



Universidad Nacional
Federico Villarreal

VRIN | VICERRECTORADO
DE INVESTIGACIÓN

FACULTAD DE INGENIERÍA GEOGRÁFICA, AMBIENTAL Y ECOTURISMO
DETECCIÓN Y CORRECCIÓN AUTOMATIZADA DE ERRORES EN LA
CONSISTENCIA LÓGICA DE DATOS GEOGRÁFICOS

Línea de investigación: Desarrollo Urbano-Rural

Informe de suficiencia profesional para optar el título profesional de Ingeniero Geógrafo

Autor:

Lopez Rengifo, Charlie Edwin

Asesor:

Aguirre Cordero, Rogelio

(ORCID: 0000-0003-3093-8756)

Jurado:

Fernández Ybarra, Felicita Nancy

Gonzales Alarcón, Angelino Oscar

Herrera Diaz, Marco Antonio

Lima - Perú

2023

DETECCIÓN Y CORRECCIÓN AUTOMATIZADA DE ERRORES EN LA CONSISTENCIA LÓGICA DE DATOS GEOGRÁFICOS

INFORME DE ORIGINALIDAD

18%

INDICE DE SIMILITUD

17%

FUENTES DE INTERNET

6%

PUBLICACIONES

4%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	docplayer.es Fuente de Internet	2%
2	bibliotecadigital.ciren.cl Fuente de Internet	2%
3	repositorio.unfv.edu.pe Fuente de Internet	1%
4	Submitted to Universidad Internacional de la Rioja Trabajo del estudiante	1%
5	www.coursehero.com Fuente de Internet	1%
6	www.ign.es Fuente de Internet	1%
7	Enric Terol Esparza. "Determinación y análisis de los factores críticos en la implantación de un SIG en la administración local.", Universitat Politecnica de Valencia, 2015 Publicación	1%

Dedicatoria

A mis padres, Carlos y Sulmira, a quienes debo mi inspiración y cuyo apoyo inquebrantable ha sido el motor detrás de mi determinación a lo largo de este desafiante viaje académico. Agradezco sinceramente su fe en mí, incluso en los momentos de duda, su amor incondicional cuando más lo necesitaba y por ser mis eternos maestros.

A mi esposa, Yessica, y a mi hija, Luana, por su apoyo incondicional y ser la fuente inagotable de inspiración en esta búsqueda de conocimiento y éxito. Su amor y aliento son mi mayor fortaleza, y este logro es un testimonio de nuestra unidad como familia.

INDICE

RESUMEN.....	6
ABSTRACT.....	7
I. INTRODUCCIÓN	8
1.1. Trayectoria Del Autor.....	9
1.2. Descripción de la Empresa	10
1.3. Organigrama de la Empresa	12
1.4. Áreas y Funciones Desempeñadas	12
II. DESCRIPCIÓN DE UNA ACTIVIDAD ESPECIFICA	15
2.1. Objetivo de la actividad.....	15
2.2. Antecedentes.....	15
2.3. Marco normativo	17
2.4. Marco conceptual	18
2.5. Procedimiento.....	21
III. APORTES MÁS DESTACABLES A LA EMPRESA.....	32
IV. CONCLUSIONES.....	34
V. RECOMENDACIONES.....	35
VI. REFERENCIAS	36
VII. ANEXOS	38
Anexo A. Funciones implementadas para la detección de errores.....	38
Anexo B. Funciones implementadas para la corrección de errores	39

ÍNDICE DE TABLAS

Tabla 1	Elementos y subelementos de calidad.....	20
Tabla 2	Datos geográficos a evaluar	22
Tabla 3	Requisitos de software	23
Tabla 4	Plantilla de especificaciones técnicas	25
Tabla 5	Reglas de consistencias lógicas	26
Tabla 6	Matriz de datos geográficos y reglas de consistencia lógica	27
Tabla 7	Funciones implementadas para la detección de errores	38
Tabla 8	Funciones implementadas para la corrección de errores	39

ÍNDICE DE FIGURAS

Figura 1	Organigrama del Área de Data y Analytics - Studio GEO (EQUIFAX)	12
Figura 2	Flujo para la detección y corrección de errores en la consistencia lógica.....	21
Figura 3	Pasos para la elaboración de las especificaciones técnicas	24
Figura 4	Estructura de las funciones.....	28
Figura 5	Ejecución de procesos utilizando programación procedimental	29
Figura 6	Errores de superposición detectados y corregidos para una capa poligonal	30
Figura 7	Errores de anillos interiores detectados y corregidos	30
Figura 8	Errores de ID duplicados detectados y corregidos	31

RESUMEN

El objetivo del presente informe es describir la metodología empleada para realizar la detección y corrección automatizada de errores que afectan la consistencia lógica de los datos geográficos. La metodología propuesta fue diseñada tomando como marco de referencia las normas ISO 19131:2022 y 19157:2013, que establece las directrices para las especificaciones de producto de datos geográficos y la calidad de la información geográfica, respectivamente. Con respecto a la automatización del proceso, se llevó a cabo empleando el lenguaje de programación Python como principal herramienta. Mediante el uso de Python y las librerías especializadas en el manejo de datos geoespaciales, se lograron identificar y resolver de manera eficiente las discrepancias y contradicciones presentes que afectaban la consistencia lógica de los datos geográficos. Finalmente, como resultado de la automatización de los procesos de detección y corrección de errores, se logró mejorar la calidad y confiabilidad de los datos espaciales, además de optimizar los tiempos de procesamiento al agilizar las tareas de corrección y preparación de datos para análisis y aplicaciones más eficientes.

Palabras clave: calidad, consistencia lógica, dato espacial, ISO 19131:2022, ISO 19157:2013

ABSTRACT

The aim of this report is to describe the methodology used to perform automated detection and correction of errors affecting the logical consistency of geographic data. The proposed methodology has been developed using as a reference framework the ISO 19131:2022 and 19157:2013 standards, which provide guidelines for the specification of geographic data products and the quality of geographic information, respectively. With regard to the automation of the process, it was carried out using the Python programming language as a central tool. By using Python and specialized geospatial libraries, discrepancies and inconsistencies affecting the logical consistency of the geographic data were identified and resolved efficiently. Finally, by automating the error detection and correction processes, the quality and reliability of the spatial data were improved, and processing times were optimized by speeding up data correction and preparation tasks for more efficient analysis and applications.

Keywords: quality, logical consistency, spatial data, ISO 19131:2022, ISO 19157:2013

I. INTRODUCCIÓN

La consistencia lógica se refiere al componente de la calidad que garantiza que las relaciones y atributos de los datos geográficos se mantengan de manera coherente, precisa y sin contradicciones internas. En términos simples, la consistencia lógica establece que las relaciones espaciales entre elementos, como puntos, líneas y polígonos, sigan una lógica coherente, y que los atributos espaciales asociados sean congruentes entre sí. De esta manera, esta propiedad permite evitar discrepancias que comprometen la confiabilidad y utilidad de los datos geográficos en la toma de decisiones.

En la actualidad, la problemática de detectar y corregir errores de consistencia lógica reside en el incremento exponencial de la cantidad y diversidad de los datos geográficos. Por lo tanto, la tarea de identificar y corregir estos errores se vuelve cada vez más complejo, laborioso y propenso a errores humanos. Bajo este contexto, surge la necesidad de la automatización de procesos que contribuyan, de forma eficiente y efectiva, en la detección y corrección de inconsistencias que afectan la calidad de los datos geográficos.

Por tal motivo, el presente informe tiene como objetivo describir la metodología utilizada para optimizar el proceso de detección y corrección de errores que afectan la consistencia lógica de los datos geográficos. Mediante la combinación de las normas ISO 19131:2022 e ISO 19157:2013, junto con la potencia del lenguaje de programación Python, se estableció un enfoque sólido y eficaz para abordar esta problemática.

De esta manera, el uso de las normas ISO 19131:2022 e ISO 19157:2013 como marco de referencia, permitió establecer las especificaciones técnicas de los datos geográficos y definir las reglas necesarias para evaluar los errores que afectan la consistencia lógica de los mismos. Además, al utilizar Python como herramienta de automatización, se obtuvo la capacidad de detectar y corregir estos errores de manera efectiva.

1.1. Trayectoria Del Autor

El autor del presente informe posee el título académico de Bachiller en Ingeniería Geográfica (desde 2013) otorgado por la Universidad Nacional Federico Villarreal. Desde su graduación profesional, el autor ha acumulado una amplia experiencia de más de 10 años en el sector privado, centrada en áreas claves como Sistemas de Información Geográfica (SIG), Bases de Datos Espaciales, Programación SIG, Análisis de Datos Espaciales, Geomarketing, Geocodificación de Direcciones y Gestión de Proyectos.

Experiencia Laboral

De mayo del 2012 a julio del 2015, laboro en la empresa peruana MAPCITY PERU S.A.C., compañía especializada en soluciones geoespaciales y consultorías de análisis territorial. En MAPCITY, comenzó con el cargo de analista SIG y más tarde ascendió al puesto de jefe del área de datos geocodificados liderando un equipo de cuatro personas. Durante este período, lideró el desarrollo y la ejecución de diversos proyectos relacionados con estudios de geomarketing, bases de datos espaciales y aplicaciones SIG.

De octubre del 2015 hasta la fecha, se desempeña en la empresa multinacional EQUIFAX ocupando el cargo de Líder de Datos Geográficos para América Latina. En esta posición, supervisa un equipo multidisciplinario de nueve personas dedicadas al desarrollo de proyectos relacionados con el mantenimiento, actualización, procesamiento y análisis de los datos geográficos de la empresa. Su responsabilidad fundamental radica en proporcionar datos actualizados y de alta calidad que son esenciales para la elaboración de estudios, ventas directas o el suministro a través de plataformas SIG.

Con esta experiencia, el autor ha demostrado un compromiso constante con el ámbito de los SIG, Bases de Datos Espaciales, Programación SIG, Análisis de Datos Espaciales,

Geomarketing, Geocodificación de Direcciones y Gestión de Proyectos, contribuyendo significativamente al éxito de las empresas en las que ha trabajado.

Estudios

- UNIVERSIDAD POLITÉCNICA DE VALENCIA
2021 - 2022, Infraestructura de Datos Espaciales y Geoportales
2021 - 2022, Base de Datos Espaciales: PostGIS
- UNIVERSIDAD PRIVADA DEL NORTE
2015 - 2017, Egresado de Ingeniería de Sistemas en Pregrado Adulto Trabajador
- PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERU
2012 - 2013, Especialización en Planificación territorial
- UNIVERSIDAD NACIONAL FEDERICO VILLAREAL
2007 - 2012, Bachiller de Ingeniería Geográfica

1.2. Descripción de la Empresa

Equifax es una empresa global de servicios de tecnología, analítica e información crediticia con presencia en más de 25 países. Equifax ofrece una amplia gama de servicios relacionados con la gestión del riesgo crediticio, protección contra el fraude, marketing dirigido, la segmentación de audiencias, analítica avanzada y soluciones tecnológicas. La empresa posee una vasta base de datos que utiliza para ayudar a las empresas a comprender mejor a sus clientes e identificar oportunidades de crecimiento.

En el año 2017, Equifax adquirió Mapcity, empresa chilena especializada en servicios de soluciones geoespaciales. La adquisición de Mapcity por parte de Equifax permitió a la empresa expandir su oferta de servicios y fortalecer su presencia en el campo de la información

espacial. En la actualidad Equifax, a través de su área “Data y Analytics - Studio GEO” proporciona las siguientes soluciones geoespaciales:

- **Datos espaciales:** Conjunto de datos georreferenciados que proporciona información detallada y actualizada del entorno geográfico.
- **Geocodificación de direcciones:** Proceso mediante el cual se transforman las descripciones de ubicaciones en coordenadas geográficas (latitud y longitud).
- **Enriquecimiento de bases de datos:** Permite agregar información espacial a conjuntos de datos preexistentes según su ubicación, mejorando así su valor y utilidad.
- **Estudios de Geo-analítica:** Elaboración de estudios que implican el uso de datos espaciales para comprender y analizar la distribución de la población, la ubicación de negocios, características socioeconómicas y demográficas y otras relevantes en un área geográfica específica. Estos estudios brindan información que contribuye a tomar decisiones informadas y precisas en una amplia variedad de áreas y sectores.
- **Plataformas SIG:** Esta solución permite a los usuarios explorar mapas interactivos, acceder a datos espaciales y realizar análisis espaciales directamente en navegadores web, sin la necesidad de instalar software especializado.

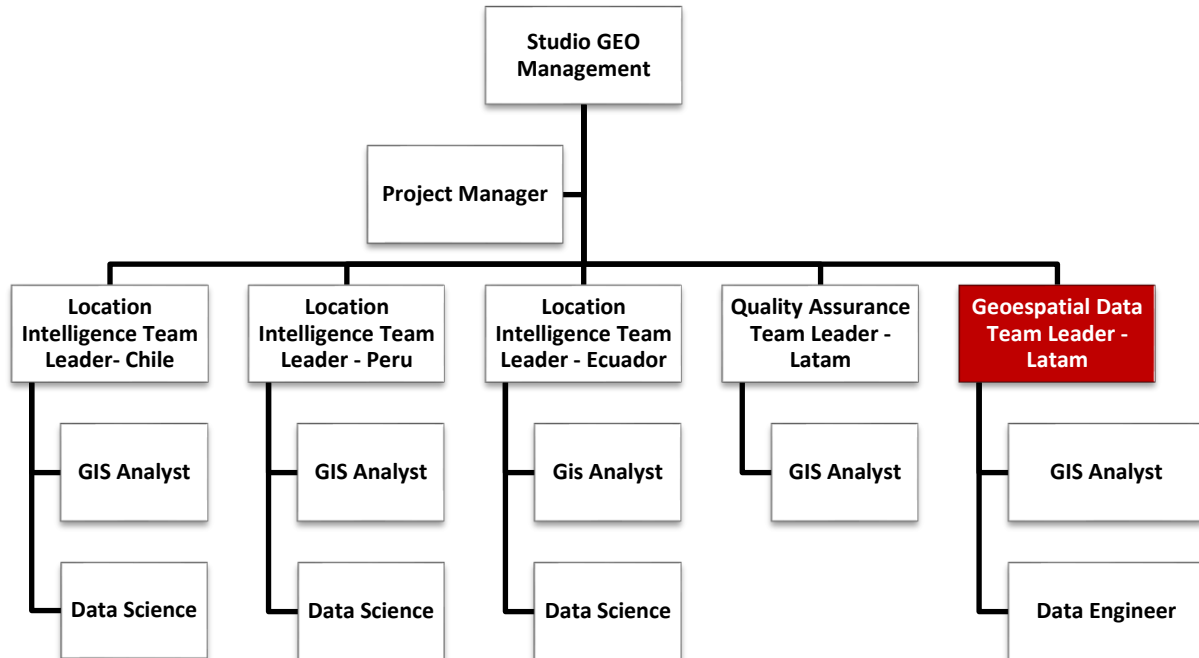
Visión

Ser un líder global confiable en datos, analytics avanzados y tecnología que crea soluciones y perspectivas innovadoras para ayudar a los clientes a impulsar el crecimiento y fomentar el progreso de las personas (EQUIFAX, 2023).

1.3. Organigrama de la Empresa

Figura 1

Organigrama del Área de Data y Analytics - Studio GEO (EQUIFAX)



1.4. Áreas y Funciones Desempeñadas

Equifax Perú (Área: Data y Analytics - Studio Geo):

El autor ocupa el cargo de Data Analytics and Wrangling desempeñando hasta la actualidad la función de Líder de datos Geográficos para América Latina y previamente la de jefe de Datos Geocodificados para Perú. Las actividades realizadas son la siguientes:

- Gestión de proyectos Geo-Analíticos para el desarrollo y ejecución de estudios de geomarketing, geocodificación de direcciones e implementación de plataformas SIG.
- Gobierno de datos geográficos, mediante la definición de políticas, procesos, estándares y procedimientos diseñados para administrar de

manera efectiva y eficiente la información geográfica de la organización.

- Elaboración y ejecución del plan anual de mantenimiento y actualización de datos geográficos utilizados para estudios, aplicaciones y sistemas de la organización.
- Procesamiento y análisis de Big Data Espacial para el mantenimiento y actualización de datos geográficos, elaboración de estudios geo analíticos e implementación de plataformas GIS. Estas tareas se llevan a cabo mediante el uso de herramientas tecnológicas avanzadas como Python y Spatial SQL.
- Ejecución del proceso de Geocodificación de direcciones de clientes externos e internos mediante el uso de algoritmo diseñados en PostgreSQL/PostGIS.
- Documentación de los datos geográficos y actualización de metadatos.
- Capacitaciones técnicas y soporte al equipo comercial y a las diferentes áreas de la organización que consumen datos geográficos.

Mapcity Perú (Área: Datos Geocodificados):

De mayo del 2012 a julio del 2015 ocupó el cargo de Analista GIS realizando las siguientes actividades:

- Ejecución del proceso de Geocodificación de direcciones de clientes externos e internos mediante el uso de algoritmo diseñados en PostgreSQL/PostGIS.
- Mantenimiento y actualización de datos geográficos utilizados para

estudios, aplicaciones y sistemas de la organización.

- Migración de datos espaciales de archivos shapefile a base de datos PostGIS.
- Elaboración de estudios de geomarketing y análisis territorial para la toma de decisiones empresariales basadas en información geográfica.
- Automatización de geoprosos para la elaboración de estudios de Geomarketing y análisis territorial utilizando Modelbuilder de ArcGIS.

II. DESCRIPCIÓN DE UNA ACTIVIDAD ESPECIFICA

En esta sección, se desarrolla la actividad específica para la detección y corrección automatizada de errores que afectan la consistencia lógica de los datos geográficos. En este contexto, se describen los objetivos de la actividad y se hace referencia a investigaciones previas relevantes. Además, se define el marco normativo utilizado y el marco conceptual de los términos claves. Finalmente, se presenta una descripción detallada del procedimiento utilizado con la presentación de los resultados obtenidos en cada una de las etapas.

2.1. Objetivo de la actividad

El desarrollo de esta actividad cuenta con dos objetivos claro. En primer lugar, se busca desarrollar una metodología que permita la detección y corrección automatizada de errores en los datos geográficos, con el propósito de asegurar la calidad y confiabilidad de los análisis y decisiones basadas en información geográfica. En segundo lugar, se pretende optimizar de manera significativa los recursos y el tiempo que el equipo de Studio GEO dedica a la corrección de estos errores.

2.2. Antecedentes

En el año 2017, Jorge Méndez en la tesis “Mejoramiento de Calidad en conjuntos de Datos Abiertos basado en la aplicación de métricas de Consistencia Lógica”, realizada en la Universidad Distrital Francisco José de Caldas (Colombia), tuvo como objetivo diseñar un conjunto de reglas que permita crear las métricas de consistencia lógica para evaluar cada conjunto de datos publicado en la plataforma distrital de datos abiertos para identificar qué registros del conjunto de datos no cumplen con las métricas de consistencia lógica y así obtener el índice de calidad según cada métrica. Como conclusión, se obtuvo que el modelo de calidad propuesto es acertado en la búsqueda de deficiencias en los datos abiertos, y además que este

puede ser mejorado para integrar reglas de consistencia lógica en los campos de los diferentes conjuntos de datos.

En el año 2019, Angie Yáñez en el trabajo de titulación “Evaluación de la calidad de datos Geográficos de la Cartografía Urbana del Cantón Pujilí escala 1:1000”, realizada en la Universidad de las Fuerzas Armadas (Ecuador), tuvo como objetivo determinar la calidad cartográfica escala 1:1000 del Cantón Pujilí, mediante la aplicación de normativas internacionales, llegando a la conclusión que el total de la cartografía presenta mala calidad (según los indicadores establecidos en esta investigación), por lo tanto, no se aceptan los datos geográficos como correctos.

En el año 2020, Juan Cuervo en la tesis titulada “Propuesta para Gestión y Control de la Información Espacial (GDB), Alfanumérica y Temática de la Cartografía Base de Isa para Incorporación en el Sistema de Información Territorial para la Construcción y Operación - SITCO”, realizada en la Universidad Distrital Francisco José de Caldas (Colombia), tuvo como objetivo estructurar la información geográfica que se encuentra en la base de datos espacial de INTERCOLOMBIA con base en la normativa NTC para Colombia con relación al metadato geográfico y a la calidad de la información geográfica. Como conclusión se obtuvo que la implementación de las normas incidió positivamente en la evaluación de los datos geográficos obteniendo un diagnóstico de la confiabilidad de la información geográfica, que a su vez determina los requisitos mínimos para que la empresa o los funcionarios relacionados con estos datos puedan valerse de ellos y puedan dar el uso pertinente para un diagnóstico acertado con el propósito de que sean datos confiables y de aplicabilidad para los usuarios.

En el año 2021, Alexander Inga en la tesis titulada “Método de Gestión de la Incertidumbre de Datos Geoespaciales de Pozos Petroleros durante la Fase Exploratoria”, realizada en la Universidad Nacional Mayor de San Marcos, tuvo como objetivo crear un

método que gestione la incertidumbre de datos geoespaciales basados en el Control de Calidad de la Geodatabase, en la Información Geográfica Voluntaria y en un artefacto que controle la incertidumbre de la georreferenciación de los pozos petroleros durante su fase exploratoria, llegando a la conclusión que el Método de Gestión de Incertidumbre de Datos Geoespaciales, basados en índices de confiabilidad, valida la calidad del dato en la Geodatabase.

2.3. Marco normativo

Norma ISO 19131

Esta norma internacional tiene como objetivo detallar los requisitos necesarios para la especificación de productos de datos geográficos, tomando como base los principios establecidos en otras normas internacionales de la familia ISO 19100. Asimismo, ofrece directrices para la elaboración de estas especificaciones de manera que sean comprensibles y adecuadas para su uso previsto (Asociación Española de Normalización y Certificación [AENOR], 2022). Esta norma fue utilizada como marco de referencia para la elaboración de las especificaciones técnicas de los datos geográficos dentro de la organización.

Norma ISO 19157

Esta norma internacional establece los principios para evaluar la calidad de los datos geográficos. Define los elementos necesarios para describirla, especifica la estructura de un registro de medidas de calidad de datos, detalla los procedimientos generales para su evaluación y establece principios para su comunicación. Además, define un conjunto de medidas para evaluar y comunicar dicha calidad. (Asociación Española de Normalización y Certificación [AENOR], 2013). Esta norma se empleó como marco de referencia para la evaluación de la consistencia lógica de los datos geográficos.

2.4. Marco conceptual

Sistema de Información Geográfica

Los Sistemas de Información Geográfica (SIG) se describen como un enfoque o técnica para el procesamiento de información geográfica, con el propósito de integrar de manera efectiva datos gráficos y numéricos para generar información espacial. Este proceso se apoya en el uso de fuentes de información y una serie de recursos informáticos, incluyendo hardware y software, todo dentro del contexto de un proyecto definido por un grupo de individuos. Este proyecto está sujeto al control de los técnicos responsables de su implementación y desarrollo (Bravo, 2000).

Componentes de un SIG

Un SIG consta de cinco componentes fundamentales. A continuación, exploraremos cada uno de estos elementos (Olaya, 2020):

- **Datos:** La materia prima necesaria para el trabajo en un SIG, y los que contienen la información geográfica vital para la propia existencia de los SIG.
- **Métodos:** Conjunto de formulaciones y metodologías a aplicar sobre los datos.
- **Software:** Aplicación informática diseñada específicamente para trabajar con los datos e implemente los métodos anteriores.
- **Hardware:** Dispositivos y recursos necesarios para ejecutar el software.
- **Personas:** Encargados de diseñar y utilizar el software, siendo el motor del sistema SIG.

Dato geográfico

Se define como aquel dato que posee una referencia geográfica precisa, ya sea a través

de coordenadas o una referencia indirecta como un código postal, que permite identificar su ubicación exacta en un mapa. Estos datos proporcionan información sobre eventos relacionados con objetos que ocupan un espacio determinado, siendo este espacio definido por su tamaño, posición y forma (Ministerio de Asuntos Económicos y Transformación Digital de España, 2020).

Modelos de datos geográficos

Los modelos de datos geográficos se refieren a la estructura o representación conceptual que se utiliza para organizar y almacenar información geográfica en un Sistema de Información Geográfica (SIG). Este modelo define cómo se capturan, almacenan, relacionan y consultan los datos geográficos. Los modelos más utilizados son:

- **Modelo ráster:** En este modelo la región de interés se fragmenta sistemáticamente en unidades mínimas, generalmente denominadas celdas, y se registra la información relevante que la caracteriza (Olaya, 2020).
- **Modelo vectorial:** Este modelo captura la variabilidad y las particularidades de la región de interés a través de entidades geométricas, cada una de las cuales mantiene sus características constantes (Olaya 2020).

Calidad

En términos generales, la calidad se define como las características o conjunto de atributos naturales de un producto que permiten evaluarlo en términos de superioridad, igualdad o inferioridad en comparación con otros productos similares (Niño, 2010).

Elemento de la calidad de datos

Se define como un componente que describe un cierto aspecto de la calidad de los

datos geográficos (AENOR, 2013). En la tabla 1 se detallan los elementos y subelementos que define la calidad de los datos geográficos.

Tabla 1

Elementos y subelementos de calidad

Elemento	Subelemento
Totalidad	- Comisión - Omisión
Consistencia lógica	- Consistencia conceptual - Consistencia de dominio - Consistencia de formato - Consistencia topológica
Exactitud posicional	- Exactitud absoluta o externa - Exactitud relativa o interna - Exactitud posicional de datos en malla
Exactitud temática	- Corrección de la clasificación - Corrección de atributos no cuantitativas - Exactitud de atributos cuantitativos
Calidad temporal	- Exactitud de una medida de tiempo - Consistencia temporal - Validez temporal
Usabilidad	- Atención a los requerimientos del usuario

Nota: Adaptado de la norma ISO 19157:2013

Consistencia Lógica

La consistencia lógica se caracteriza por el grado de conformidad con las reglas lógicas que rigen la estructura de los datos, sus atributos y sus relaciones. Esta dimensión de la calidad de los datos está compuesta por cuatro elementos (AENOR, 2013):

- Consistencia conceptual: Adherencia a las reglas del modelo conceptual.
- Consistencia de dominio: Adherencia a los valores de su dominio.
- Consistencia de formato: Grado en que los datos se almacenan de acuerdo con la

estructura física del conjunto de datos.

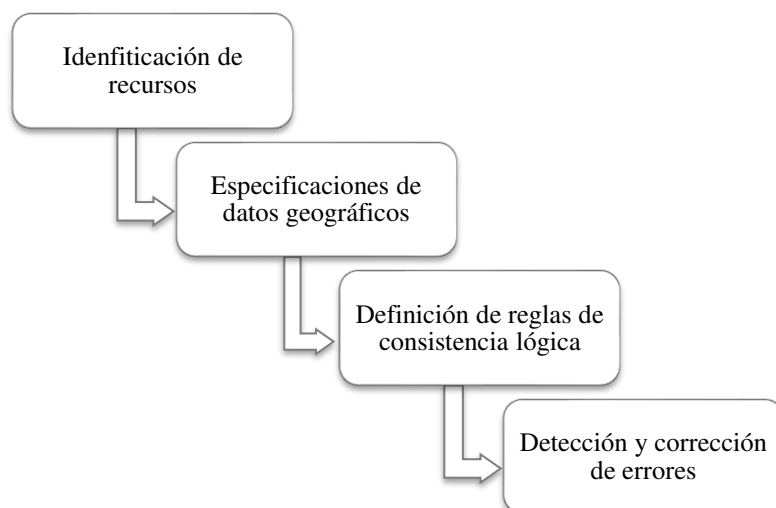
- Consistencia topológica: Corrección de las características topológicas codificadas explícitamente.

2.5. Procedimiento

Para llevar a cabo esta actividad, se implementó un procedimiento que se fundamentó en las directrices y recomendaciones establecidas en las normas ISO 19131:2022 e ISO 19157:2013. Es preciso indicar que estas normativas se adecuaron para satisfacer las necesidades específicas de la organización. Para una visualización clara y detallada, se presenta una representación gráfica en la Figura 2, donde se ilustra flujo de las etapas que componen este proceso.

Figura 2

Flujo de proceso para la detección y corrección de errores en la consistencia lógica



Nota: Adaptado de las normas ISO 19131:2022 e ISO 19157:2013

Identificación de recursos

La primera etapa de este proceso consistió en la identificación de los recursos de datos y software necesarios para la ejecución del proyecto. Los datos fueron seleccionados

considerando el nivel de importancia y la recurrencia en la elaboración de proyectos dentro de la organización. En la Tabla 2, se encuentran listados y descritos los datos geográficos que se emplearon en el desarrollo de esta actividad.

Tabla 2

Datos geográficos evaluados

N°	Objeto geográfico	Definición
1	CB01_DISTRITOS	Capa de polígonos que establece la división política y administrativa de una provincia
2	CB02_BARRIOS	Capa de polígonos que delimitan una superficie geográfica dentro de un Municipio y que contienen una o más manzanas
3	CB03_MANZANAS	Capa de polígono que delimitan un espacio geográfico que está constituido por uno o un grupo de viviendas, edificios, predios o terrenos
4	CB04_PARQUES	Capa de polígonos de manzanas destinados a uso público o recreativo
5	CB05_EJES_VIALES	Capa de líneas que representan un segmento de una vía de comunicación
6	CB07_APT	Capa de puntos compuesto por un conjunto de atributos que definen y ubican de manera única domicilios, comercios, industrias, entidades públicas, entre otros
7	PO01_POIS	Capa de puntos que identifican un lugar de características útiles o interesantes, principalmente orientado al ámbito comercial, de servicios, gobierno, cultural o religioso
8	RN01_SISMOS	Capa de polígonos que delimitan áreas geográficas en base a su nivel de riesgo sísmico
9	RN02_INUNDACION	Capa de polígonos que delimitan áreas geográficas en base a su nivel de riesgo de inundación
10	RN03_DESLIZAMIENTO	Capa de polígonos que delimita áreas geográficas en base a su nivel de riesgo de deslizamiento o huaycos

Nota. Prefijos de clases. CB: Cartografía base; PO: Punto de interés; RN: Riesgos naturales

Por otro lado, en la Tabla 3 se listan y describen los requisitos de software indispensables para el desarrollo de esta actividad.

Tabla 3*Requisitos de softwares*

N°	Nombre	Versión	Descripción
1	Anaconda	>= 22.9	Plataforma de código abierto diseñada para la gestión y distribución de paquetes y entornos de desarrollo en Python y otros lenguajes de programación
2	Python	>= 3.9	Lenguaje de programación de alto nivel y de código abierto. Cuenta con numerosas bibliotecas especializadas para la manipulación, análisis y visualización de datos geográficos
3	Numpy	>= 1.23	Biblioteca de Python especializada en el cálculo numérico y el análisis de datos para grandes volúmenes de datos
4	Pandas	>= 1.5	Biblioteca de Python especializada en la manipulación y el análisis de datos
5	Geopandas	>= 0.13	Biblioteca de Python especializada para el trabajo con datos geoespaciales
6	Shapely	>= 2.0	Biblioteca de Python para la manipulación y análisis de objetos geométricos planos
7	Contextily	>= 1.2	Biblioteca de Python diseñada para recuperar mapas de mosaicos de internet.
8	Matplotlib	>= 3.6	Biblioteca diseñada para crear visualizaciones estáticas, animadas e interactivas en Python
9	Seaborn	>= 0.12	Biblioteca de visualización de datos de Python basada en matplotlib
9	JupyterLab	>= 3.5	Entorno de desarrollo interactivo basado en web para cuadernos, códigos y datos. Permite configurar, organizar y documentar flujos de trabajo

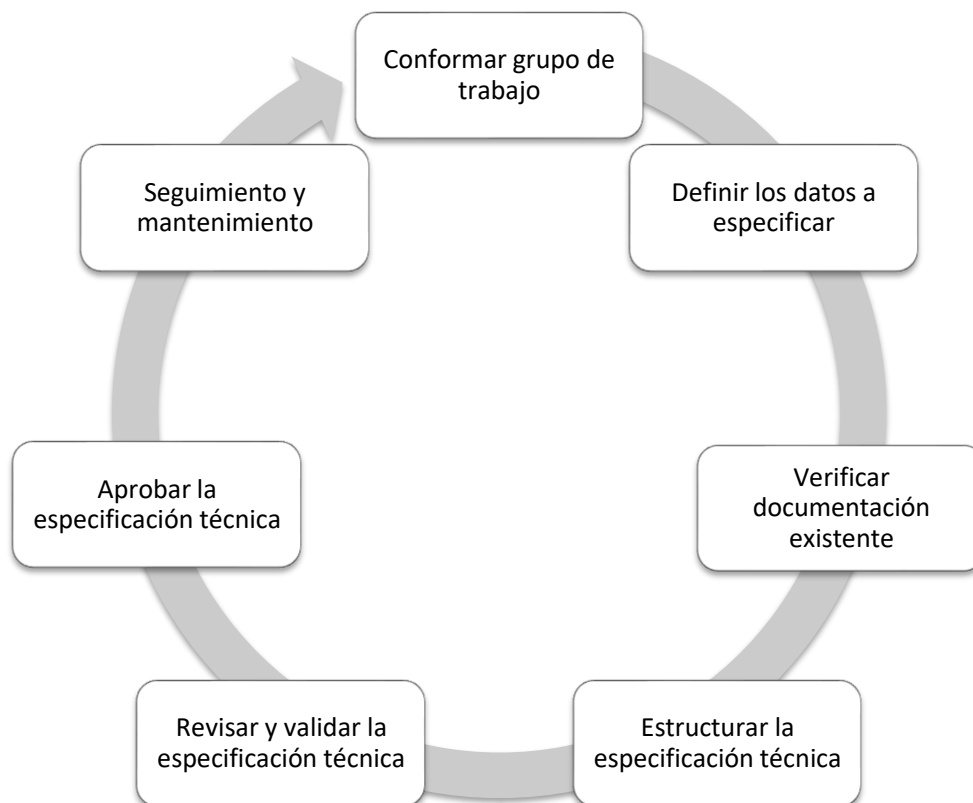
Especificaciones de datos geográficos

Durante esta fase del proceso, se elaboraron las especificaciones técnicas destinadas a la documentación de los datos geográficos, utilizando como guía las directrices establecidas en la norma ISO 19131:2022. La creación de estas especificaciones técnicas implicó seguir una

serie de pasos para garantizar que la documentación resultara concisa y de fácil comprensión. La Figura 3 ofrece una representación visual de los pasos que guiaron la elaboración de estas especificaciones.

Figura 3

Pasos para la elaboración de las especificaciones técnicas de datos geográficos



Nota. Tomado de la ISO 19131:2022

Como resultado, se generaron fichas de especificaciones técnicas para cada uno de los datos geográficos previamente identificados (ver Tabla 2), donde se definieron los atributos y relaciones considerados de importancia y que deberían estar presentes necesariamente en los datos (AENOR, 2013). A continuación, en la Tabla 4, se muestra la plantilla empleada para la elaboración de estas especificaciones.

Tabla 4*Plantilla de especificaciones técnicas para los datos geográficos*

1. DESCRIPCIÓN DE LA ESPECIFICACIÓN	
Título	Nombre que identifica la especificación
Versión	Referencia para identificar y controlar el versionamiento
Fecha de referencia	Fecha de creación o modificación de la especificación
Responsable	Dependencia que ha creado o modificado la especificación
Idioma	Idioma de la especificación
2. IDENTIFICACIÓN DEL DATO GEOGRÁFICO	
Título	Nombre del dato geográfico
Resumen	Descripción del contenido y características del dato geográfico
Categoría del tema	Tema al que se hace referencia
Alcance espacial	Extensión del área geográfica
Tipo de representación espacial	Forma de representación espacial: vectorial o ráster
Formato	Formato de archivo en el que se proporcionará el dato geográfico
Fuente	Origen del dato geográfico
3. ESTRUCTURA Y CONTENIDO (DATOS VECTORIALES)	
Geometría	Tipo de geometría
Diccionario de datos	Descripción de los atributos del dato geográfico: Nombre, definición, tipo de dato y dominio
Codificación de caracteres	Reglas y estándares para representar caracteres alfanuméricos y símbolos en los atributos del dato geográfico
4. SISTEMA DE REFERENCIA	
Identificador	Código EPSG
Descripción	Tipo de sistema de referencia y los parámetros que los definen

Nota. Adaptado de la ISO 19131:2022

Definición de reglas de consistencia lógica

Una vez definidas las especificaciones técnicas, se procedió a la fase de definición de reglas de consistencia lógica. En este proceso, se siguieron las pautas de la norma ISO 19157:2013 como referencia para identificar y desarrollar las reglas que asegurarían la

consistencia lógica de los datos geográficos.

Para llevar a cabo esta identificación de reglas, se evaluaron los cuatro elementos que componen la consistencia lógica de los datos geográficos. A continuación, en la tabla 5 se presenta un resumen de las reglas que se establecieron para garantizar la consistencia lógica de los datos geográficos de la organización.

Tabla 5

Reglas de consistencias lógicas establecidas para los datos geográficos

Código	Regla
C1	El atributo identificador no debe tener valores duplicados
C2	Los atributos deben respetar los tipos de datos y la longitud de caracteres indicados en las especificaciones técnicas (ítem diccionario de datos)
C3	Los valores del atributo UBIGEO deben tener 6 caracteres obligatoriamente
C4	Los atributos de tipo texto deben estar en mayúscula
C5	El sistema de referencia es WGS84
D1	Los atributos deben respetar los dominios indicados en las especificaciones técnicas (ítem diccionario de datos)
D2	El atributo UBIGEO solo acepta los valores que se encuentren en el objeto geográfico BC01_DISTritos
D3	Debe estar contenida en la extensión de la capa BC01_DISTritos
T1	No deben existir geometrías duplicadas (a menos que se establezca una excepción)
T2	No deben existir geometrías inválidas
T3	No deben existir geometría multipartes
T4	No deben existir anillos interiores
T5	No debe existir superposición de entidades
T6	Las capas de puntos deben interceptarse o estar contenidos dentro del objeto BC03_MANZANAS
T7	Las capas de líneas deben cortarse en las intersecciones

Después de definir las reglas generales, se procedió a crear una matriz que identifica los objetos geográficos y asigna a cada uno de ellos las reglas correspondientes. Esta matriz se encuentra detallada en la Tabla 6 para su referencia.

Tabla 6*Matriz de datos geográficos y reglas de consistencia lógica*

Dato geográfico	C1	C2	C3	C4	C5	D1	D2	D3	T1	T2	T3	T4	T5	T6	T7
BC01_DISTritos	X	X	X	X	X	X			X	X			X		
BC02_BARRIOS	X	X	X	X	X	X	X	X	X	X	X	X	X		
BC03_MANZANAS	X	X	X	X	X	X	X	X	X	X	X	X	X		
BC04_PARQUES	X	X	X	X	X	X	X	X	X	X	X	X	X		
BC05_EJES_VIALES	X	X	X	X	X	X	X	X	X	X	X				X
BC07_APT	X	X	X	X	X	X	X	X	X	X	X			X	
PO01_POIS	X	X	X	X	X	X	X	X		X	X			X	
NR01_SISMOS	X	X		X	X	X		X	X	X	X		X		
NR02_INUNDACION	X	X		X	X	X		X	X	X	X		X		
NR03_DESLIZAMIENTO	X	X		X	X	X		X	X	X	X		X		

Detección y corrección de errores

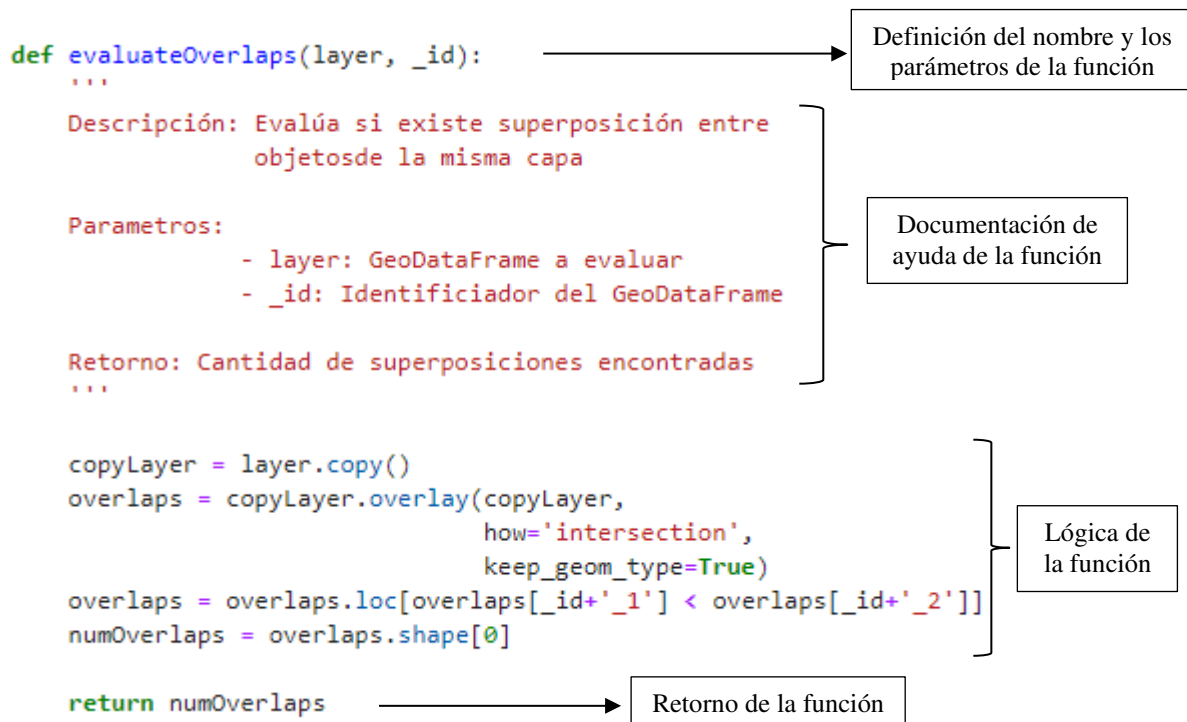
En esta fase final del proceso, se utilizaron las reglas de consistencia lógica establecidas como base para el desarrollo de los procesos automatizados destinados a la detección y corrección de estos errores. En este proceso se utilizaron técnicas de análisis espacial y algoritmos de análisis de datos con la finalidad de detectar y posteriormente corregir cualquier irregularidad o discrepancia en los datos geográficos. Para llevar a cabo la evaluación de los datos, se optó por el método de evaluación directa, que implica la inspección detallada de los elementos que conforman el conjunto de datos (AENOR, 2013).

Para lograr la automatización de estos procesos, se utilizaron los paradigmas de programación funcional y procedimental. La programación funcional se empleó para la creación de funciones diseñadas para detectar y corregir los errores en la consistencia lógica. Estas funciones fueron desarrolladas de manera modular, de modo que puedan ser aprovechadas por múltiples datos geográficos que compartan las mismas reglas de consistencia lógica e inclusive reutilizarse en nuevos conjuntos de datos. En la figura 4 se muestra la

estructura de las funciones utilizando las buenas prácticas de programación.

Figura 4

Estructura de las funciones implementadas



Por otro parte, la programación procedimental permitió ejecutar estas funciones de forma ordenada y estructurada. Para lograrlo, se utilizaron cuadernos de trabajo en JupyterLab, donde se ejecutaron los procesos de manera organizada, tal como se ejemplifica en la Figura 5. Estos cuadernos proporcionaron el entorno propicio para integrar y ejecutar las funciones de corrección de manera efectiva.

Bajo este enfoque, se han diseñado funciones personalizadas destinadas a identificar y solucionar los errores en la consistencia lógica de los datos geográficos. En el Anexo 1, se detallan las funciones diseñadas para detectar dichos errores, mientras que en el Anexo 2, se enumeran y describen las funciones diseñadas específicamente para corregir estos errores presentes en los datos geográficos.

Figura 5

Ejemplo de ejecución de procesos utilizando programación procedimental

4.5. Evaluar CRS

```
isCRS = de.evaluateCRS(layer, CRS)
isCRS
```

False

4.6. Evaluar geometrías duplicadas

```
geomsDup = de.evaluateGeomDup(layer)
geomsDup
```

0

4.7. Evaluar geometrías inválidas

```
geomsInv = de.evaluateGeomDup(layer)
geomsInv
```

0

4.8. Evaluar geometrías multipartes

```
geomMulti = de.evaluateGeomMulti(layer, GEOMTYPE)
geomMulti
```

5

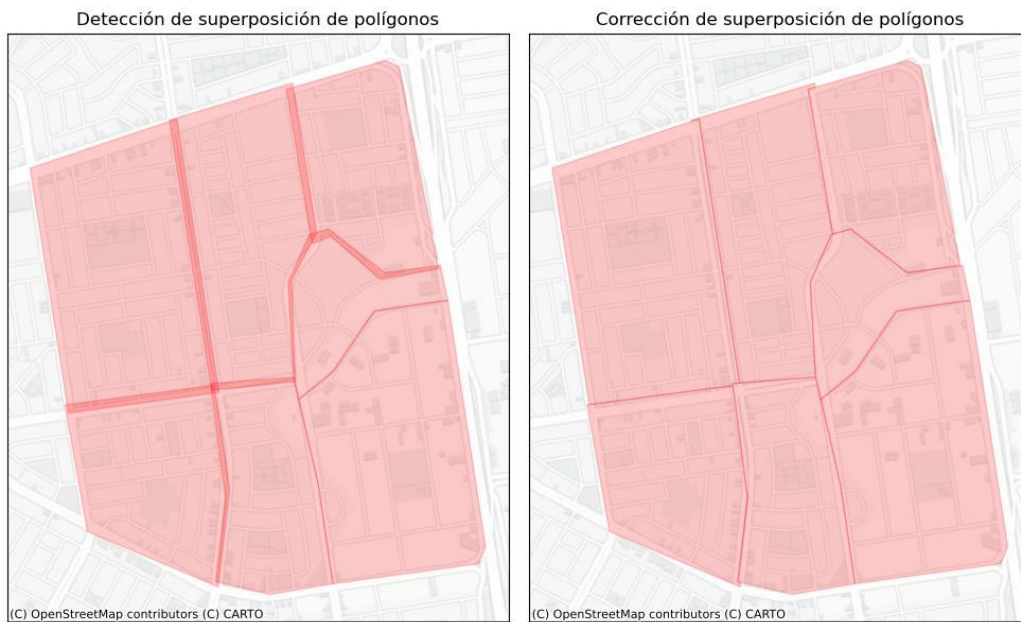
Como resultado de la detección y corrección automatizada, se generan nuevos datos geográficos que cumplen con las reglas de consistencia lógica previamente establecidas. A continuación, se presentan unos ejemplos de los errores detectados y corregidos de los datos geográficos analizados:

- Detección y corrección de superposición

La superposición de polígonos es una regla topológica que establece que dos o más elementos no deben superponerse. Esta regla se aplica porque las capas poligonales se utilizan para enriquecer datos puntuales georreferenciados, y la presencia de superposiciones puede dar lugar a duplicación de puntos.

Figura 6

Errores de superposición detectados y corregidos para una capa poligonal

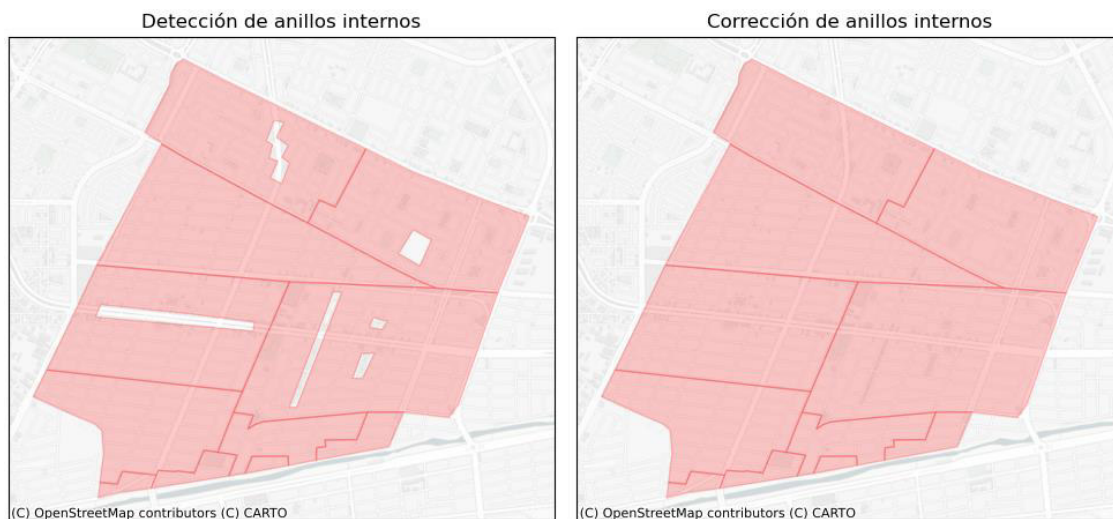


- Detección y corrección de anillos interiores

De manera similar, los anillos internos en una capa poligonal pueden generar errores al enriquecer capas puntuales, ya que, si un punto se encuentra dentro de un anillo interior, no se enriquecerá y se obtendrán valores nulos. En la Figura 7 se muestra los errores de anillos interiores detectados y corregidos

Figura 7

Errores de anillos interiores detectados y corregidos

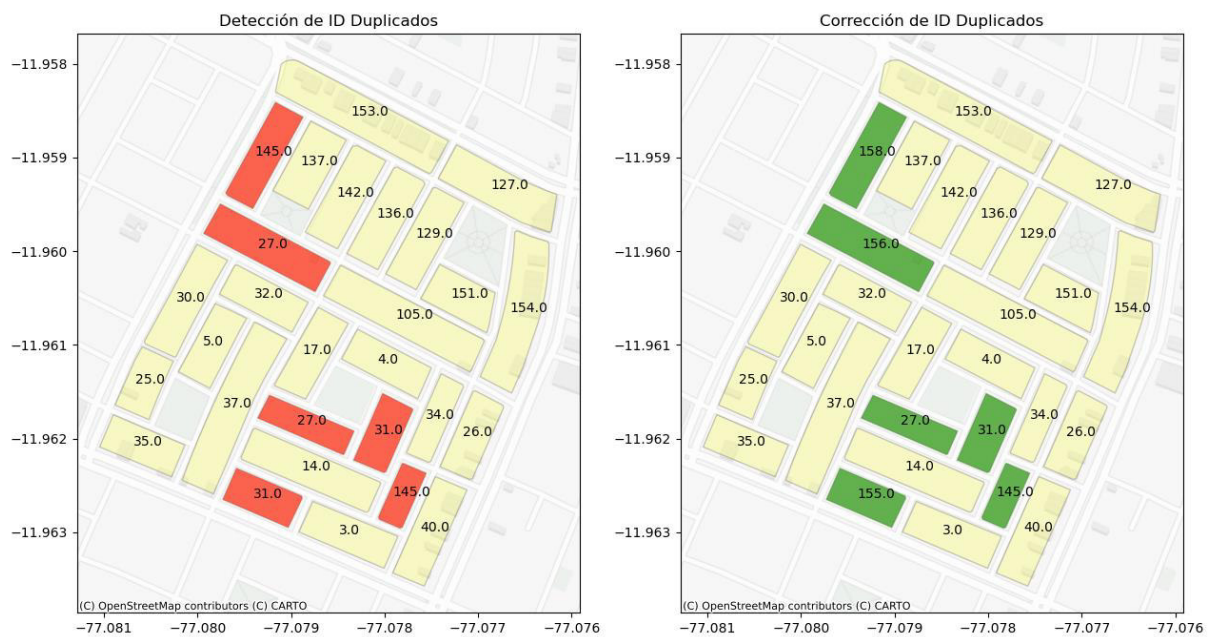


- Detección y corrección de ID Duplicados

Los errores de ID duplicados generan ambigüedades en la identificación de los registros, lo que afecta a las operaciones de búsqueda, consulta, análisis y gestión de datos. En la Figura 8, se resaltaron en color rojo los errores detectados de identificadores duplicados en el lado izquierdo, mientras que en el lado derecho se muestra la corrección de estos errores.

Figura 8

Errores de ID duplicados detectados y corregidos



III. APORTES MÁS DESTACABLES A LA EMPRESA

Entre los principales aportes a la empresa, se destacan los siguientes

- Gobierno de datos espaciales: Establecimiento de políticas, estándares y procedimientos para la recolección, almacenamiento, análisis, distribución y documentación de datos espaciales con la finalidad de garantizar la calidad, integridad, seguridad y uso adecuado de los datos espaciales dentro de la empresa.
- Integración y consolidación de datos espaciales: Recopilación, clasificación y unificación de los datos espaciales disponibles en la empresa, creando una base de datos coherente y accesible con la finalidad de facilitar su gestión y análisis.
- Optimización de los procesos de actualización y mantenimiento de datos espaciales: Diseño e implementación de herramientas de procesamiento automatizado para limpiar, filtrar y transformar los datos espaciales. Estas herramientas se utilizan para eliminar datos duplicados, corregir errores o completar información faltante.
- Automatización de procesos de carga de datos espaciales: Implementación de procesos ETL (Extracción, Transformación y Carga) para la recopilación de datos espaciales de diferentes fuentes, transformación en un formato común y carga a las diferentes bases de datos de los sistemas y aplicaciones de la empresa.
- Control de calidad de datos espaciales: Establecimiento e implementación de procesos de control de calidad para verificar la precisión y la integridad de los datos espaciales.
- Mejora del motor de Geocodificación de direcciones: Reestructuración de la base de datos del motor de Geocodificación de direcciones. De esta manera, se logró optimizar la actualización de datos, aumentar el porcentaje de efectividad de acierto y asegurar

los datos espaciales.

- Migración de datos espaciales a Google Cloud Platform (GCP): Traslado y actualización del conjunto de datos espaciales a la plataforma de almacenamiento y análisis de datos en la nube de Google.

IV. CONCLUSIONES

- La implementación de un sistema de detección y corrección automatizada de errores en la consistencia lógica ha resultado ser sumamente eficaz para elevar la calidad de los datos geográficos. Este sistema ha demostrado su capacidad para abordar y corregir de manera integral los errores de consistencia lógica presentes en los conjuntos de datos examinados, lo que contribuye de manera significativa a la fiabilidad y utilidad de la información geoespacial generada.
- La automatización de los procesos de detección y corrección de errores que afectan la consistencia lógica ha incrementado la eficiencia y productividad del equipo encargado de mantener y gestionar los datos geográficos. Esta implementación ha llevado a una significativa reducción de la carga de trabajo manual relacionada con la corrección de errores, liberando recursos y tiempo para la realización de tareas más estratégicas y de mayor valor agregado.

V. RECOMENDACIONES

- Ejecutar de forma periódica el proceso de detección y corrección automatizada para mejorar la gestión de datos geográficos, con la finalidad de asegurar que los conjuntos de datos se mantengan actualizados y precisos.
- Continuar con el proceso de mejora continua del proceso de detección y corrección automatizada, implementando nuevas reglas de consistencia lógica y/o mejorando las existentes.
- Capacitar al personal en el uso y la interpretación del proceso de detección y corrección automatizada. Esto asegurará que el equipo tenga una comprensión sólida de la calidad de los datos geográficos y fomentará la toma de decisiones basadas en la información geoespacial.

VI. REFERENCIAS

- Asociación Española de Normalización y Certificación [AENOR]. (2013). *Información geográfica. Calidad de datos* (UNE 19157:2013)
- Asociación Española de Normalización y Certificación [AENOR]. (2022). *Información geográfica. Especificaciones de producto de datos* (UNE 19131:2022)
- Bravo, J. D. (2000). *Breve introducción a la cartografía ya los sistemas de información geográfica (SIG)*. Ciemat.
- Cuervo, J. P. (2020). *Propuesta para Gestión y Control de la Información Espacial (GDB), Alfanumérica y Temática de la Cartografía Base de Isa para Incorporación en el Sistema de Información Territorial para la Construcción y Operación - SITCO*. [Trabajo de grado, Universidad Distrital Francisco José de Caldas].
<https://repository.udistrital.edu.co/handle/11349/22973>
- Equifax (2023). *Acerca de Equifax. Gobierno corporativo*. <https://www.equifax.pe/acerca-de-equifax/gobierno-corporativo/>
- Inga, A. (2021). *Método de Gestión de la Incertidumbre de Datos Geoespaciales de Pozos Petroleros durante la Fase Exploratoria*. [Tesis de maestría, Universidad Nacional Mayor de San Marcos]. <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16792>
- Méndez, J. H. (2017). *Mejoramiento de Calidad en conjuntos de Datos Abiertos basado en la aplicación de métricas de Consistencia Lógica*. [Tesis de maestría, Universidad Distrital Francisco José de Caldas].
<https://repository.udistrital.edu.co/handle/11349/8032>
- Ministerio de Asuntos Económicos y Transformación Digital a través de la Entidad Pública Empresarial Red.es. (2020). *Guía práctica para la publicación de Datos Espaciales*.
https://datos.gob.es/sites/default/files/doc/file/guia_publicacion_datos_espaciales_2.pdf

Niño, E. N. (2010). *Calidad de datos geoespaciales básicos. UD y la geomática*, (4), 110-116. <http://hdl.handle.net/11349/21202>

Olaya, V. (2020). *Sistemas de Información Geográfica*. España: Creative Commons, 123. <https://volaya.github.io/libro-sig/>

Yanes, A. V. (2019). *Evaluación de la calidad de datos geográficos de la cartografía urbana del cantón Pujilí escala 1:1000*. [Tesis de pregrado, Universidad de las Fuerzas Armadas ESPE]. <http://repositorio.espe.edu.ec/xmlui/handle/21000/21011>

VII. ANEXOS

Anexo A. Funciones implementadas para la detección de errores

Tabla 7

Funciones implementadas para la detección de errores

N°	Función	Descripción
1	<i>evaluateDupId(layer, id)</i>	Evalúa identificadores duplicados.
2	<i>evaluateStructure(layer, dicTypes)</i>	Evalúa que los datos cumplan con la estructura definida
3	<i>evaluateUbigeo(layer, ubigeo)</i>	Evalúa que el ubigeo cuente con 6 caracteres
4	<i>evaluateText(layer)</i>	Evalúa que todos los atributos de tipo texto se encuentren en mayúscula
5	<i>evaluateCRS(layer, CRS)</i>	Evalúa que sistema de referencia se encuentre en la especificación solicitada
6	<i>evaluateDomain(layer)</i>	Evalúa que los atributos cumplan con los dominios definidos
7	<i>evaluateDomUbigeo(layer, ubigeo)</i>	Evalúa que los valores del atributo de UBIGEO coincidan con los valores de BC01_DISTRITOS
8	<i>evaluateSpatExt(layer)</i>	Evalúa la extensión espacial de la capa
9	<i>evaluateGeomDup(layer)</i>	Evalúa geometrías duplicadas
10	<i>evaluateGeomInv(layer)</i>	Evalúa geometrías inválidas
11	<i>evaluateGeomMulti(layer, gType)</i>	Evalúa geometría multipartes
12	<i>evaluateGeomRings (layer)</i>	Evalúa si la geometría tiene anillos interiores
12	<i>evaluateOverlaps(layer, id)</i>	Evalúa la superposición de entidades
14	<i>evaluatePntPlg(layerPnt, layerPlg)</i>	Evalúa si una capa de puntos se intersecta con un polígono de referencia
15	<i>evaluaLineSplit(layerLine)</i>	Evalúa si una capa de línea se corta en la intersección

Anexo B. Funciones implementadas para la corrección de errores

Tabla 8

Funciones implementadas para la corrección de errores

N°	Función	Descripción
1	<i>correctDupId(layer, id, type)</i>	Corrige o elimina identificadores duplicados.
2	<i>correctStructure(layer, dicTypes)</i>	Corrige la estructura de datos según se especifique
3	<i>correctUbigeo(layer, ubigeo)</i>	Completa los ubigeo a 6 caracteres completando con 0 a la izquierda
4	<i>correctText(layer)</i>	Convierte los campos de tipo texto en mayúscula
5	<i>correctCRS(layer, CRS)</i>	Transforma la geometría a un nuevo sistema de referencia de coordenadas
6	<i>correctDomain(layer, field, dicTxt)</i>	Corrige errores de dominio según se especifique
7	<i>correctDomUbigeo(layer, distrito)</i>	Corrige los valores del atributo UBIGEO en función de la relación espacial con la capa BC01_DISTRITOS
8	<i>correctGeomDup(layer)</i>	Elimina geometrías duplicadas
9	<i>correctGeomInv(layer)</i>	Corrige geometrías inválidas
10	<i>correctGeomMulti(layer)</i>	Convierte geometrías de multiparte a monoparte
11	<i>correctGeomRings (layer)</i>	Elimina anillos interiores
12	<i>correctOverlaps(layer)</i>	Elimina la superposición de entidades
13	<i>correctPntPlg(layerPnt, layerPlg)</i>	Mueve el punto al polígono más cercano
14	<i>correctLineSplit(layer)</i>	Corta la línea en la intersección y genera un nuevo identificador