometrías inválidas | 1 | Exclusión del conjunto consolidado | Garantía de coherencia espacial |

| Tipo de Inconsistencia | Descripción del Problema Identificado | Registros Afectados (n) | Acción Aplicada | Resultado |
|-------------------------------|--|-------------------------|---|-------------------------------------|
| Duplicidad administrativa | Repetición del mismo código SGC | 3 | Eliminación de registros duplicados | Unicidad administrativa garantizada |
| Duplicidad espacial | Registros con coordenadas planas coincidentes (Este–Norte) | 265 | Eliminación de registros duplicados | Puntos únicos |
| Inconsistencias territoriales | Puntos sin correspondencia válida con departamento o municipio | 11125 | Asignación automática mediante cruce espacial con capas oficiales | Integridad territorial garantizada |

Implementación de la Base de Datos

Para materializar el diseño conceptual en un sistema funcional, capaz de garantizar la integridad, consistencia y trazabilidad de la información hidrogeológica, así como de soportar consultas se realizó la implementación técnica de la Base de Datos Hidrogeológica (BDH) a partir del modelo lógico y relacional definido en las etapas previas.

La implementación se realizó en un entorno local, permitiendo validar de forma controlada la estructura de la base de datos, las reglas de integridad y los mecanismos de control definidos durante el diseño.

Elección del SGBD y Configuración del Entorno

La base de datos fue implementada utilizando el Sistema de Gestión de Bases de Datos PostgreSQL, versión 15.6, complementado con la extensión espacial PostGIS versión 3.5.2, la selección de dicho entorno se fundamentó en criterios técnicos y funcionales para la gestión de información hidrogeológica.

PostgreSQL es un SGBD de código abierto ampliamente utilizado en contextos científicos e institucionales, que ofrece soporte robusto para integridad referencial, dominios, restricciones avanzadas y escalabilidad. Por su parte, PostGIS extiende estas capacidades al incorporar tipos de datos espaciales y funciones geográficas, adecuada para el manejo de información hidrogeológica georreferenciada.

El entorno de implementación correspondió a un equipo local con sistema operativo Windows, lo que permitió reproducir condiciones similares a las de análisis y desarrollo utilizadas en entornos técnicos del Servicio Geológico Colombiano (SGC). Se trabajó con un único usuario de base de datos, configurado con permisos completos para la creación de esquemas, tablas, dominios y restricciones, dado el carácter académico y controlado del ejercicio.

La base de datos se organizó en seis esquemas temáticos, definidos con el fin de mejorar la organización lógica, la mantenibilidad y la claridad del modelo:

- contexto_espacial, para información territorial y geométrica;
- catalogos, para tablas de referencia;
- inventario, para el registro principal de puntos de agua subterránea;
- diagnóstico, para información asociada al diagnósticos sanitaria que se realiza

durante las visitas de los puntos de agua subterránea;

- caudal_nivel, para almacenar mediciones dinámicas;
- características_agua, para parámetros fisicoquímicos y organolépticos.

Creación de Tablas, Dominios y Restricciones

La implementación del modelo relacional se realizó mediante la creación explícita de esquemas, tablas, dominios y restricciones, utilizando sentencias SQL estructuradas y ejecutadas directamente desde la consola psql. En total, se ejecutaron 4 archivos SQL independientes, cada uno correspondiente a la creación de una tabla o conjunto de objetos relacionados, lo que favoreció la trazabilidad del proceso y la reutilización del código.

Se crearon 25 dominios personalizados, utilizados para controlar los valores permitidos en variables categóricas. Estos dominios incluyen, entre otros, tipos de punto de agua, permanencia, uso del suelo, parámetros medidos y unidades de medida. Aunque los dominios no se agruparon en un único esquema denominado explícitamente “dominios”, sí se organizaron de manera coherente dentro de los esquemas temáticos correspondientes, garantizando su reutilización y consistencia.

Las tablas fueron definidas incorporando de forma sistemática restricciones de integridad, incluyendo:

- claves primarias (PRIMARY KEY), para garantizar la unicidad de los registros;
- claves foráneas (FOREIGN KEY), para asegurar la coherencia entre tablas relacionadas;
- restricciones UNIQUE, para evitar duplicidades;
- restricciones CHECK, para validar rangos físicos, valores permitidos y coherencia básica;
- restricciones NOT NULL, aplicadas únicamente a atributos obligatorios.

La tabla inventario.punto_agua fue implementada como la tabla padre del modelo, almacenando la información básica e identificadora de cada punto de agua subterránea. A partir de esta tabla se estableció la relación con la tabla de visitas hidrogeológicas, la cual registra las diferentes observaciones realizadas a un mismo punto en fechas distintas.

Las tablas temáticas de caudal, nivel, parámetros fisicoquímicos y demás mediciones dependen relacionamente de la tabla de visitas, lo que permite estructurar la información de manera temporal y evitar la duplicación de atributos. Esta jerarquía fue reforzada mediante restricciones de integridad referencial con reglas explícitas de ON UPDATE CASCADE y ON DELETE RESTRICT, garantizando que las actualizaciones de identificadores se propaguen correctamente y que no se eliminen registros padre con información dependiente.

Desde el punto de vista espacial, se incorporó un campo geométrico definido como GEOMETRY(MULTIPOLYGON, 4686), asegurando la compatibilidad con el sistema de referencia MAGNA-SIRGAS y habilitando el uso de funciones espaciales para análisis geográficos posteriores.

Pruebas de Integridad y Consistencia

Una vez implementada la estructura de la base de datos, se realizaron pruebas orientadas a evaluar el comportamiento del sistema frente a la inserción de datos inconsistentes, así como a validar la efectividad de las restricciones definidas.

Las pruebas incluyeron intentos deliberados de inserción de registros con valores fuera de los rangos permitidos, uso de categorías no contempladas en los dominios definidos y referencias a claves foráneas inexistentes. En todos estos casos, el SGBD respondió de manera adecuada, rechazando los registros y emitiendo mensajes de error asociados a violaciones de restricciones CHECK, dominios inválidos o integridad referencial. Ejemplos de pruebas realizadas:

Inserción Correcta. En la figura 2, se presenta un ejemplo de la sintaxis para la inserción correcta de un registro en la tabla punto_agua.

Figura 2

Ejemplo Inserción Correcta

```
INSERT INTO inventario.punto_agua VALUES ('110-I-A-1', 'Angosturas 1', -72.8889325644827,
7.41030559507461, 3549, NULL, '110-I-A', 1, 1131223, 1311378, '68', '68780', '687800203',
'2319', 'planicie', 'ninguno', 'pozo', '2022-02-04',
'01010000204E120000BB9BA73AE43852C00A698D4127A41D40');
```

| Data Output | Messages | Notifications |
|--|----------|---------------|
| INSERT 0 1 | | |
| Query returned successfully in 188 msec. | | |

Prueba de Dominio Inválido. En la figura 3, se presenta un ejemplo de una inserción errónea al intentar ingresar un valor que no pertenece a un dominio válido; en este caso, se viola el dominio definido para el tipo de punto, 'PIEZOMETRO' no es un valor valido.

Figura 3

Prueba de Dominio

```
INSERT INTO inventario.punto_agua VALUES ('110-I-A-1', 'Angosturas 1', -72.8889325644827,
7.41030559507461, 3549, NULL, '110-I-A', 1, 1131223, 1311378, '68', '68780', '687800203',
'2319', 'planicie', 'ninguno', 'PIEZOMETRO', '2022-02-04',
'01010000204E120000BB9BA73AE43852C00A698D4127A41D40');
```

| Data Output | Messages | Notifications |
|---|----------|---------------|
| ERROR: el valor para el dominio inventario.dominio_punto_agua viola la restricción «check» «dominio_punto_agua_check» | | |
| SQL state: 23514 | | |

Prueba de Violación de CHECK (Rango Geográfico). En la figura 4, se muestra un ejemplo de la violación de la restricción asociada al rango geográfico, dado que existe un *check* que limita las coordenadas al territorio de Colombia y las coordenadas indicadas no son válidas.

Figura 4

Prueba Restricción Rango Geográfico

```
INSERT INTO inventario.punto_agua VALUES ('110-I-A-1', 'Angosturas 1', -120, 40, 3549, NULL,
'110-I-A', 1, 1131223, 1311378, '68', '68780', '687800203', '2319', 'planicie', 'ninguno',
'PIEZOMETRO', '2022-02-04', '01010000204E120000BB9BA73AE43852C00A698D4127A41D40');
```

Data Output [Messages](#) [Notifications](#)

ERROR: La fila que falla contiene (110-I-A-1, Angosturas 1, -120, 40, 3549, null, 110-I-A, 1, 1131223, 1311378, 68, 68780, 687800203, 2319, planicie, ninguno, pozo, 2022-02-04, 01010000204E120000BB9BA73AE43852C00A698D4127A41D40).el nuevo registro para la relación «punto_agua» viola la restricción «check» «check_coordenadas_colombia»

ERROR: el nuevo registro para la relación «punto_agua» viola la restricción «check» «check_coordenadas_colombia»
SQL state: 23514

Detail: La fila que falla contiene (110-I-A-1, Angosturas 1, -120, 40, 3549, null, 110-I-A, 1, 1131223, 1311378, 68, 68780, 687800203, 2319, planicie, ninguno, pozo, 2022-02-04, 01010000204E120000BB9BA73AE43852C00A698D4127A41D40).

Prueba de Violación de Llave Foránea (FK). En la figura 5, se presenta un ejemplo en el que se intenta registrar un código de municipio que no corresponde al departamento indicado, lo que viola la restricción municipio-departamento. Adicionalmente, en la figura 6, se presenta un segundo caso en el que se vulnera una restricción de llave foránea al intentar registrar un manantial asociado a un punto que no existe en la tabla punto agua.

Figura 5

Prueba Restricción Llave Foránea y Relación Municipio-Departamento

```
INSERT INTO inventario.punto_agua VALUES ('110-I-A-1', 'Angosturas 1', -72.8889325644827,
7.41030559507461, 3549, NULL, '110-I-A', 1, 1131223, 1311378, '68', '99999', '687800203',
'2319', 'planicie', 'ninguno', 'pozo', '2022-02-04',
'01010000204E120000BB9BA73AE43852C00A698D4127A41D40');
```

Data Output [Messages](#) Notifications

```
ERROR: La fila que falla contiene (110-I-A-1, Angosturas 1, -72.8889325644827, 7.41030559507461, 3549, null, 110-I-A, 1,
1131223, 1311378, 68, 99999, 687800203, 2319, planicie, ninguno, pozo, 2022-02-04,
01010000204E120000BB9BA73AE43852C00A698D4127A41D40).el nuevo registro para la relación «punto_agua» viola la restricción
«check» «check_municipio_departamento»
```

Figura 6

Prueba Llave Foránea cod_sgc

```
INSERT INTO inventario.manantial VALUES ('110-I-C-8', 'manantial', 'perenne', 'suelo',
'corriente');
```

Data Output [Messages](#) Notifications

```
ERROR: La llave (cod_sgc)=(110-I-C-8) no está presente en la tabla «punto_agua».inserción o actualización en la tabla
«manantial» viola la llave foránea «manantial_cod_sgc_fkey»
```

```
ERROR: inserción o actualización en la tabla «manantial» viola la llave foránea «manantial_cod_sgc_fkey»
SQL state: 23503
Detail: La llave (cod_sgc)=(110-I-C-8) no está presente en la tabla «punto_agua».
```

Prueba de Unicidad Espacial. En la figura 7, se presenta el restado de intentar registrar un punto con una geometría idéntica a la de un punto previamente almacenado, por lo que la base retorna un error indicando la violación de la restricción de unicidad espacial.

Figura 7

Prueba Unicidad Espacial

```
INSERT INTO inventario.punto_agua VALUES ('110-I-C-8', 'La Catalina', -72.9020643597331,
7.36233626409302, 3063, NULL, '110-I-C', 8, 1129787, 1306068, '68', '68132', '681320023',
'2319', 'ladera', 'ninguno', 'manantial', '2022-02-04',
'01010000204E120000BB9BA73AE43852C00A698D4127A41D40');
```

| Data Output | Messages | Notifications |
|-------------|--|---------------|
| ERROR: | Ya existe la llave (geom)=(01010000204E120000BB9BA73AE43852C00A698D4127A41D40).llave duplicada viola restricción de unicidad «unique_ubicacion_espacial» | |
| ERROR: | llave duplicada viola restricción de unicidad «unique_ubicacion_espacial» | |
| SQL state: | 23505 | |
| Detail: | Ya existe la llave (geom)=(01010000204E120000BB9BA73AE43852C00A698D4127A41D40). | |

Prueba de Llave Primaria. En la figura 8, se muestra la respuesta de la base al intentar registrar un punto con un código ya existente, la base genera un error debido a la violación de la restricción de unicidad, dado que el campo *cod_sgc* corresponde a la llave primaria de la tabla *punto_agua*.

Figura 8

Prueba Unicidad de Punto

```
INSERT INTO inventario.punto_agua VALUES ('110-I-A-1', 'M. Angosturas 1', -
72.8922084754678, 7.41801666873051, 3794, NULL, '110-I-A', 2, 1130859, 1312230, '68',
'68780', '687800203', '2319', 'ladera', 'ninguno', 'manantial', '2022-02-04',
'01010000204E12000082CAF8F7193952C029B000A60CAC1D40');
```

| Data Output | Messages | Notifications |
|-------------|--|---------------|
| ERROR: | Ya existe la llave (cod_sgc)=(110-I-A-1).llave duplicada viola restricción de unicidad «punto_agua_pkey» | |
| ERROR: | llave duplicada viola restricción de unicidad «punto_agua_pkey» | |
| SQL state: | 23505 | |
| Detail: | Ya existe la llave (cod_sgc)=(110-I-A-1). | |

Análisis de Datos Espaciales

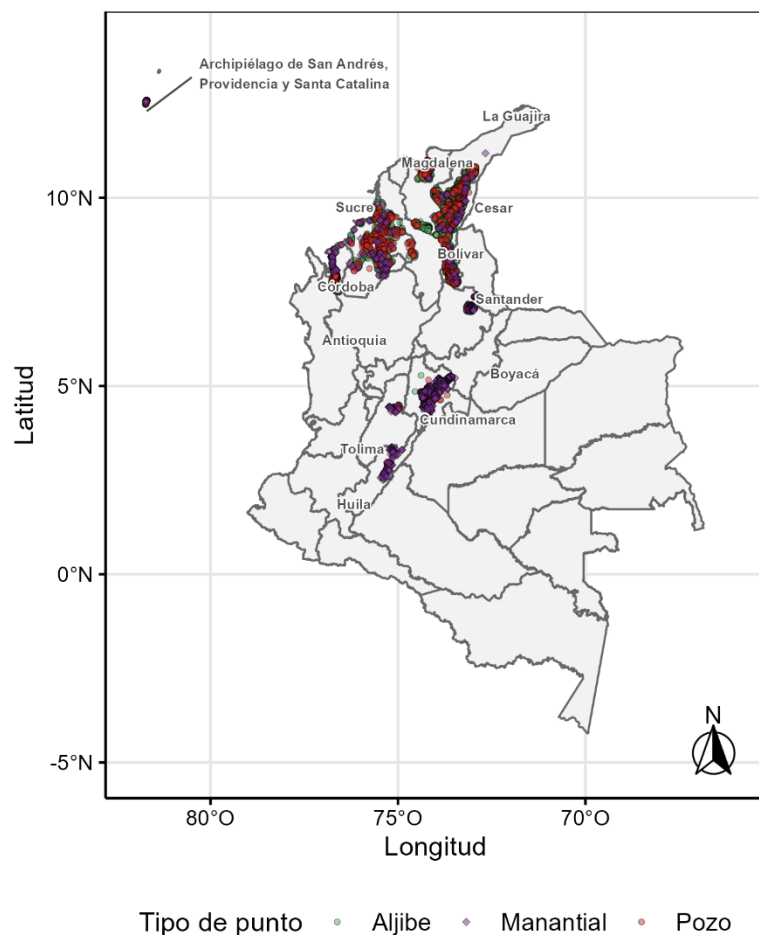
Distribución Espacial de los Puntos de Agua Subterránea en Colombia

Con el propósito de conocer la cobertura espacial del inventario de puntos de agua subterránea, se realizó un análisis de la distribución geográfica de los registros disponibles en la base de datos, considerando su clasificación en aljibes, manantiales y pozos.

La figura 9 presenta la localización espacial de los puntos inventariados. Se observa que la mayor concentración de registros se encuentra en la región Caribe y en sectores de la región Andina, mientras que amplias zonas del sur y oriente del país presentan baja representatividad de información. Este comportamiento sugiere una distribución heterogénea de la información hidrogeológica, asociada posiblemente a factores como el tiempo para el registro de información de a la base de datos y la accesibilidad a las fuentes de agua subterránea.

Figura 9

Distribución Espacial de los Puntos de Agua Registrados en la Base De Datos



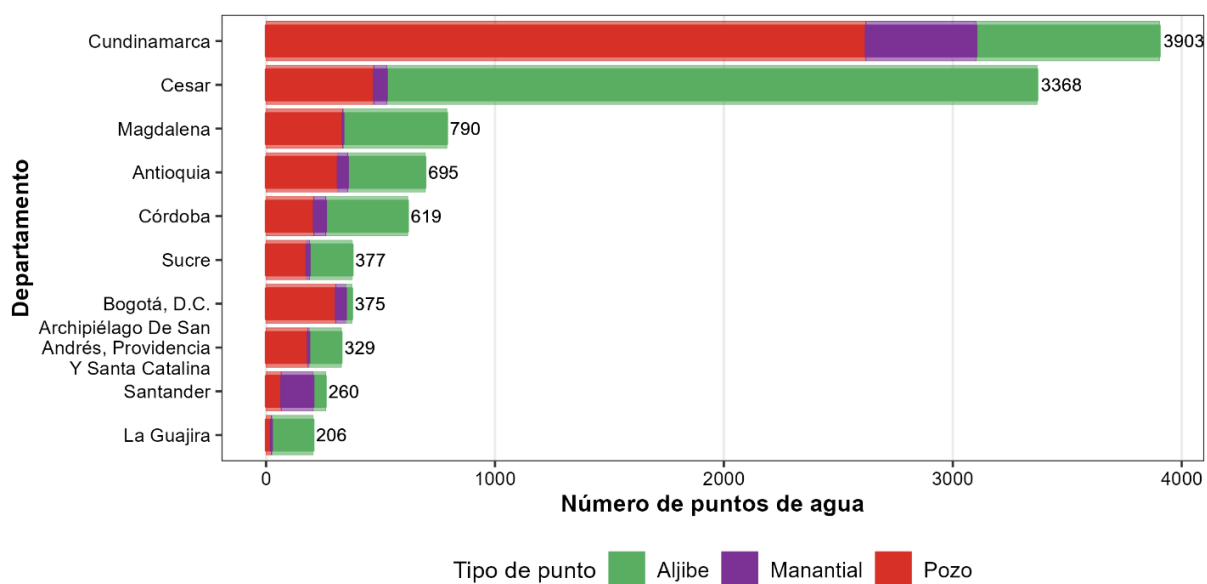
De manera complementaria, la figura 10 muestra el número de puntos por departamento, discriminados según el tipo de captación. Los resultados indican que Cundinamarca concentra el mayor número de registros con 3903 puntos, seguido por el departamento del Cesar con 3368 registros. En ambos departamentos predominan los aljibes, lo cual evidencia el uso tradicional de sistemas de captación somera para abastecimiento doméstico y actividades productivas de pequeña escala.

Otros departamentos como Magdalena, Antioquia y Córdoba presentan valores intermedios, con más de 600 puntos registrados, mientras que Sucre, Bogotá D.C., el Archipiélago de San Andrés, Providencia y Santa Catalina, Santander y La Guajira presentan menores cantidades de registros.

En general, el predominio de aljibes sobre pozos y manantiales evidencia la importancia de las captaciones poco profundas en el aprovechamiento del agua subterránea en las regiones analizadas, lo que puede estar relacionado con condiciones hidrogeológicas favorables, costos de construcción y prácticas tradicionales de abastecimiento.

Figura 10

Departamento con Mayor Cantidad de Registros



Análisis Espacial Detallado del Departamento del Cesar

Dado que el departamento del Cesar presenta uno de los mayores números de registros dentro del inventario nacional, se seleccionó como área de análisis con el fin de evaluar los patrones espaciales de distribución de los puntos de agua subterránea.

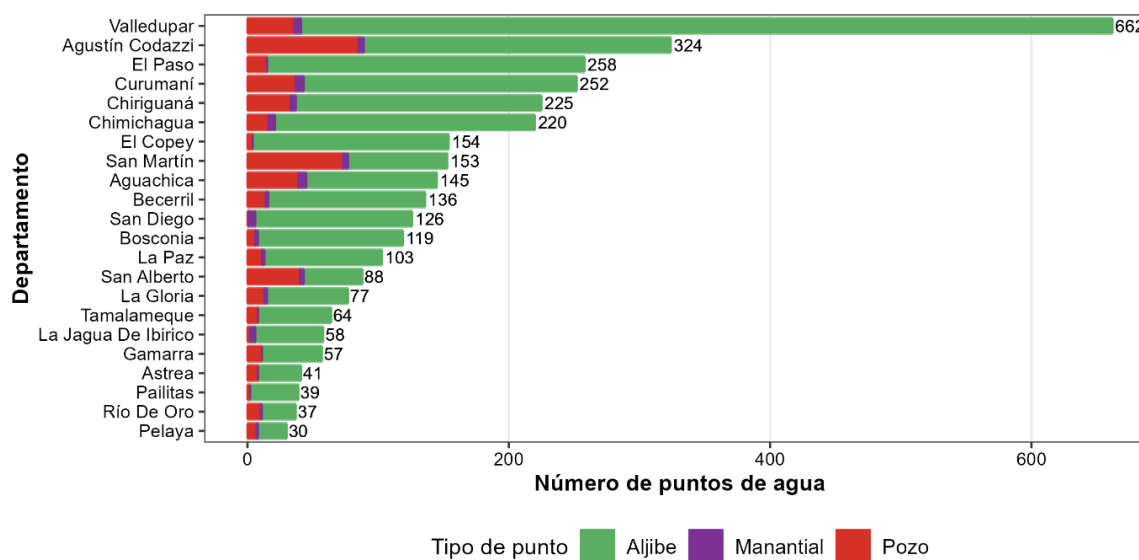
Distribución de Puntos por Municipio y Tipo de Captación. La figura 11 presenta la distribución del número de puntos de agua subterránea por municipio en el departamento del Cesar. Los resultados muestran que Valledupar concentra el mayor número de registros, con 662 puntos, seguido por Agustín Codazzi con 324 registros, El Paso con 258 y Curumaní con 252 puntos.

En la mayoría de los municipios se observa un claro predominio de aljibes frente a los pozos y manantiales, lo cual sugiere una fuerte dependencia de captaciones someras para el abastecimiento en el departamento y puede estar asociado a la facilidad constructiva y disponibilidad de acuíferos someros.

Municipios como Chiriguaná, Chimichagua y El Copey presentan igualmente una cantidad significativa de registros, mientras que municipios como Pelaya, Río de Oro y Pailitas muestran una menor cantidad de puntos, lo que podría indicar vacíos de información hidrogeológica o menor aprovechamiento del recurso subterráneo.

Figura 11

Cantidad de Registros por Municipio

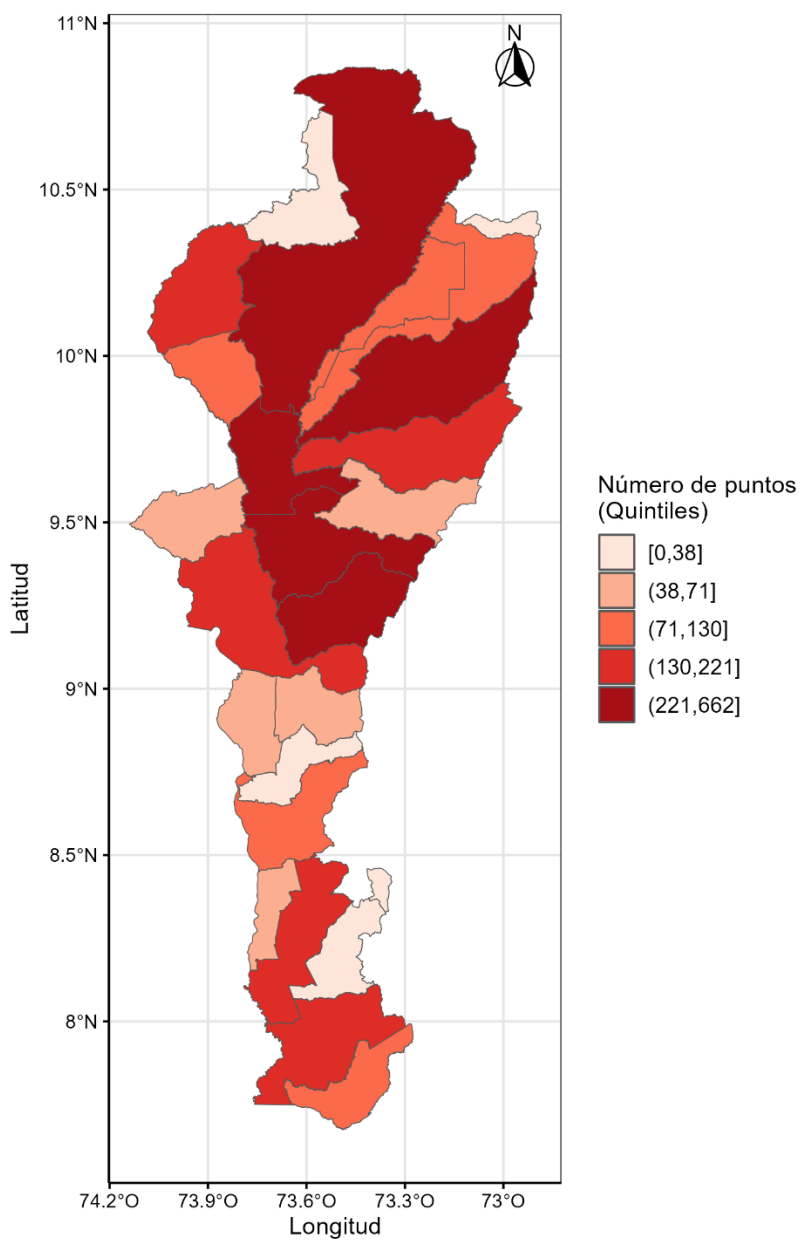


Análisis Espacial de Densidad de Puntos. La figura 12 muestra la distribución espacial del número de puntos de agua subterránea por municipio, clasificados mediante quintiles. Se identifican zonas con alta concentración de registros en el sector centro y norte del departamento, donde se ubican municipios como Valledupar y Agustín Codazzi.

Por el contrario, los municipios del sur del departamento presentan menores densidades de puntos, lo que evidencia posibles áreas con baja cobertura de información hidrogeológica. Estas diferencias espaciales pueden estar relacionadas con variaciones en la disponibilidad del recurso, el desarrollo socioeconómico y la intensidad de las actividades productivas.

Figura 12

Distribución de los Puntos en el Departamento del Cesar



Análisis Estadístico Espacial de Parámetros Fisicoquímicos

Con el fin de evaluar la variabilidad de las características fisicoquímicas del agua subterránea, se analizaron los parámetros de conductividad eléctrica, pH y temperatura, discriminados según el tipo de punto de captación.

Estadística Descriptiva por Tipo de Punto. Los resultados de la estadística descriptiva, de la tabla 4, muestran que la conductividad eléctrica presenta valores medios de 336,19 $\mu\text{S}/\text{cm}$ en aljibes, 308,59 $\mu\text{S}/\text{cm}$ en manantiales y 292,58 $\mu\text{S}/\text{cm}$ en pozos. El coeficiente de variación indica una alta dispersión en los tres tipos de captación, siendo mayor en aljibes, lo que sugiere una mayor heterogeneidad en la mineralización de las aguas en este tipo de captación.

La asimetría positiva observada en todos los casos indica la presencia de valores elevados ocasionales, posiblemente asociados a procesos locales de mineralización.

Tabla 4

Estadística Descriptiva de los Valores de Conductividad Eléctrica ($\mu\text{S}/\text{cm}$) en el Departamento del Cesar

| Tipo de Punto | N | Media | Mediana | Min | Max | Rango | Varianza | SD | CV | Asimetría | Curtosis |
|---------------|------|--------|---------|------|-----|-------|----------|--------|-------|-----------|----------|
| Aljibe | 2781 | 336,19 | 280 | 5,0 | 999 | 994,0 | 81022,34 | 284,64 | 84,67 | 0,57 | -0,83 |
| Manantial | 50 | 308,59 | 270 | 14,0 | 887 | 873,0 | 61038,32 | 247,06 | 80,06 | 0,70 | -0,51 |
| Pozo | 197 | 292,58 | 310 | 6,8 | 930 | 923,2 | 43077,75 | 207,55 | 70,94 | 0,55 | -0,09 |

Nota. SD: desviación estándar; CV: coeficiente de variación.

En cuanto al pH, los valores promedio se encuentran cercanos a la neutralidad, con medias alrededor de 7 en los tres tipos de captación. La baja variabilidad observada sugiere condiciones geoquímicas relativamente estables, compatibles con sistemas acuíferos dominados

por interacciones agua-roca de carácter carbonatado que mantienen las aguas neutras (ver tabla 5).

Tabla 5

Estadística Descriptiva de los Valores de pH en el Departamento del Cesar

| Tipo de Punto | N | Media | Mediana | Min | Max | Rango | Varianza | SD | CV | Asimetría | Curtosis |
|---------------|------|-------|---------|-----|-----|-------|----------|------|-------|-----------|----------|
| Aljibe | 2548 | 7,05 | 7 | 4 | 10 | 6 | 0,77 | 0,88 | 12,47 | -0,45 | -0,21 |
| Manantial | 51 | 6,78 | 7 | 5 | 8 | 3 | 0,64 | 0,80 | 11,82 | -0,15 | -0,69 |
| Pozo | 208 | 7,03 | 7 | 5 | 9 | 4 | 0,37 | 0,61 | 8,65 | -0,29 | 1,18 |

SD: desviación estándar; CV: coeficiente de variación.

Respecto a la temperatura, los valores medios oscilan entre 29 °C y 30 °C, reflejando condiciones térmicas propias de ambientes tropicales. Los bajos coeficientes de variación indican estabilidad térmica del agua subterránea (tabla 6).

Tabla 6

Estadística Descriptiva de los Valores de Temperatura (°C) en el Departamento del Cesar

| Tipo de Punto | n | Media | Mediana | Min | Max | Rango | Varianza | SD | CV | Asimetría | Curtosis |
|---------------|------|-------|---------|------|------|-------|----------|------|------|-----------|----------|
| Aljibe | 2922 | 29,79 | 30,00 | 9,2 | 37,6 | 28,4 | 2,57 | 1,60 | 5,38 | -1,22 | 16,16 |
| Manantial | 48 | 29,00 | 29,00 | 25,0 | 35,2 | 10,2 | 3,56 | 1,89 | 6,51 | 0,80 | 2,07 |
| Pozo | 194 | 29,67 | 29,95 | 19,7 | 38,0 | 18,3 | 3,49 | 1,87 | 6,30 | -0,75 | 8,65 |

SD: desviación estándar; CV: coeficiente de variación.

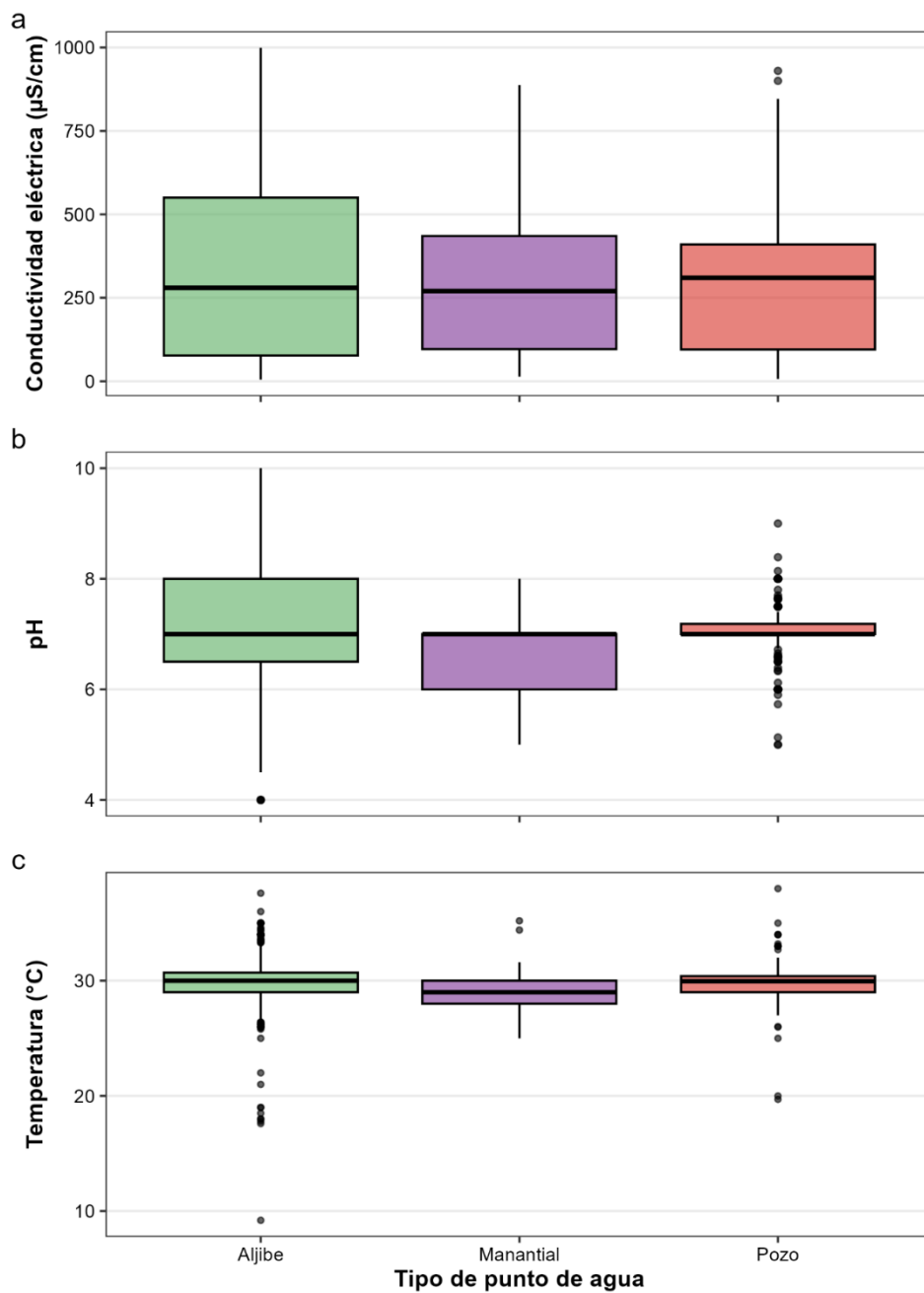
Variabilidad de los Parámetros Físicoquímicos. La figura 13 presenta la distribución de la conductividad eléctrica (CE), el pH y la temperatura del agua, diferenciados por tipo de punto de captación (aljibe, manantial y pozo). El número de observaciones es desigual entre

categorías, con un mayor número de registros en aljibes (CE: $n = 2781$; pH: $n = 2548$; temperatura: $n = 2922$), frente a manantiales (CE: $n = 50$; pH: $n = 51$; temperatura: $n = 48$) y pozos (CE: $n = 197$; pH: $n = 208$; temperatura: $n = 194$).

En cuanto a la conductividad eléctrica, los tres tipos de punto presentan medianas comparables, situadas aproximadamente entre 250 y 300 $\mu\text{S}/\text{cm}$. No obstante, los aljibes muestran la mayor dispersión, con un rango intercuartílico amplio y valores máximos cercanos a 1000 $\mu\text{S}/\text{cm}$. Los pozos presentan una variabilidad intermedia, con algunos valores altos aislados, mientras que los manantiales exhiben la menor dispersión, reflejando una distribución más homogénea.

El pH se concentra mayoritariamente en valores neutros a ligeramente ácidos. Los aljibes presentan la mayor variabilidad, con registros que abarcan desde valores ácidos (<5) hasta valores claramente básicos (>9). En contraste, los pozos muestran una distribución más concentrada alrededor de pH neutro, con un rango intercuartílico reducido. Los manantiales presentan valores intermedios tanto en dispersión como en tendencia central.

La temperatura del agua presenta medianas similares entre los tres tipos de punto, alrededor de 29–30 °C. Sin embargo, los aljibes muestran una mayor variabilidad térmica, incluyendo valores notablemente bajos y algunos altos, estos valores atípicos podrían estar asociados a anomalía en los puntos de medición o a errores instrumentales o de digitación. Los manantiales exhiben la distribución más estrecha, mientras que los pozos presentan una dispersión moderada, con algunos valores elevados.

Figura 13*Distribución de los Parámetros en Campo por Tipo de Punto*

Clasificación Espacial de la Calidad del Agua Según Conductividad Eléctrica

La figura 14 muestra la distribución espacial de la calidad del agua subterránea clasificada a partir de la conductividad eléctrica (CE), empleando rangos que permiten diferenciar aguas no salinas ($< 700 \mu\text{S/cm}$), ligeramente salinas ($700\text{--}2000 \mu\text{S/cm}$), moderadamente salina ($2000\text{--}10000 \mu\text{S/cm}$), Altamente salina ($10000\text{--}25000 \mu\text{S/cm}$), Extremadamente salina ($25000\text{--}45000 \mu\text{S/cm}$) y salmueras ($>45000 \mu\text{S/cm}$), de acuerdo con la escala propuesta por Rodier *et al.* (1981).

A escala regional, se identifica una predominancia de aguas no salinas, las cuales se distribuyen de forma amplia y continua en la mayor parte del departamento. Este patrón es consistente con los valores centrales de conductividad eléctrica identificados en el análisis estadístico descriptivo, y sugiere niveles moderados de mineralización.

Las aguas ligeramente salinas aparecen de manera localizada y discontinua, la distribución espacial fragmentada indica que el incremento de la conductividad eléctrica responde a condiciones locales, posiblemente asociadas a variaciones litológicas, diferencias en el tiempo de residencia del agua, o a la influencia puntual de procesos antrópicos. La ausencia de grandes extensiones continuas con esta categoría sugiere que no se trata de un proceso regionalizado, sino de fenómenos de escala local.

En particular, la dominancia de la categoría no salina, junto con la aparición puntual de aguas de mayor conductividad, refuerza la interpretación de un sistema en el que predominan condiciones naturales, pero con anomalías locales.

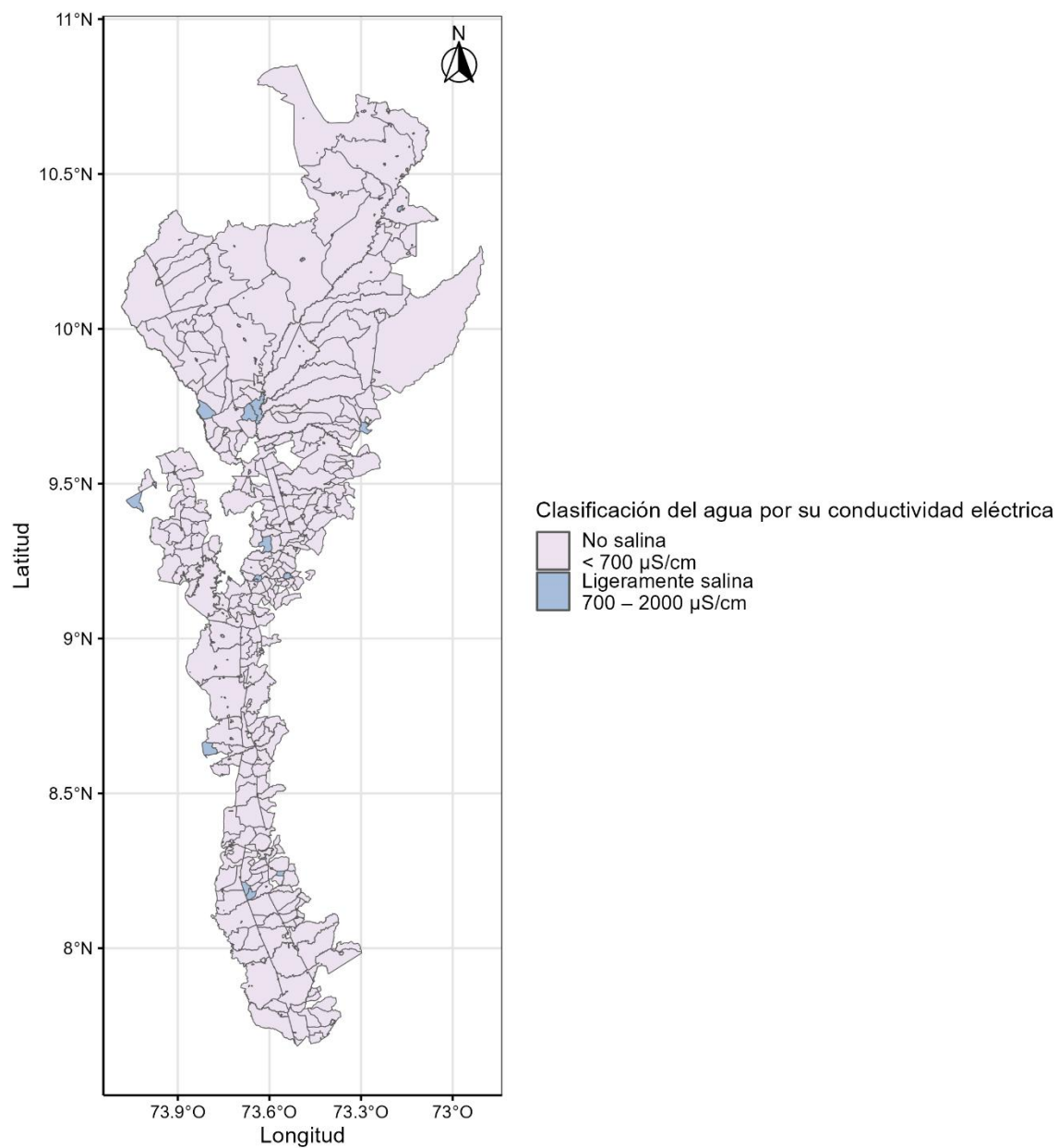
Es importante señalar que esta clasificación se basa en datos históricos medidos *in situ*, para los cuales no fue posible aplicar un control de calidad posterior. En este contexto, los valores elevados de conductividad eléctrica no se interpretan como errores, sino como registros

válidos que reflejan la variabilidad reportada, al igual que lo valores de conductividad bajos.

Dado el enfoque descriptivo del análisis, el mapa permite identificar patrones espaciales generales y áreas con particularidades, sin pretender establecer relaciones causales ni evaluar tendencias temporales.

Figura 14

Clasificación del Agua por su Conductividad Eléctrica



El mapa de clasificación por tipo de agua es un ejemplo de la aplicabilidad del análisis espacial para abordar problemáticas asociadas a la calidad del agua subterránea. En particular, la distribución espacial de categorías relacionadas con niveles de salinidad permite identificar zonas donde el uso del agua podría representar limitaciones para el consumo humano o la actividad agrícola.

Este resultado es posible gracias a la integración de la información hidrogeológica en la base de datos, la cual permite realizar consultas orientadas a discriminar puntos o áreas según su clasificación. De esta manera, la base de datos trasciende el almacenamiento de información y se consolida como una herramienta de apoyo a la toma de decisiones sobre el recurso hídrico.

Conclusiones

El diseño de la base de datos relacional logró la estructuración sistemática y normalizada de la información hidrogeológica del Servicio Geológico Colombiano, respondiendo a la complejidad y diversidad de los datos asociados al inventario de agua subterránea y resultados de muestreos químicos. La aplicación de principios de normalización y la definición explícita de entidades, relaciones y dominios temáticos redujeron redundancias y mejoraron la consistencia de los registros.

La separación de la información en esquemas temáticos y la definición de relaciones jerárquicas claras, particularmente entre la entidad el punto de agua, las visitas y sus mediciones asociada, permitió representar adecuadamente la lógica hidrogeológica del sistema, diferenciando entre características estáticas del punto y observaciones dinámicas en el tiempo.

El proceso de consolidación de los datos evidenció la heterogeneidad estructural y semántica de las fuentes históricas, así como la ausencia de mecanismos de validación en los sistemas de origen. Mediante la aplicación de un proceso sistemático de extracción, transformación y carga (ETL), fue posible integrar información proveniente de múltiples tablas y formularios, preservando la trazabilidad de los registros originales y mejorando significativamente su calidad.

La implementación de dominios, restricciones de integridad y validaciones automáticas permitió identificar y corregir inconsistencias relacionadas con duplicidades, errores temporales, valores fuera de rango y categorías no estandarizadas.

La implementación técnica de la base de datos en PostgreSQL con extensión PostGIS permitió gestionar atributos alfanuméricos y geométricos, habilitando la realización de análisis espaciales sobre información hidrogeológica previamente dispersa y heterogénea. La validación

de geometrías, la estandarización del sistema de referencia espacial y la incorporación de relaciones territoriales mediante cruces espaciales garantizaron la coherencia espacial de los datos.

En conjunto, los resultados confirman que la base de datos desarrollada es una herramienta estratégica para el fortalecimiento de la gestión de la información y el conocimiento hidrogeológico en el SGC, al facilitar el acceso, la consulta y la reutilización de los datos de agua subterránea en Colombia.

Recomendaciones

Se recomienda continuar con el proceso de actualización y poblamiento de la base de datos, incorporando nuevos puntos, campañas de muestreo y resultados de análisis de laboratorio.

Se recomienda formalizar protocolos institucionales para la captura, estandarización y almacenamiento de la información hidrogeológica, alineados con la estructura y los dominios definidos en la base de datos, con el fin de asegurar la consistencia de los datos desde su origen y reducir la necesidad de procesos de depuración posteriores.

Se sugiere potenciar la interoperabilidad de la BDH con sistemas de información geográfica (SIG), lo que permitiría ampliar su uso en análisis espaciales y apoyo a procesos institucionales.

Se propone profundizar en la aplicación de técnicas de análisis espacial y estadístico más avanzadas, como análisis de autocorrelación espacial, modelos geoestadísticos o análisis multivariado, aprovechando la estructura y calidad de la información almacenada.

Finalmente, se recomienda promover la capacitación del personal técnico en el uso y administración de la base de datos, así como en el análisis de la información hidrogeológica, con el propósito de garantizar la sostenibilidad del sistema y maximizar su impacto en la gestión del agua subterránea a nivel institucional y nacional.

Referencias Bibliográficas

- Agencia Nacional de Licencias Ambientales (ANLA). (2021). *Términos de referencia para la exploración de hidrocarburos*. Agencia Nacional de Licencias Ambientales.
https://www.anla.gov.co/documentos/normativa/terminos_referencia/TdR_exploracion_HC_VF_Para_adopcion_sin_CC.pdf
- Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2), 93-115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- Buzai, G. D., y Baxendale, C. (2006). *Análisis Socioespacial con Sistemas de Información Geográfica*. Lugar Editorial.
- Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6), 377-387.
- Connolly, T. M., y Begg, C. E. (2015). *Database Systems: A Practical Approach to Design, Implementation, and Management* (Sixth). Pearson.
- Custodio, E., y Llamas, M. R. (1996). Hidrología Subterránea. En Ediciones Omega S. A. (Ed.), *Omega S. A.* (Número Tomos I y II).
- Date, C. J. (2004). *An Introduction to Database Systems*. Addison-Wesley.
- De Corso Sicilia, G. B., y Pinilla Rivera, M. (2017). Métodos gráficos de análisis exploratorio de datos espaciales con variables espacialmente distribuidas. *Cuadernos Latinoamericanos de Administración*, 13(25), 92-104. <https://doi.org/10.18270/cuaderlam.v13i25.2417>
- Elmasri, R., y Navathe, S. B. (2015). *Fundamentals of Database Systems* (7th ed.). Addison-Wesley.
- Instituto de Hidrología Meteorología y Estudios Ambientales. (2023). *Estudio Nacional del Agua 2022*. Ideam.

- Mendoza, A., y López, R. (2018). *Bases de datos*.
- Ministerio de Ambiente y Desarrollo Sostenible. (2014). Programa Nacional de Aguas Subterráneas. *Ministerio de Ambiente y Desarrollo Sostenible, 1*, 3-22.
- Nina, C., y Gissela, M. (2017). *Sistema de control de personal y tareaje para el proyecto de nutrición infantil del Gobierno Regional Puno-2015*.
- Rodier, J. A. (Jean A. ., Geoffray, C., Rodi, L., y Balagué Dolz, S. (1981). *Análisis de las aguas: aguas naturales, aguas residuales, agua de mar : química, fisicoquímica, bacteriológica, biología*. <https://dialnet.unirioja.es/servlet/libro?codigo=94445>
- Servicio Geológico Colombiano. (s. f.). *Programa de Aguas Subterráneas*.
- Silberschatz, A., Korth, H. F., Sudarshan, S., Pérez, F. S., Santiago, A. I., y Sánchez, A. V. (2002). *Fundamentos de bases de datos* (Vol. 11). McGraw-Hill Ciudad de México, México.
- Velasco, V., Criollo, R., Vázquez-Suñè, E., Alcaraz, M., Serrano, A., García-Gil, A., y Gogu, R. (2013). GIS-based Hydrogeological database and analysis tools. *Environmental Modelling & Software*.

Apéndices

Apéndice A

Modelo de Entidad-Relación

Este apéndice contiene el modelo entidad–relación de la base de datos desarrollada. Dada la complejidad técnica y el nivel de detalle del esquema, su representación excede las dimensiones del formato del documento; por esta razón, el modelo se presenta en el siguiente enlace: [Modelo-ER](#).

Apéndice B

Diccionario de la Base de Datos

El presente apéndice contiene el diccionario de la base de datos, el cual describe las entidades, atributos, tipos de datos, dominios y la restricciones de integridad del sistema. Debido a su extensión, el diccionario se presenta en formato digital y puede ser consultado a través del siguiente enlace: [Diccionario BaseDatos](#).