

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**  
**Facultad de Ingeniería en Electricidad y Computación**

**“PLATAFORMA DE VISUALIZACIÓN DE DATOS  
GEOESPACIALES PARA LA CIUDAD DE GUAYAQUIL QUE  
INTEGRA DATOS DE SALUD, TRÁNSITO Y CLIMA”**

**PROYECTO INTEGRADOR**

Previo a la obtención del Título de:

**Ingeniero en Computación**

Presentado por:

Leonardo Xavier Kuffó Rivero y

Leonardo Javier Larrea Díaz

**GUAYAQUIL – ECUADOR**

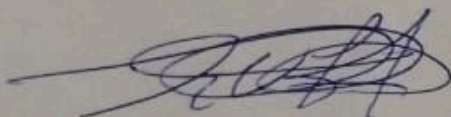
**AÑO: 2019**

## **DEDICATORIA**

El presente proyecto lo dedicamos a nuestras familias y amigos.

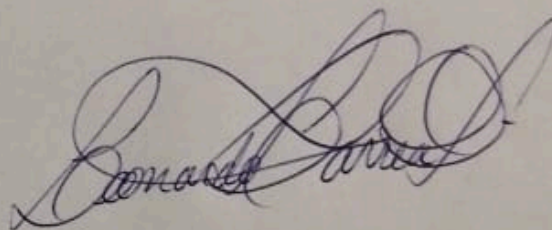
## DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; Leonardo Xavier Kuffo Rivero y Leonardo Javier Larrea Díaz damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"



---

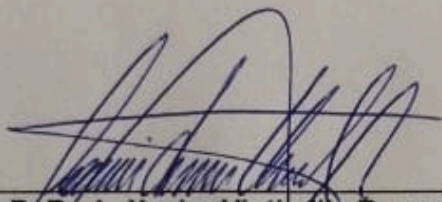
Leonardo Xavier Kuffo Rivero



---

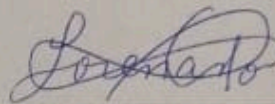
Leonardo Javier Larrea Díaz

## EVALUADORES



---

**Ph. D. Boris Xavier Vintimilla Burgos**  
PROFESOR DE LA MATERIA



---

**Ph. D. María Lorena Carló Unda**  
PROFESOR TUTOR

## RESUMEN

Actualmente, los sistemas de información nos permiten transformar datos de registros médicos de salud en conocimiento de manera rápida y eficaz. Esto puede comprender la detección temprana de brotes de enfermedades y la evolución de cuadros clínicos a través del tiempo. Uno de los usos secundarios más relevantes de los registros médicos es la generación de mapas de salud. Los mapas de salud permiten analizar la incidencia y evolución de patologías dentro de un contexto geográfico tomando en consideración una variedad de factores externos. Sin embargo, en Guayaquil, la segunda ciudad más poblada de Ecuador, un país en vías de desarrollo, no se da uso a los datos de registros médicos más allá del guardado histórico de la información. La falta de una herramienta de este tipo limita la detección de focos de enfermedades e impide el potencial análisis de la interacción de factores externos dados por la geografía de la ciudad con las incidencias de enfermedades en los pacientes. En este trabajo se propone una arquitectura de software y un esquema de almacenamiento de datos robusto, escalable y seguro para lograr el desarrollo de un mapa epidemiológico para la ciudad de Guayaquil. Esto, partiendo de registros de pacientes no geolocalizados de instituciones de salud. Para esto, se dividió Guayaquil en 177 sectores y se desarrolló un método para geolocalizar los pacientes a uno de los sectores de la ciudad. Este método de geolocalización calcula el porcentaje de similaridad entre n-gramas obtenidos a partir de las direcciones de los pacientes y los nombres de los sectores de la ciudad para geolocalizar una dirección dentro de la ciudad. Utilizando este método, se logró geolocalizar un 97.53% de 74,118 direcciones de pacientes con una precisión del 92%. Además, se demostró a través de pruebas de usabilidad conducidas con epidemiólogos, médicos e investigadores en el área de salud que el mapa desarrollado es útil y fácil de utilizar.

Palabras Clave: Mapa de salud, visualización de datos, ingeniería de software, datos geoespaciales, Guayaquil.

## ***ABSTRACT***

Nowadays, information systems can provide us with knowledge obtained from medical records in an efficient fashion. This knowledge may comprise the early detection of disease outbreaks and the evolution of diseases throughout time. One of the most relevant uses of medical records is the development of health maps. Health maps help users to analyze the incidence and evolution of a disease within a geographic context, taking into account a great variety of factors. However, in Guayaquil, the 2nd biggest city of Ecuador, a developing country, the secondary use of medical records is inexistent. The lack of health maps limits the detection of disease outbreaks and prevents the analysis of the interaction between external factors and disease incidences. This thesis proposes a software architecture and a robust, scalable and secure database schema to achieve the development of a health map for Guayaquil, starting from non-geolocated medical records taken from Guayaquil hospitals. In order to do this, Guayaquil was divided into 177 geo zones, and a geolocation methodology was developed to determine the geo zone in which a patient resides. This geolocation method computes the percentage of similarity between n-grams obtained from the patient addresses and the city sectors names to geolocate an address inside the city. Using this method, 97.53% from 74,118 addresses were geo-located with an accuracy of 92%. Furthermore, we demonstrated through usability tests conducted with medical doctors, researchers and epidemiologists that the developed map is useful and easy to use.

Keywords: Health map, data visualization, software engineering, geospatial data, Guayaquil.

# ÍNDICE GENERAL

RESUMEN.....	I
ABSTRACT .....	II
ÍNDICE GENERAL .....	III
ABREVIATURAS.....	V
SIMBOLOGÍA .....	VI
ÍNDICE DE FIGURAS .....	VII
ÍNDICE DE TABLAS.....	VIII
1 INTRODUCCIÓN.....	10
1.1 Descripción del problema.....	11
1.2 Justificación e impacto del proyecto.....	13
1.3 Objetivos .....	14
1.3.1 Objetivo general.....	14
1.3.2 Objetivos específicos.....	14
1.4 Marco teórico.....	15
1.4.1 Mapas de salud globales y regionales.....	17
1.4.2 Casos de estudio .....	18
2 METODOLOGÍA .....	20
2.1 Visión general de la solución.....	21
2.2 Consideraciones éticas .....	23
2.2.1 Anonimización de datos.....	24
2.3 Sistema de almacenamiento de datos .....	29
2.4 Arquitectura de servidor .....	33
2.4.1.1 Diccionario de términos de APIs .....	35
2.4.1.2 Definición de autenticación y cabeceras de API .....	38
2.4.1.3 Definición del API de ingreso y procesamiento de datos (Tabla 2.3)	38

2.4.1.4	Definición del API de búsqueda de datos (Tabla 2.4).....	39
2.5	Datos externos: datos meteorológicos y datos de tránsito.....	42
2.6	Diseño del visualizador .....	43
3	RESULTADOS Y ANÁLISIS .....	45
3.1	Arquitectura de servidor .....	45
3.2	Visualizador.....	47
3.2.1	Tiempo de carga de interfaz .....	53
3.2.2	Pruebas de usabilidad .....	54
3.3	Anonimización de datos .....	63
CAPÍTULO 4	.....	65
4	CONCLUSIONES Y RECOMENDACIONES.....	65
4.1	Conclusiones.....	65
4.2	Limitaciones .....	66
4.3	Recomendaciones y trabajo futuro.....	67
BIBLIOGRAFÍA	.....	69
APÉNDICES	.....	74



## ABREVIATURAS

ESPOL	Escuela Superior Politécnica del Litoral
HLB	Hospital León Becerra
MEG	Mapa Epidemiológico de Guayaquil
INAMHI	Instituto Nacional de Meteorología e Hidrología
OMS	Organización Mundial de Salud
ONU	Organización de las Naciones Unidas
SISPRO	Sistema Integral de Información de Protección Social
SISA	Sistema Integrado de Información Sanitaria Argentino
CENS	Centro Nacional en Sistemas de Información de Salud
GPS	Global Positioning System
API	Application Programming Interface
UX	User Experience
HTTPS	Hypertext Transfer Protocol (Secure)
SSL	Secure Socket Layer
NoSQL	No Structured Query Language
PSQL	Postgre Structured Query Language Database
JSON	Javascript Object Notation

## **SIMBOLOGÍA**

ms	Milisegundos
----	--------------

## ÍNDICE DE FIGURAS

Figura 2.1: Esquema de Ingeniería. ....	23
Figura 2.2. Guayaquil dividida geográficamente tomando como referencia: a) Distritos zonales [36], b) Distritos educativos [37]. c) Distritos policiales [38]. d) Parroquias urbanas [39]. e) Sectores urbanos [40]. f) Barrios urbanos (acercado) [40].....	26
Figura 2.3. Esquemas de base de datos relacional propuesto para el MEG. ....	31
Figura 2.4. Ciclo de vida de la jerarquía de cachés. ....	33
Figura 2.5. Arquitectura de Docker del MEG sobre Docker Compose.....	34
Figura 2.6. Abstracción de ingreso de datos y abstracción de búsqueda. ....	35
Figura 2.7. Ejemplo de encabezado de requerimiento al API. ....	38
Figura 3.1. Visualizador del MEG.....	48
Figura 3.2. Panel de Acciones del MEG. a) Sección de datos. b) Sección de búsqueda de enfermedades. c) Sección de zonas de interés.....	50
Figura 3.3. Filtros rápidos para aplicar sobre los pacientes en el MEG. ....	51
Figura 3.4. Ventana flotante de filtros para aplicar sobre los pacientes en el MEG..	52
Figura 3.5. Capa de tránsito vehicular activada en el MEG. ....	53

## ÍNDICE DE TABLAS

Tabla 2.1. Proceso de anonimización. ....	25
Tabla 2.2. Diccionario de términos de API. ....	35
Tabla 2.3. Endpoints de ingreso y procesamiento de datos.....	38
Tabla 2.4. Endpoints de búsqueda de datos. ....	39
Tabla 3.1. Datos almacenados en el esquema de almacenamiento del MEG.....	46
Tabla 3.2. Resultados de tiempos de respuesta del servidor de API utilizando y no utilizando el mecanismo de cacheo DataLoader en el endpoint de /incidences. La celda resaltada representa la consulta en la cual la caché del DataLoader no se encuentra lleno. ....	46
Tabla 3.3. Resultados de Apache JMETTER en /incidences. Observación: Experimentos conducidos sobre 2 clusters en un servidor 2.7 GHz Intel Core i5 con 8 GB 1867 MHz DDR3. ....	47
Tabla 3.4. Resultados de tiempos de carga del visualizador del Mapa de Salud en la versión anterior del MEG y la versión desarrollada en este proyecto. ....	54
Tabla 3.5. Información acerca de los antecedentes de los sujetos de prueba.....	56
Tabla 3.6. Resultados de las pruebas de usabilidad. Parte 1: Realización de tareas dentro del mapa por parte de los sujetos de prueba. ....	57
Tabla 3.7. Resultados de las pruebas de usabilidad. Parte 2: Entrevista de discusión acerca del mapa de salud. ....	58
Tabla 3.8. Resultados de las pruebas de usabilidad. Parte 3: Encuesta de usabilidad del mapa de salud. ....	60

Tabla 3.9. Resultados de la metodología propuesta para la geolocalización de direcciones. Observación: La capa 2 del algoritmo no pasó por etapa de validación debido a la alta fiabilidad del método [47]. .....	64
---	----

# CAPÍTULO 1

## 1 INTRODUCCIÓN

Desde comienzos de la humanidad la salud ha sido uno de los factores más importantes para la prosperidad y el avance de las sociedades. Es así, que el registro médico de enfermedades y su evolución a través del tiempo es enfoque de estudio de especialistas de salud. Gracias a registros médicos históricos hoy existe la posibilidad de ver hacia el pasado y analizar patologías pandémicas que van desde la Plaga de Atenas (Siglo V a.C, documentada en la obra de Tucídides Historia de la Guerra del Peloponeso), pasando por la Plaga de Justiniano (Siglo VI) y su reaparición como la Peste Negra (Siglo XIV) [1], hasta la enfermedad conocida como la pandemia del siglo XXI, la Diabetes Mellitus Tipo 2.

Los registros médicos también permiten analizar grupos de personas con anomalías genéticas o degenerativas. Tal es el caso de estudio de personas de una comunidad con un alto porcentaje de incidencias del Síndrome de Laron (enanismo tipo Laron) en una comunidad de Loja, Ecuador [2]. Además, con ayuda de los registros médicos se pueden determinar correlaciones entre las enfermedades y factores medio ambientales. Un ejemplo de esto se encuentra en registros como el Tercer Libro de las Epidemias de Hipócrates, en donde ya se expresaba una correlación entre el estado del aire, los cambios de estación y “la peste” [3].

Las diversas anomalías patológicas actuales, sean mortales o no, hacen necesario el registro informático de las incidencias observadas. Ya que, a través de estos registros informáticos se pueden realizar análisis y visualizaciones de estos, en conjunto con los diversos factores subyacentes a situaciones de salud en una población. Esta información, es de ayuda para las entidades médicas de investigación y gubernamentales al momento de tomar decisiones oportunas, precautelar la salud de la población e investigar patologías más a fondo.

Ecuador, país en vías de desarrollo ubicado en la región costera de América Latina, carece en sus distritos provinciales de una herramienta que realice lo antes

expuesto. Perjudicando así la detección de focos de enfermedades y brotes repentinos de patologías infecciosas, parasitarias o genéticas.

Este trabajo se enfoca en el desarrollo de una herramienta de salud digital que combina datos de registros médicos de enfermedades, datos meteorológicos y datos de tránsito, para una de las ciudades más importantes dentro de Ecuador: Guayaquil. Guayaquil, a pesar de su importancia económica y turística dentro de la geografía de Ecuador, carece de una herramienta de salud digital de este tipo.

Este desarrollo comprende una plataforma de visualización de datos geoespaciales anonimizados de pacientes que presentan patologías de cualquier tipo a través del tiempo. Además, de permitir la integración de datos de factores externos que afectan continuamente el estado epidemiológico de la ciudad: factores meteorológicos y tránsito vehicular. Siendo también una plataforma que se encuentra construida bajo una arquitectura de software robusta, moderna, escalable y accesible. La cual se encuentra plasmada en la forma de una aplicación web, haciendo uso de mapas interactivos para la visualización de la información.

En este primer capítulo se presenta la descripción del problema, en conjunto con la justificación y el impacto de este trabajo. Seguido a esto, se define el objetivo general y los objetivos específicos del proyecto. Finalmente, se introduce un estado del arte con trabajos previamente realizados enfocados en herramientas de salud digital. Incluyendo proyectos globales y regionales, así como también, artículos científicos publicados en conferencias y revistas.

## **1.1 Descripción del problema**

Guayaquil es la segunda ciudad más poblada de Ecuador con 2'644.891 habitantes [4], y se considera uno de los focos de economía y turismo más importantes del Ecuador. Guayaquil es una ciudad ubicada a 4 metros sobre el nivel del mar, sin embargo, su geografía es cambiante a lo largo de sus diferentes distritos. Existiendo sectores con una altitud considerablemente diferente, al resto de la ciudad. Así como también, con niveles de congestionamiento vehicular variantes a lo largo de los diferentes distritos de esta.

A pesar de su relevancia dentro del país, Guayaquil no posee una herramienta de salud digital en la cual se pueda visualizar su estado epidemiológico

de manera rápida. Esto limita la detección de focos de enfermedades y brotes repentinos de patologías infecciosas. Además de impedir la observación y el potencial análisis de la interacción de factores externos dados por la geografía de la ciudad con las incidencias de enfermedades en los pacientes.

Actualmente, en Ecuador, el uso secundario de los datos clínicos en las entidades de salud es limitado. El Ministerio de Salud Pública del Ecuador utiliza los datos de salud históricos recolectados por sus instituciones, para realizar reportes del estado epidemiológico del país, de forma ocasional. De la misma manera, se realiza un reporte epidemiológico cuando existen brotes de patologías recurrentes o mortales como el dengue, tuberculosis y el VIH [6, 7]. Esta subutilización de los datos de salud limita el monitoreo en tiempo de real del estado epidemiológico de la ciudad. Así como también, reduce la posible detección temprana de brotes de enfermedades y patrones de transmisión de patologías.

A partir de esta necesidad de la ciudadanía y autoridades médicas de la ciudad, nació la iniciativa de la creación de un mapa epidemiológico de la ciudad de Guayaquil. El Mapa Epidemiológico de la Ciudad de Guayaquil (MEG) [5] fue una iniciativa de la Escuela Superior Politécnica del Litoral (ESPOL) que se trabajó como un proyecto comunitario en colaboración con el Hospital León Becerra (HLB) de Guayaquil.

La primera versión del MEG era una aplicación web interactiva en el área de la salud. Esta herramienta procesaba y visualizaba, a través de mapas interactivos, registros anonimizados de los pacientes del HLB. La funcionalidad principal de esta primera versión del MEG era visualizar la incidencia de patologías, utilizando los datos geoespaciales de pacientes agregados a un nivel de sectores y parroquias urbanas.

Con la primera versión del MEG, se intentaba mitigar los problemas previamente expuestos. Sin embargo, a pesar de que era un avance en los sistemas de visualización de información de salud del país, la primera versión del MEG se encontraba limitada a solamente monitorear incidencias de enfermedades. De tal manera que, esta primera versión del MEG no explotaba todo el espectro de posibilidades que un mapa epidemiológico puede brindarle a la comunidad científica, investigadores y autoridades de salud.



Además, poseía otras limitaciones tales como: tiempos de respuesta y renderizado no óptimos, división geográfica de barrios poco granular, carencia de un sistema almacenamiento de datos robusto y un limitado algoritmo de geolocalización de direcciones. Este algoritmo de geolocalización de direcciones se encargaba de encontrar el área geográfica a la cual pertenecía una dirección dada en un formato de texto. En la primera versión del MEG, este algoritmo era capaz de solo reconocer un 71.40% de las direcciones de pacientes disponibles.

## **1.2 Justificación e impacto del proyecto**

La nueva versión del MEG desarrollada en este proyecto contribuye en la visualización y recaudación de información para futuras investigaciones y reportes, sobre incidencias epidemiológicas en la ciudad de Guayaquil. Además de facilitar la detección de patrones de transmisión, brotes y focos de enfermedades. Con la posibilidad de focalizar las observaciones y análisis en enfermedades que afectan a la niñez o a las personas de avanzada edad. Finalmente, esta nueva versión del MEG facilita la observación para generar evidencias de correlaciones entre cuadros patológicos y la geografía que rodea a estos cuadros, con la posibilidad de extender la herramienta a otras ciudades del Ecuador.

De igual manera, el MEG contribuye en la transparencia de la salud pública hacia la población. Esto quiere decir, que a través del MEG el estado epidemiológico de la ciudad es de conocimiento público, sin comprometer la privacidad y la información de los pacientes. Esta herramienta es robusta y se encuentra actualizada con la información de los registros de pacientes del HLB. Produciendo un impacto positivo al poder contribuir en un futuro en el desarrollo de investigaciones y en el desarrollo de planes de contingencia. Además, tiene el potencial de contribuir a la creación de nuevas políticas de salud pública.

Por otra parte, el MEG colabora directamente con el proyecto DataCity. DataCity es una iniciativa del Centro de Visión y Robótica (CVR) dentro de ESPOL, que busca recolectar, procesar y visualizar información de: clima, tránsito, temperatura, establecimientos educativos, deslizamientos de tierra, inundaciones y

otros factores, de diferentes ciudades de Ecuador. El MEG provee a DataCity de una capa de información de salud de la ciudad de Guayaquil.

### **1.3 Objetivos**

#### **1.3.1 Objetivo general**

Desarrollar una nueva versión del MEG, compuesta por visualizaciones de mapas agregados a nivel de sectores urbanos, que muestren la incidencia de diferentes enfermedades a lo largo de la ciudad de Guayaquil. Información que es obtenida a partir del procesamiento de registros anonimizados de pacientes de instituciones de salud dentro de la ciudad.

Al mismo tiempo, integrando en las visualizaciones información de factores externos relacionados con la salud: tránsito vehicular y factores meteorológicos [8, 9]. Con el fin de proveer una plataforma tecnológica abierta y útil para el estudio y detección temprana de enfermedades, con posibilidad de contribuir al desarrollo de planes de contingencia y políticas de salud pública.

#### **1.3.2 Objetivos específicos**

##### **Objetivos específicos de usabilidad:**

- Crear una aplicación web interactiva y dinámica que integre datos de salud de la ciudad de Guayaquil con información meteorológica y de tránsito. De tal manera que, se pueda observar la forma en que interactúan este trío de factores a lo largo del tiempo. Con la posibilidad de utilizar filtros adicionales para la visualización de información basados en las características de los pacientes, tales como: el rango de edad de los pacientes y el género de los pacientes.
- Diseñar una herramienta útil desde un punto de vista médico y epidemiológico, para lograr alcanzar el objetivo principal de manera satisfactoria.

##### **Objetivos específicos de ingeniería:**

- Recolectar datos de salud de la ciudad de Guayaquil directo de la

fuentes (instituciones de salud).

- Geolocalizar al menos un 90% de los registros de salud de pacientes, con un sector urbano dentro de la ciudad de Guayaquil.
- Visualizar datos geoespaciales de salud, tráfico y datos meteorológicos a través del tiempo, en un mapa cuyo manejo sea fluido e interactivo.
- Desarrollar un API (Application Programming Interface) bajo una arquitectura cliente-servidor que cumpla con las siguientes características: a) Disponibilidad el 99% del tiempo. b) Tiempos de respuesta menores a 5 segundos en el 95% de los requerimientos.
- Diseñar un esquema de base de datos que cumpla con las siguientes características: a) Garantizar la integridad de los datos, por consultas masivas b) Almacenamiento de datos indexados, con un tiempo de extracción promedio menor a 1 segundo c) Soportar hasta 20 consultas concurrentes sin sacrificar el tiempo de respuesta.

#### **1.4 Marco teórico**

Existen varios estudios en la literatura que reportan como la visualización y el análisis de datos epidemiológicos geoespaciales han ayudado a realizar mejoras de salud en diferentes poblaciones. Estas mejoras de salud incluyen aspectos tales como: identificación de zonas de riesgo y brotes de enfermedades [10, 11], mejoras de planes de contingencia y prevención de cuadros clínicos de distintas enfermedades [12, 13] y contribuciones a las estadísticas de salud de una población [14, 15, 16].

Dentro de esta miscelánea colección de trabajos es recurrente el énfasis en la importancia y los beneficios de la exploración visual y el análisis de datos geoespaciales de salud. Así como también, la importancia de una arquitectura robusta por detrás de los componentes visuales [10, 14] y las pruebas de usabilidad de los sistemas [13, 17].

En 2009, Cinnamon et al. [13] realizó un estudio de usabilidad de mapas que muestran diferentes factores de salud. Las pruebas de usabilidad fueron orientadas a usuarios ocupando cargos concernientes a salud pública. Se consideraron tres

principales tipos de mapas: Estáticos, interactivos y animados, definidos de la siguiente manera:

a) Estáticos: Mapas crudos sin ningún tipo de interacción ni cambios a través del tiempo. Muestran la misma información de la misma manera siempre.

b) Interactivos: Mapas estáticos que permiten la interacción del usuario para realizar acciones que afectan la forma en que se visualizan y la información que se muestra.

c) Animados: Mapas cuya información cambia dinámicamente, permitiendo así la integración de una o más variables, generalmente variables temporales.

A pesar de que se encontró que los tres tipos de mapas resultan “útiles” desde la perspectiva de los usuarios, los mapas interactivos son los preferidos por estos. Esto último se debe al control que permiten sobre las visualizaciones a lo largo de las N variables que muestra la visualización. Por otro lado, los mapas animados funcionan mejor cuando los usuarios desean tener control sobre las variables que regulan las transiciones visuales.

No obstante, Slocum et al. (2004) [18] demostraron que diferentes tipos de mapas pueden cumplir distintos objetivos. Los mapas animados son ideales para identificar tendencias a través del tiempo, mientras que los mapas estáticos son la opción ideal para comparar ventanas de tiempo específicas. Por esta razón, y por las necesidades de las partes-interesadas (stakeholders), este trabajo se enfoca principalmente en el desarrollo de mapas interactivos.

La importancia del tipo de visualizaciones a mostrar es igual de relevante que el esquema de desarrollo utilizado para lograr un proyecto exitoso que satisfaga las necesidades de las partes interesadas y de los usuarios finales.

Por un lado, se tiene a las partes-interesadas, que en el contexto de este proyecto corresponden a los diferentes doctores, investigadores y epidemiólogos de instituciones de salud. Por otro lado, se tiene a los usuarios finales, que corresponde a cualquier persona natural que desee saber sobre el estado epidemiológico de la ciudad de manera rápida y fácil.

Por encima de esto, se tienen los objetivos de usabilidad e ingeniería detallados en la subsección 1.3.2 que deben ser superados. Kushniruk et. al (2004) [17] presentó un estudio detallado sobre interacción humano-computador en el uso

de sistemas computacionales orientados a la salud, que Cinamon et. al (2009) [13] adaptó hacia un esquema más simplificado de 5 etapas.

- 1) Obtención de requerimientos
- 2) Investigación de campo
- 3) Diseño de ingeniería y desarrollo
- 4) Pruebas de usabilidad con las partes-interesadas: Proceso iterativo de pruebas en el cual se puede regresar a las etapas 2 o 3 dependiendo de la retroalimentación.
- 5) Publicación, mantenimiento y reiteraciones (diseminación)

A continuación, se mencionan sistemas de información similares que presentan mapas de salud a nivel global y a nivel regional (Sudamérica). Finalmente se exponen casos de estudios exitosos en los cuales la generación de mapas de salud ha contribuido a diversas comunidades en diferentes aspectos como los mencionados al inicio de la sección 1.4.

#### **1.4.1 Mapas de salud globales y regionales**

La Organización Mundial de Salud (OMS) [20], fundada en 1948, es una entidad perteneciente a la Organización de las Naciones Unidas (ONU) encargada de realizar un monitoreo de salud de manera global, recolectando información relevante a temas de salud pública. Lamentablemente, la información relevante a muchos países en vías de desarrollo de América Latina como Ecuador, se encuentra desactualizada, incompleta, con poca granularidad y problemas de calidad [21].

Sin embargo, entre uno de los casos de éxito de mapas de salud globales se puede destacar HealthMap [22]. HealthMap es una iniciativa desarrollada por un equipo de investigadores epidemiológicos del Boston Children's Hospital en 2006. HealthMap es una plataforma web que integra varias fuentes de datos y permite la visualización interactiva y animada de brotes de enfermedades alrededor del mundo. En marzo del 2014, la plataforma encontró brotes de fiebre hemorrágica en Guinea 9 días antes de la identificación oficial de la epidemia de ébola en África [11].

Dentro de un contexto regional (Sudamérica) se pueden encontrar sistemas como SISPRO [23] (Sistema Integral de Información de Protección Social) desarrollado en el 2013 con información relacionada a diferentes variables de salud dentro del territorio de Colombia. Por otro lado, en Argentina existe el SISA [24]

(Sistema Integrado de Información Sanitaria Argentino), fundado en 2014, que se encarga de recolectar información acerca de establecimientos públicos de salud de Argentina.

Actualmente en Chile se encuentra en desarrollo la iniciativa CENS [25] (Centro Nacional en Sistemas de Información de Salud), impulsada por 5 universidades chilenas desde el 2016. Se prevé que sea una plataforma que recaude la información de diferentes establecimientos públicos y privados de salud de Chile para la visualización pública. Dentro de Ecuador el Ministerio de Salud pública posee una herramienta similar a SISA [24], denominada GeoSalud [26]. En esta herramienta se presenta la ubicación de los establecimientos de salud pública del país, teniendo la posibilidad de realizar búsquedas de estas instituciones por diferentes filtros.

#### **1.4.2 Casos de estudio**

Noon et al. (2001) [12] utilizó visualizaciones y datos de registros médicos de nacimientos para poder determinar cuales recintos poblacionales necesitaban de un aumento en el número de instituciones de atención para nacimientos. Además de, determinar cuáles poblaciones necesitaban una mejora en sus instalaciones de cuidados médicos.

Por otro lado, Yang et al. (2012) [14], presenta un trabajo enfocado en la parte ingenieril de un sistema de monitoreo en tiempo real de la transmisión del cuadro clínico de esquistosomiasis en la provincia de Jiangsu (China). Yang et al. utilizaron mapas provistos por Google Earth y dentro de la herramienta se realizan análisis de predicciones para determinar el nivel de riesgo de un recinto en diferentes épocas del año. Este nivel de riesgo se determina basado en parámetros meteorológicos, distribución de los caracoles (principales transmisores de la enfermedad) y medidas de control implementadas en los recintos. Técnicas similares de ingeniería fueron utilizadas por Parmento et al. (2008) [15] para poder obtener estadísticas de salud en un pueblo de Indonesia y visualizarlas en una herramienta a través de mapas y gráficos de barras.

Las enfermedades más devastadoras llaman la atención de los investigadores. No es sorpresa, ya que estas son las que más afectan a la población en diferentes partes del mundo. Highfield et al. (2011) [19] tenía como objetivo de su

investigación proveer a la comunidad científica, residentes y organizaciones, acceso público a datos de salud de las comunidades de Houston, USA. En Julio del 2011 lanzaron una herramienta con mapas interactivos de salud, que realiza rápidas consultas bajo demanda. El sistema informático permite filtrar las consultas por edad, etnia, género y patología. Este trabajo se apega en gran manera al objetivo principal de este proyecto.

Por otro lado, Bell et al (2006) [16] realiza un estudio de mapas de salud con estadísticas de cáncer. Sin embargo, el objetivo de este trabajo no fue desarrollar, sino analizar de manera objetiva mapas de salud públicos y determinar los factores que influyen en la mala interpretación de estos. Entre los análisis realizados, destacan las siguientes características que deberían estar presentes en las visualizaciones de mapas de salud: a) Dejar claras las limitaciones de las visualizaciones y las limitaciones en los datos. b) Siempre mantener la confidencialidad de la información. c) Definir y comunicar el nivel de incertidumbre de los resultados. d) Realizar pruebas de usabilidad.

# CAPÍTULO 2

## 2 METODOLOGÍA

El MEG es una herramienta con el potencial de cumplir un rol importante en la toma de decisiones oportunas y en el cuidado de la salud de la población de Guayaquil. Además, puede llegar a ser de gran utilidad para las entidades médicas de investigación y gubernamentales, estudiantes, y la ciudadanía. Por lo tanto, es de vital importancia la optimización de la parte ingenieril detrás del procesamiento, almacenamiento y visualización de los datos.

El diseño de ingeniería propuesto hace énfasis en una plataforma escalable que en un futuro pueda extenderse a la visualización de datos epidemiológicos de diversas ciudades. Así como también, diversas instituciones de salud, subdepartamentos de salud y grupos de geo zonas geográficas de interés. Además de, considerar en todo momento como punto crítico la confidencialidad de los registros médicos almacenados.

Para esto, en este capítulo se propone una metodología de desarrollo que comprende desde la extracción, anonimización y almacenamiento de datos, hasta el diseño y desarrollo de la plataforma de visualización de datos geospaciales para la ciudad de Guayaquil.

En la primera sección de este capítulo, se introduce una visión general de la solución y el esquema a seguir para el desarrollo del nuevo MEG. A continuación, se exponen las consideraciones éticas que conlleva manejar datos médicos. Estas consideraciones derivan en el desarrollo de una metodología para la correcta anonimización y geolocalización de los registros médicos.

En la sección 2.3 se detalla el esquema de almacenamiento y recuperación de datos para realizar las consultas de información de manera rápida. De la mano con el sistema de persistencia de datos, se detallan las tecnologías utilizadas en la arquitectura del servidor web y la definición de un API público para que plataformas de terceros puedan integrarse a los datos de la plataforma.

Luego de esto, se detalla brevemente la utilización de fuentes externas de datos meteorológicos y de tránsito. Finalmente, se describe el diseño y mecanismo



de validación de la interfaz gráfica de la plataforma de visualización de datos geoespaciales para la ciudad de Guayaquil.

Cabe destacar que se utilizó la metodología SCRUM para el desarrollo de este proyecto. Gracias a esto, se pudo ejecutar un desarrollo ágil y rápido a lo largo de la implementación.

## 2.1 Visión general de la solución

El desarrollo del MEG utiliza el esquema de 5 etapas para el desarrollo de mapas de salud adaptado por Cinamon et. al (2009) [13], definido de la siguiente manera: (1) Obtención de requerimientos, (2) investigación de campo, (3) diseño de ingeniería y desarrollo (4) pruebas de usabilidad con las partes-interesadas y, (5) publicación, mantenimiento y reiteraciones (diseminación). Este trabajo hace énfasis en los pasos 3 y 4. Esto es debido a que las dos primeras etapas ya han sido completadas satisfactoriamente por la primera versión del MEG.

El diseño de ingeniería y desarrollo del MEG puede dividirse en 4 grandes desafíos tecnológicos (componentes) definidos de la siguiente manera:

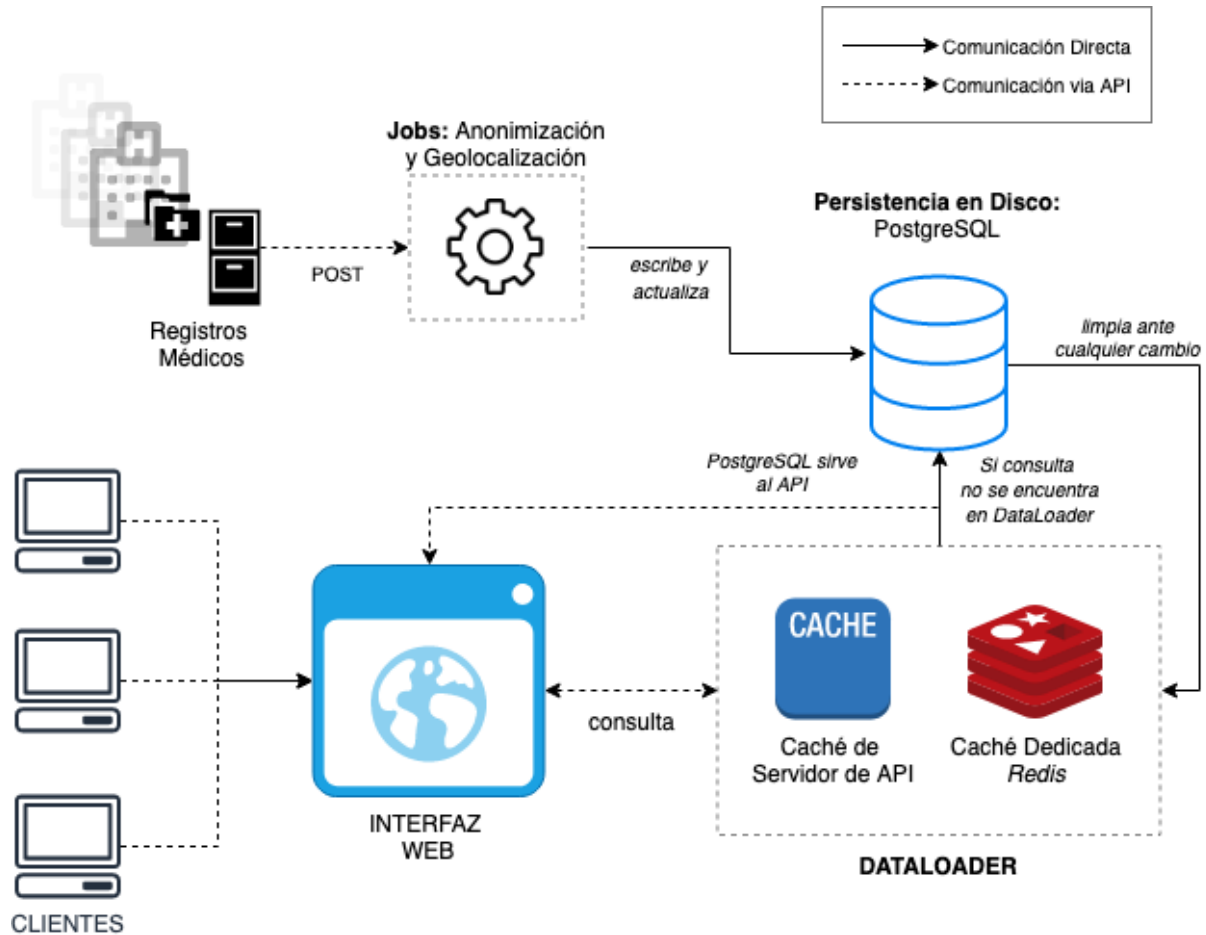
1. **Recolección de datos:** Comprende la recolección, procesamiento y transformación de los registros médicos recolectados de las instituciones de salud. En este componente, se toma en consideración las implicaciones éticas que conlleva trabajar con información sensible, como lo son los registros médicos de pacientes (Sección 2.2).
2. **Persistencia de datos:** Implica la ingeniería detrás del almacenamiento de los datos necesarios para la generación del mapa epidemiológico. Este componente es clave para la escalabilidad y rápida consulta de los datos (Sección 2.3).
3. **Servidor:** Comprende la arquitectura y tecnología utilizada para la intercomunicación de los componentes a través de un API. Además, este componente maneja la recuperación de información e interfaces gráficas a través de solicitudes cliente-servidor (Sección 2.4).
4. **Visualizador:** Es el componente gráfico que muestra la información epidemiológica de la ciudad al usuario final. Engloba el desarrollo de

una buena experiencia de usuario (User Experience) y componentes gráficos responsivos y reutilizables (Sección 2.6).

El correcto funcionamiento de los cuatro componentes en conjunto es vital para poder ofrecer buenas experiencias de uso a los usuarios finales. La Figura 2.1 detalla cómo interactúan cada uno de estos componentes, y qué rol cumplen dentro del sistema.

Primero, la información de las instituciones de salud es recolectada y enviada a través de un requerimiento Hypertext Transfer Protocol Secure (HTTPS) vía un API hacia una arquitectura de procesos (Jobs). Esta arquitectura anonimiza y geolocaliza la información de los pacientes y las persiste en un esquema de base de datos relacional dentro del sistema.

En el momento que los clientes ingresan al dominio donde se encuentra el MEG, se realiza un requerimiento cliente-servidor, a través del cual el API sirve a los clientes con las interfaces gráficas y la información del MEG. Durante este requerimiento de información, el servidor realiza las búsquedas de la información dentro de una caché en memoria para servirlos de manera rápida. En caso de no encontrarla, el servidor utiliza la base de datos para servir la información y actualiza la caché para optimizar las futuras consultas de otros usuarios. Para mantener la consistencia de la información, la memoria caché se limpia ante cualquier cambio en la base de datos. Este esquema de cacheo de información descrito es denominado DataLoader. En las siguientes secciones se hablará en detalle sobre cada uno de los procesos mencionados anteriormente.



**Figura 2.1: Esquema de Ingeniería.**

## 2.2 Consideraciones éticas

Un mapa epidemiológico permite la visualización de datos de salud, los cuales están relacionados a individuos dentro de una población. Estos datos son de carácter sensible, más aún cuando los individuos pueden estar padeciendo cuadros clínicos de carácter crónico, genético o degenerativo. Por estos motivos, es importante asegurar la confidencialidad de la información a través de la anonimización de los datos. El objetivo del proceso de anonimización es que ningún usuario final pueda saber la procedencia de un registro médico de manera individual. En la siguiente subsección se detalla la metodología para realizar el proceso de anonimización de los datos recolectados.

### **2.2.1 Anonimización de datos**

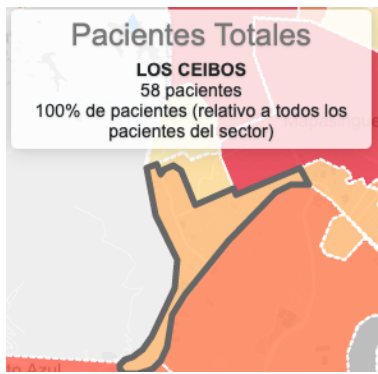
El objetivo de la anonimización de los registros médicos es salvaguardar la información privada de los pacientes. De tal manera que ningún usuario final pueda inferir la procedencia de un registro médico individual a través de la visualización del mapa epidemiológico.

Bell et. al. [16] describen cuatro diferentes métodos para mantener la confidencialidad de datos de salud que deseen ser mostrados en mapas geográficos. Entre los cuales se encuentran (1) la integración de variables etiológicas a los registros, removiendo cualquier localizador geográfico (2) la agregación temporal y espacial, (3) la limitación del acceso de los usuarios a la información a través entornos controlados (p. ej. el uso de credenciales para poder acceder a una visualización), y (4) la redistribución aleatoria de los localizadores geográficos de los pacientes (p. ej. cambiar intencionalmente la ubicación geográfica del paciente de manera aleatoria hasta 20 metros a la redonda del punto original).

Para esta nueva versión del MEG se utilizó una combinación de los primeros tres métodos: (1) Utilización de variables contextuales de carácter etiológico y geográfico dentro de cada registro médico, (2) agregación espaciotemporal y (3) limitación de acceso de los usuarios a la información.

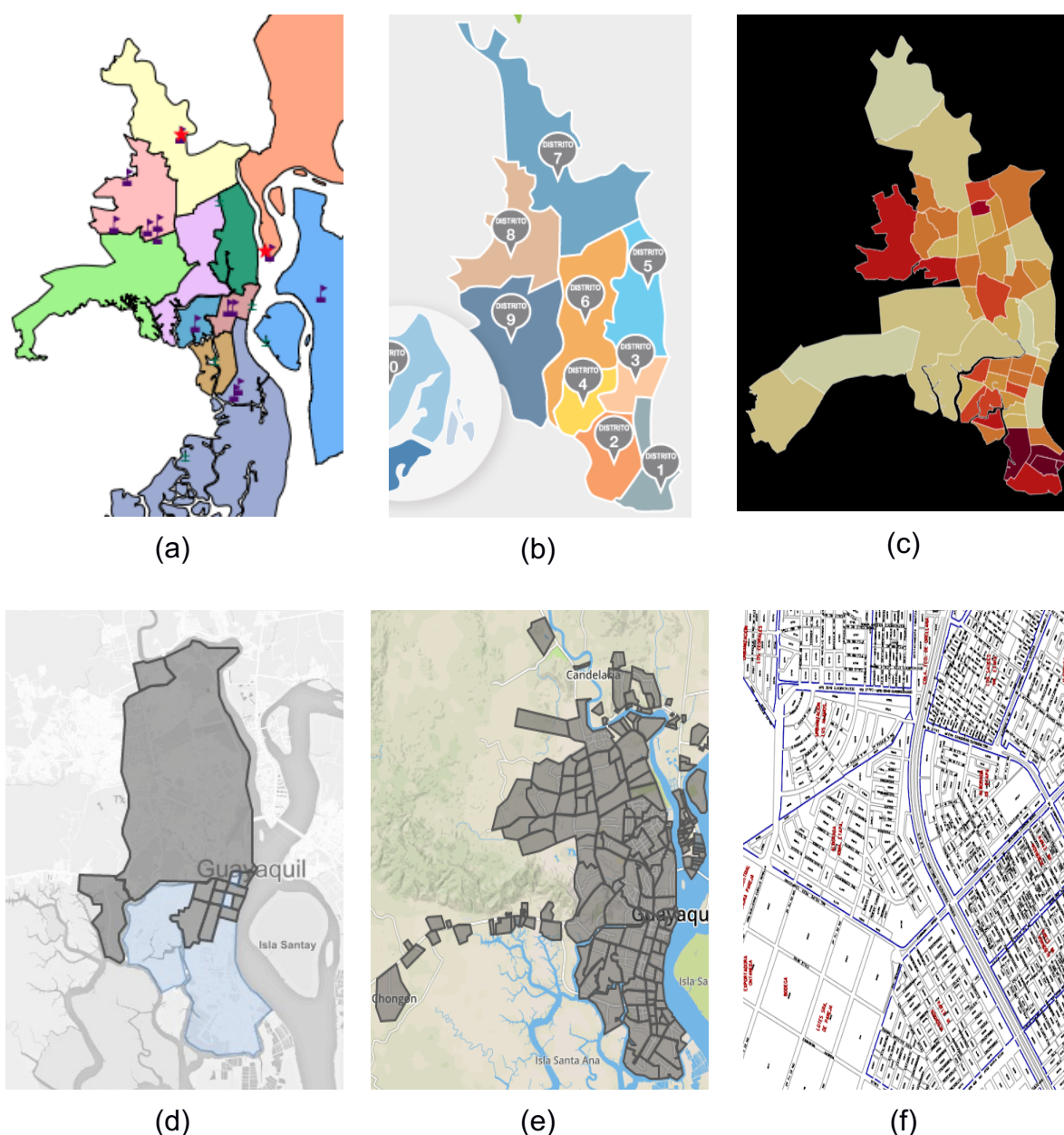
La Tabla 2.1 detalla el proceso de anonimización de datos haciendo uso de los tres métodos anteriormente expuestos. Como variables contextuales se utiliza el sector geográfico al cual pertenece el paciente, su género, y la discretización de su edad. Además, la exposición de esta información al usuario final es agregada temporal y espacialmente.

**Tabla 2.1. Proceso de anonimización.**

	Registro Médico →	Registro Anonimizado →	Registro Agregado
<b>D E S C R I P C I Ó N</b>	Registro médico crudo obtenido de las instituciones de salud.	Utilización de variables contextuales: Sector urbano, género, rango de edad para anonimizar un registro. Esta información será almacenada en la base de datos.	Agregación espaciotemporal utilizada en el visualizador. Esta representación de la información será visible públicamente solamente si existe más de 1 registros dentro de la agregación espaciotemporal. [16]
<b>U S O</b>	<b>Nombre:</b> Leonardo Kuffo <b>Edad:</b> 23 <b>CIE10 (patología):</b> Z99 <b>Fecha de Ingreso:</b> 10/10/15 <b>Dirección de Domicilio:</b> Los Ceibos Etapa Y Manzana Z #0	<b>Género:</b> Masculino <b>Edad:</b> 18 - 25 <b>CIE10:</b> Z99 <b>Fecha de Ingreso:</b> 10/10/15 <b>Sector:</b> Los Ceibos	

Dentro del proceso de anonimización de la información, se hace uso de una variable contextual de carácter geográfico para anonimizar los registros médicos. Es importante dejar clara la definición de esta variable dentro del contexto de este proyecto (es decir, Guayaquil).

Guayaquil, posee diversos esquemas oficiales de división geográfica en los cuales se puede segmentar la ciudad (Figura 2.2). La visualización de incidencias de enfermedades a un nivel de sectores urbanos (Figura 2.2.e) es ideal para las autoridades de salud, ya que esta división provee un óptimo nivel de granularidad geográfica. Además de facilitar la lectura y búsquedas en el mapa, debido a que estas divisiones son de conocimiento general para la población. Es por esto por lo que se utilizaron los sectores urbanos como unidad geográfica principal de análisis.



**Figura 2.2. Guayaquil dividida geográficamente tomando como referencia: a) Distritos zonales [36], b) Distritos educativos [37]. c) Distritos policiales [38]. d) Parroquias urbanas [39]. e) Sectores urbanos [40]. f) Barrios urbanos (acercado) [40].**

Una vez definido el espectro de la variable contextual geográfica para la anonimización, se realizó la geolocalización de direcciones de los pacientes. Lograr una correcta geolocalización fue uno de los desafíos más relevantes del proyecto.

En el contexto de este proyecto, geolocalizar es poder obtener la ubicación geográfica de un paciente dada su dirección. Esto no fue una tarea sencilla debido a diversos factores. Dentro del más relevante se puede destacar la carencia de un sistema de normalización de direcciones dentro de Guayaquil. Esto quiere decir, que

una dirección puede ser escrita de más de una forma sin perder su validez. Así como también, un mismo lugar, puede ser referenciado por dos nombres totalmente diferentes.

Finalmente, también existen problemas en la calidad de las direcciones ingresadas manualmente en los sistemas de información. Por ejemplo, hay direcciones que no existen o se encuentran escritas de manera errónea.

La correcta geolocalización de direcciones también es un problema recurrente dentro de los modelos de negocios de entregas a domicilio por medio de crowdshipping [27]. El crowdshipping es un modelo de logística de envío de paquetes basado en economía compartida. En este modelo los repartidores generan ganancias realizando entregas bajo-demanda requeridas por los clientes.

Las empresas de crowdshipping muchas veces solucionan el problema de geolocalización dándole la posibilidad a sus clientes de poder geolocalizar la ubicación de la entrega dentro de un mapa digital. Sin embargo, las instituciones de salud en Guayaquil no proveen estas facilidades. Dentro de las instituciones de salud las direcciones de los pacientes son ingresadas de forma manual, sin tener un punto geográfico de referencia.

Considerando lo anterior, se desarrolló para la primera versión del MEG una metodología de geolocalización de direcciones de pacientes partiendo de direcciones en formato de texto plano. El propósito de esta metodología fue inferir el sector urbano de Guayaquil al cual pertenece una dirección de un paciente.

La metodología de geolocalización de esta nueva versión del MEG consiste en una tokenización de las direcciones en n-gramas de una, dos y tres palabras (es decir, uni-grams, bi-grams y tri-grams). Cada token es comparado uno a uno contra un levantamiento manual en formato de texto del nombre de todos los sectores urbanos de la ciudad. En caso de encontrar una coincidencia exacta entre un token y un sector en el levantamiento manual, se utiliza este sector como variable contextual geográfica para la anonimización. Utilizando esta estrategia, un 71.40% de las direcciones disponibles pueden ser geolocalizadas a un sector. Mientras que el resto de los registros deben ser descartados, debido a que no pueden ser anonimizados.

En esta nueva versión del MEG, se propuso un esquema de dos capas. Cada capa intenta realizar la geolocalización de la dirección de manera consecutiva, en caso de fallar la anterior.

La primera capa consiste en una mejora a la metodología de tokenización de direcciones implementada en la versión anterior del MEG. Al igual que la metodología anterior, se tokeniza la dirección original en uni-grams, bi-grams y tri-grams. Sin embargo, cada token pasa por un mecanismo de filtrado. Este filtrado consiste en eliminar tokens que correspondan a palabras “neutras” dentro de una dirección. Es decir, palabras que no aporten al contexto. Tales como: entre, los, callejón, cooperativa, manzana, villa, parroquia, etc.

Luego, los tokens restantes son comparados con los nombres de todos los sectores de la ciudad. Esta comparación se realiza mediante el cálculo de una métrica de distancia de texto que represente cuán similares son el token y el nombre de un sector de la ciudad. De tal manera que, no se descarten los registros para los que una coincidencia exacta entre el token y un sector no pudo ser encontrada.

La métrica de distancia que se utilizó es la Distancia de Levenshtein. La Distancia de Levenshtein es una métrica que puede ser utilizada para cuantificar la similitud entre dos cadenas de texto [41]. Esta distancia se basa en el número de ediciones necesarias para convertir un *texto A* en un *texto B*. La Distancia de Levenshtein ha sido utilizada previamente en la detección de plagio [42] y la cuantificación de diferencias de dialectos [43]. Para su correcta implementación en el contexto de este trabajo, se definió un límite (threshold) de similitud que deben tener el token y el sector para ser considerados una coincidencia. La precisión de aciertos de esta capa fué validada de forma manual con una muestra aleatoria de 100 direcciones.

En caso de que esta capa no pueda geolocalizar la dirección, entra en juego la segunda capa: Google Maps Geocoding API. Se utilizó Google Maps Geocoding API [28] en una modalidad “no estricta” con un enfoque de búsqueda “bottom-up search”. Esto quiere decir que, el Geocoding API intentará encontrar un punto aproximado de correspondencia dada una dirección, en caso de fallar al encontrar el punto exacto.

Google Maps Geocoding API es utilizado como segunda capa, y no como



primera debido a que, en una ciudad dentro de un país en vías de desarrollo como Guayaquil, los APIs convencionales de geolocalización de direcciones no poseen una alta precisión. Debido a que, suelen generalizar en gran manera los resultados de las búsquedas al intentar relacionar una entidad geográfica a una dirección en formato de texto. Esto sucede debido a la falta de normalización de direcciones que existe dentro de estas ciudades.

Por otro lado, la anonimización de un registro también fue lograda utilizando el género y la edad del paciente como variables contextuales adicionales.

Cabe recalcar que este proceso de anonimización y geolocalización, se realiza en una arquitectura basada en Jobs. Dentro de esta arquitectura, la información original es recibida a través de Socket Secure Layer (SSL) dentro de un requerimiento POST de tipo HTTPS. La información procesada anonimizada es almacenada en la base de datos. Mientras que la información original es desechada, y no persiste en ningún tipo de almacenamiento. La ventaja de realizar este procesamiento en una arquitectura basada en Jobs consiste en la separación de responsabilidades y el control sobre los procesos que se ejecuten. Además de, no depender de las limitaciones del protocolo HTTPS, tales como un tiempo de respuesta limitado.

Finalmente, la extracción de la información desde las instituciones de salud no entra en el espectro de este trabajo. Sin embargo, se desarrolló la receta de un daemon (servicio) genérico a ser implementado dentro de los servidores de las instituciones de salud. Este daemon tiene como propósito extraer cada cierto periodo de tiempo los registros médicos actualizados, normalizarlos y transmitirlos a la arquitectura de Jobs previamente descrita.

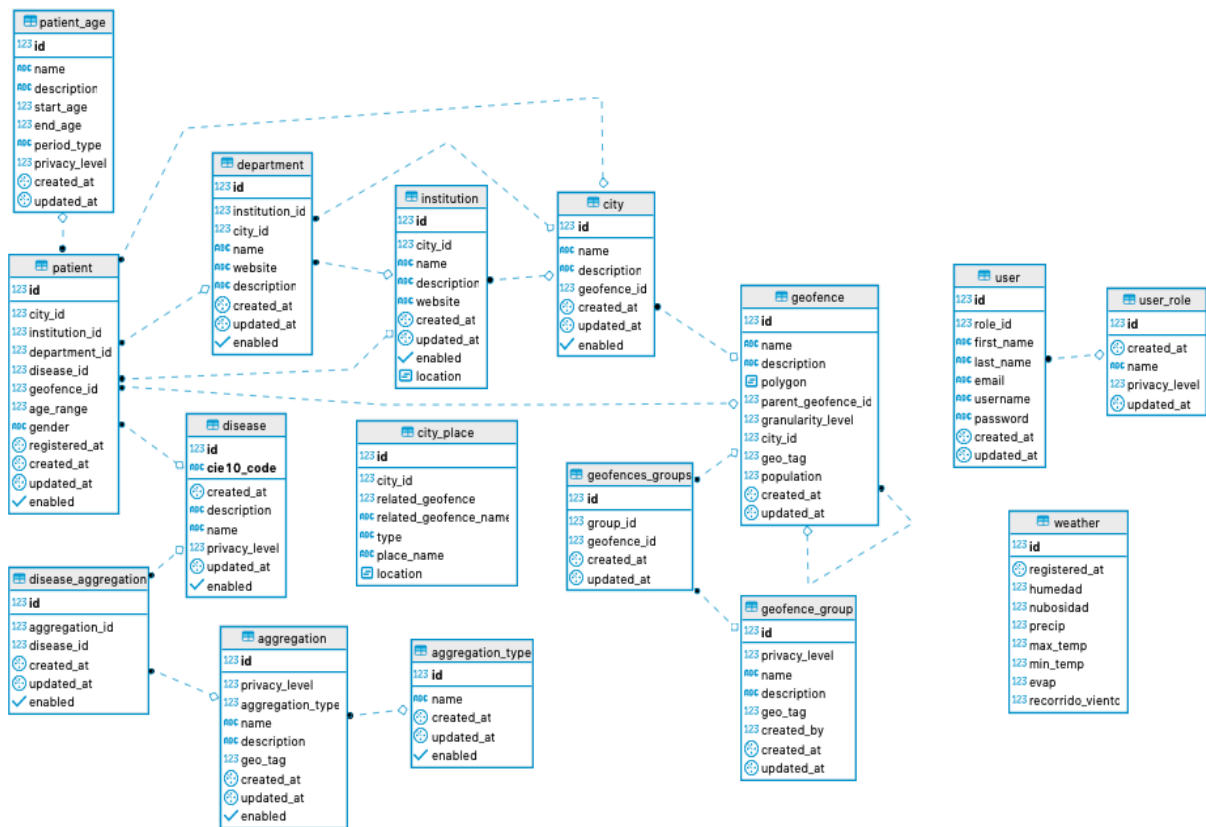
### **2.3 Sistema de almacenamiento de datos**

El sistema de almacenamiento de datos fue diseñado pensando en la consistencia y en la velocidad de búsqueda de los datos como núcleos del sistema. Es por esto por lo que, se divide el sistema de almacenamiento de datos en dos componentes: (1) Persistencia en disco: PostgreSQL y (2) el esquema de cacheo de datos: DataLoader (Redis y memoria de servidor).

El primer componente es una base de datos PostgreSQL (PSQL). Esta base de datos es conocida por ser robusta, relacional y con alta tolerancia a fallos [29]. Esto proporciona consistencia e indexación óptima de los datos dentro de las tablas. El uso de PSQL como pilar para el almacenamiento de los datos, trae consigo algunas ventajas. Entre estas se encuentra la posibilidad de agregar funciones en la base de datos utilizando el lenguaje de programación PSQL, ejecutar disparadores (triggers) al momento de detectar algún cambio en la información y ejecutar rutinas para la creación de respaldos automáticos.

El esquema de base de datos relacional propuesto contempla los diferentes requerimientos de arquitectura y las necesidades de las partes-interesadas. El esquema soporta el almacenamiento de la información anonimizada de pacientes de diferentes ciudades e instituciones de salud. Cada paciente almacenado es asociado con la información del cuadro clínico (enfermedades) que padece y con el sector urbano (definido como geo zona) al cual fue contextualizado. Además, se tomó en consideración la posibilidad de definir y almacenar sectores de interés personalizados, compuestos por diferentes geo zonas. Así como también la administración de usuarios con diferentes niveles de acceso a la información, entre otros.

En el capítulo 3 se discutirá de cómo estas características del almacenamiento se reflejan en las funcionalidades del visualizador. El esquema de base de datos completo se puede observar en la Figura 2.3.



**Figura 2.3. Esquemas de base de datos relacional propuesto para el MEG.**

El segundo componente del sistema de almacenamiento es un esquema de caché de datos, llamado DataLoader [30]. Este esquema utiliza un sistema de jerarquías de memoria caché, que comienza con la memoria del propio servidor web, seguido de una base de datos No Structured Query Language (NoSQL) clave-valor con alta velocidad en consultas (Redis), hasta el almacenamiento en disco (PSQL).

El orden de la jerarquía de las cachés es dado por la velocidad de transmisión de bits que cada caché posee y la frecuencia de las consultas. De tal manera que, las consultas realizadas con mayor frecuencia se almacenen en la caché que posea una mayor velocidad de transmisión de bits, acelerando así las consultas de los usuarios.

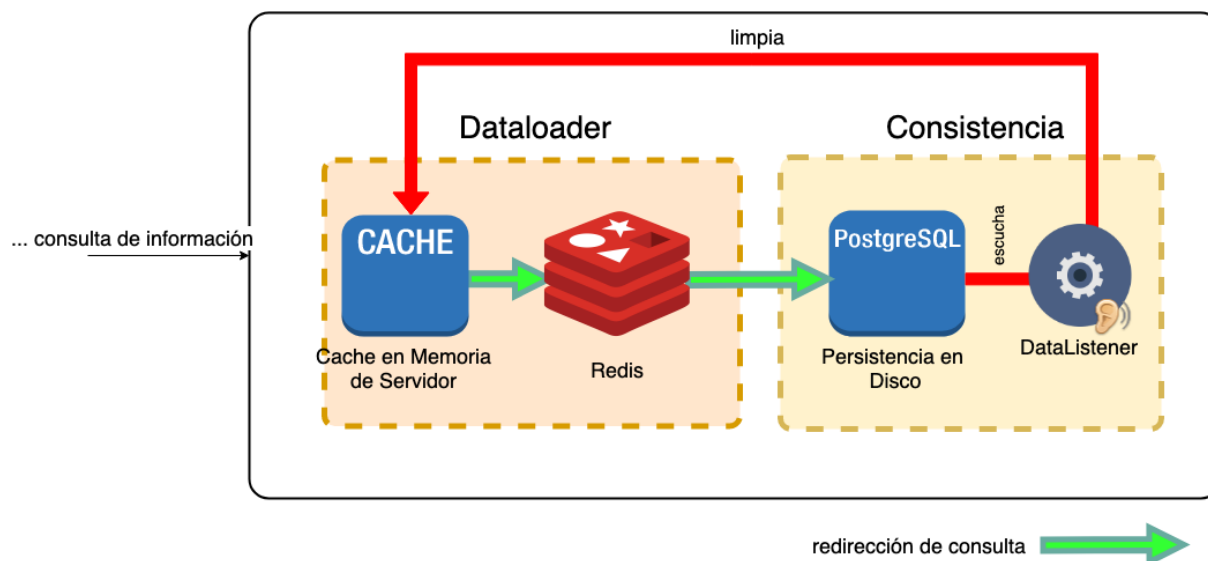
Es por esto por lo que, las consultas que tengan una mayor frecuencia son cargadas en la memoria del servidor web. Las consultas realizadas con una frecuencia media son alojadas en Redis y las consultas realizadas con una

frecuencia baja son extraídas de PSQL (almacenamiento en disco). Si la consulta realizada no se encuentra en los primeros niveles de jerarquía, la información se obtiene de PSQL y se alimenta a los primeros niveles de la jerarquía para acelerar las siguientes consultas de la misma información.

Este sistema jerárquico de caché fue diseñado, como parte de un conjunto de soluciones planteadas para mejorar la experiencia de la velocidad de renderizado del MEG. Esta velocidad de renderizado es la sumatoria de tiempo que toman dos procesos principales al momento de cargar el mapa. Estos procesos son: la respuesta de los requerimientos del API y el renderizado de los componentes en el navegador. El DataLoader está pensado para incrementar la velocidad de respuesta de los requerimientos del API.

Dentro del esquema de DataLoader (Figura 2.4) se plantea tener dos niveles de caché ordenados por velocidad de transferencia de bits. De tal manera que la caché más rápida sea consultada en primera instancia. Esta primera caché se encuentra alojada en la memoria del mismo servidor consultado. En caso de no encontrarse la consulta en esta primera caché, se realiza la búsqueda en el segundo nivel de caché: Redis. Finalmente, de no encontrarse la consulta en ninguno de los niveles, esta se realiza en la base de datos PSQL y se actualizan los niveles de memoria caché con el dato consultado.

Para completar el ciclo de las consultas de datos, las actualizaciones de la base de datos de PSQL serán monitoreadas con un proceso (es decir, DataListener) que se encarga de limpiar toda la jerarquía de cachés. El objetivo de este proceso es poder mantener la consistencia de la información. Este proceso consiste en una conexión a PSQL que escucha y detecta cualquier cambio o actualización dentro de la misma. Al suscitarse este escenario, el proceso limpia la jerarquía de cachés asociadas a los datos actualizados. Por lo que, en las nuevas consultas la información se va a obtener directamente de PSQL (consulta lenta), actualizando así la caché para las futuras consultas (consulta rápida).



**Figura 2.4. Ciclo de vida de la jerarquía de cachés.**

## 2.4 Arquitectura de servidor

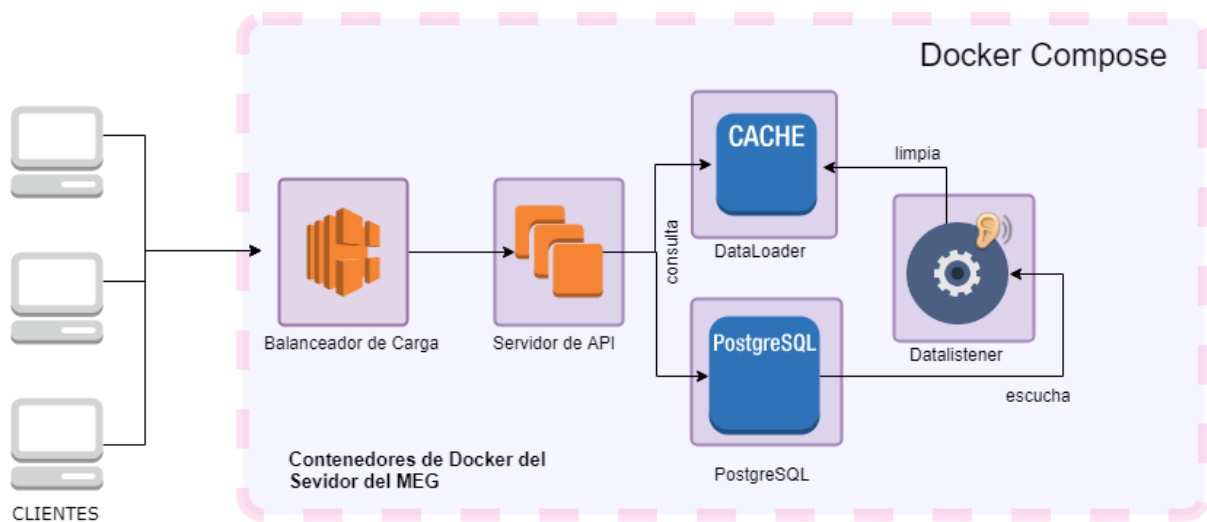
El tercer componente del sistema comprende la arquitectura utilizada para la intercomunicación de los componentes. Este componente maneja la recuperación de información e interfaces gráficas a través de solicitudes cliente-servidor. Las solicitudes son distribuidas con un balanceador de carga (LoadBalancer) hacia el API (Node.js + Express.js) que recupera la información del esquema de almacenamiento de datos.

El servidor se encuentra instanciado sobre una arquitectura llamada Docker. Docker permite levantar instancias de un servidor o un conjunto de servidores basándose en una receta definida, dando la habilidad de levantar todo el ambiente de manera rápida [31]. Esto permite poder desarrollar cada componente del software con facilidad y agilidad, ya que se dispone de todo el ambiente de desarrollo en una misma máquina. Además de esto, Docker facilita la ejecución de pruebas previas al lanzamiento de una nueva versión del software.

El servidor web se encuentra declarado sobre un administrador de recetas de Docker llamado Docker Compose [32]. Docker Compose administra la creación de los diferentes servicios a crearse en la infraestructura y su interconexión.

Los componentes de la infraestructura de servidor instanciados dentro de Docker son los siguientes (Figura 2.5):

- Balanceador de carga (LoadBalancer) en un servidor de NGINX: Se encarga de balancear los requerimientos de los clientes en caso de escalar a una arquitectura con varios servidores [33].
- Servidor de API: Node.js + Express.js
- Caché NoSQL Redis (DataLoader)
- Almacenamiento en disco (PSQL)
- DataListener (Proceso de Node.js)



**Figura 2.5. Arquitectura de Docker del MEG sobre Docker Compose.**

### 2.4.1 Definición de API

El API posee dos abstracciones principales: a) API de ingreso y procesamiento de datos y b) API de búsqueda y visualización de datos. La primera abstracción comprende una API REST HTTPS, para el ingreso de nuevos datos para su posterior procesamiento. Es decir, este API se encarga de la anonimización y guardado mediante la arquitectura de Jobs discutida en la sección 2.2. La segunda abstracción, comprende una API REST HTTP para la obtención de la información a renderizar el mapa epidemiológico.

Las dos APIs están soportadas por un servidor NGINX dockerizado cuyo objetivo es actuar como balanceador de cargas del servidor de API. De esta manera se realiza el respectivo redireccionamiento de los endpoints, soportando un

escalamiento horizontal de ser necesario en un futuro. La Figura 2.6 muestra la abstracción de API en detalle.

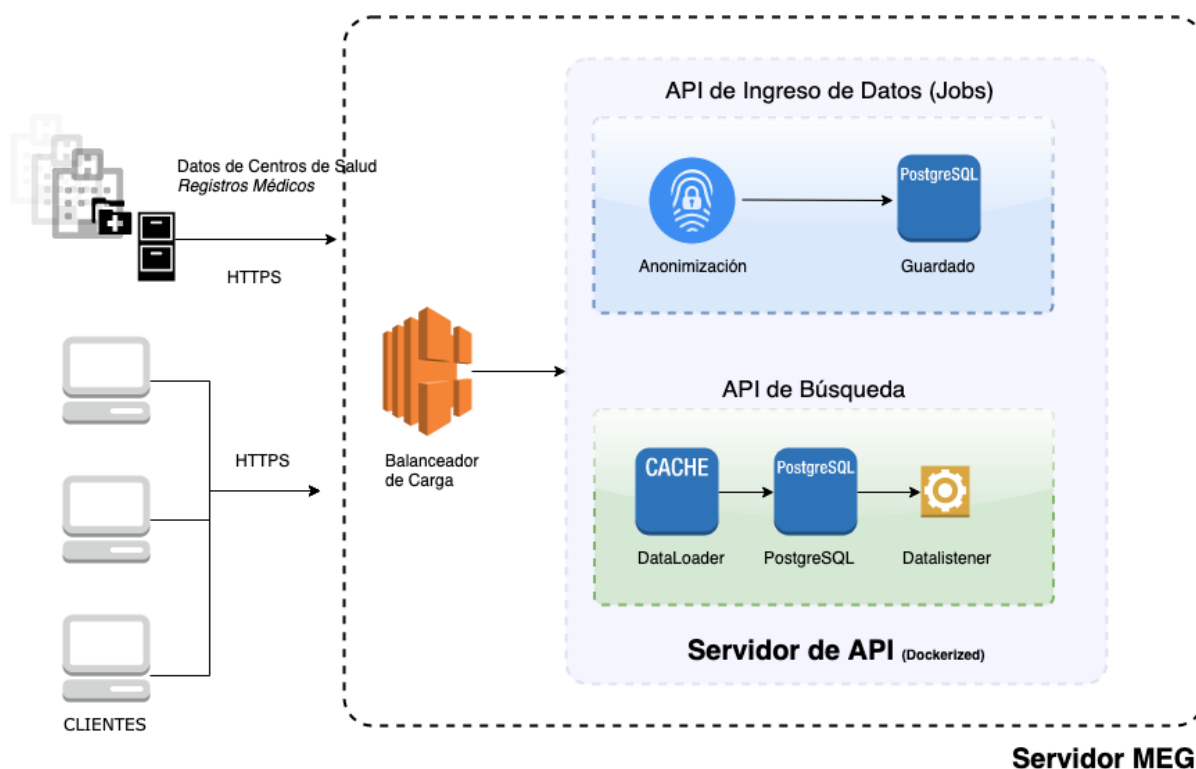


Figura 2.6. Abstracción de ingreso de datos y abstracción de búsqueda.

#### 2.4.1.1 Diccionario de términos de APIs

Esta subsección comprende la construcción de un diccionario de términos utilizados dentro de la definición programática del API. Este diccionario de términos se encuentra descrito en detalle en la Tabla 2.2.

Tabla 2.2. Diccionario de términos de API.

Término	Definición
<b>geofence</b> (Geo zona)	Define la geometría del polígono que delimita la zona geográfica a la cual pertenece un conjunto de datos.
<b>granularityLevel</b> (Granularidad)	Define la granularidad de la geo zona. Los niveles se definen de la siguiente manera: <ul style="list-style-type: none"> <li>Continente: 1</li> </ul>

	<ul style="list-style-type: none"> <li>• País: 2</li> <li>• Región: 3</li> <li>• Provincia/Estado: 4</li> <li>• Ciudad: 5</li> <li>• Parroquia/Comuna: 6</li> <li>• Sector/Micro barrio: 7</li> <li>• Barrio: 8</li> </ul>
<b>ageRange</b> (Rango de edad)	<p>Define el rango de edades. Sus identificadores van desde el 0 hasta el 10:</p> <ul style="list-style-type: none"> <li>• 0 - 1 meses</li> <li>• 1 - 12 meses</li> <li>• 1 - 3 años</li> <li>• 3 - 10 años</li> <li>• 10 - 18 años</li> <li>• 18 - 24 años</li> <li>• 25 - 34 años</li> <li>• 35 - 44 años</li> <li>• 45 - 54 años</li> <li>• 55 - 64 años</li> <li>• + 64 años</li> </ul>
<b>userRole</b> (Rol de usuario)	<p>Define los diferentes roles que se encuentran asociados a los usuarios. Los roles de usuario se definen de la siguiente manera:</p> <ul style="list-style-type: none"> <li>• Usuario visitante</li> <li>• Estudiante</li> <li>• Investigador</li> <li>• Doctor</li> <li>• Autoridad</li> <li>• Administrador</li> </ul>
<b>privacyLevel</b> (Nivel de privacidad)	<p>Define los niveles de privacidad a los cuales tienen accesos los usuarios dentro del sistema. Se representa como un número entero que va desde el 0 hasta el 100.</p> <p>Cada entidad dentro del sistema (es decir, enfermedades, instituciones, ciudades, pacientes, departamentos) se encuentra asociada a un nivel de privacidad. Y, solamente se pueden visualizar las entidades cuyo nivel de privacidad sea menor o igual al nivel de privacidad del usuario que consulta la información. Por ejemplo, al momento de configurar el nivel de privacidad de la ciudad de Guayaquil a 50, solamente los usuarios con un rol que posea un nivel de privacidad mayor o igual a 50 pueden consultar datos sobre esta ciudad.</p>



	<p>Por defecto, todas las entidades se crean con un nivel de privacidad igual a 0. Y, los usuarios visitantes poseen un nivel de privacidad igual a 0.</p> <p>Para el MEG se han definido los siguientes niveles de privacidad según los diferentes roles de usuario:</p> <ul style="list-style-type: none"> <li>• 0: usuario visitante.</li> <li>• 40: estudiante.</li> <li>• 60: investigador.</li> <li>• 90: doctor.</li> <li>• 100: autoridad.</li> <li>• 100: administradores.</li> </ul>
<b>cie10Code</b> (Código de enfermedad)	Define el código CIE10 de la enfermedad.
<b>aggregation</b> (Grupo de enfermedades)	Define una agregación de enfermedades. Sus identificadores van desde el 0 hasta el 270. Esto representa las 270 agrupaciones de los códigos CIE10 en capítulos y subcapítulos.
<b>aggregationType</b> (Tipo de categoría)	Define el identificador de la categoría. Actualmente existen 2: capítulos (1) y subcapítulos (2).
<b>Institution</b> (Institución de salud)	Define una institución pública o privada de salud.
<b>departament</b> (Subdepartamento de salud)	Define un subdepartamento de salud pública o privada de una unidad de salud más grande.
<b>city</b> (Ciudad)	Define el identificador de alguna ciudad.
<b>geogroup</b> (Conjunto de geo zonas)	Conjunto de geo zonas que definen una geo zona de interés.
<b>absolute</b>	Incidencias de enfermedades absolutas en una geo zona.
<b>relativeToPopulation</b>	Incidencias de enfermedades relativas a la población total de una geo zona.

<b>relativeToPatients</b>	Incidencia de enfermedades relativa al número de pacientes totales de una geo zona.
<b>per1000Habitants</b>	Incidencia de enfermedades absolutas por cada 1000 habitantes.

#### 2.4.1.2 Definición de autenticación y cabeceras de API

El método seleccionado para la autenticación de las APIs de ingreso y procesamiento de datos y el API de búsqueda y visualización de datos fue Basic-Auth. Basic-Auth es ideal para este proyecto debido a su simplicidad y facilidad de implementación sin comprometer la seguridad del sistema. Por otro lado, el formato de transmisión de datos elegido fue Javascript Object Notation (JSON). A pesar de no ser el formato de transmisión de datos de mayor rapidez, es un formato ampliamente utilizado. Esto favorece a la futura integración de terceros con el API.

Dadas estas consideraciones, los encabezados HTTPS de las peticiones al API quedan definidos de la siguiente manera (Figura 2.7):

```
Accept-Charset:utf-8
Authorization:Basic
aWp1azhyZnE1ZmN0d3J0cDZtMjBydWRpOjA5MjA4NjkyYmU2ZTVjODh0YTczYTQ4YzMwYzc2MGEw
Content-Type:application/json
```

**Figura 2.7. Ejemplo de encabezado de requerimiento al API.**

En las siguientes subsecciones (2.4.1.3 y 2.4.1.4) se detalla la definición de los endpoints del API.

#### 2.4.1.3 Definición del API de ingreso y procesamiento de datos (Tabla 2.3)

**Tabla 2.3. Endpoints de ingreso y procesamiento de datos.**

Endpoint	Parámetros(Eg.)	Respuesta	Descripción
<b>POST</b> <b>/data</b> Basic-Auth Application /json	Archivo con Información	{ code: 'OK', message: 'Cargado con éxito'	Crea un proceso (JOB) para realizar la geolocalización, anonimización e ingreso de pacientes en la

		<pre>data: {   jobId: 20,   rows: 300 }</pre>	base de datos. Permite la carga de un archivo en formato texto o un cuerpo de tipo JSON para el envío de información.
--	--	---	---

#### 2.4.1.4 Definición del API de búsqueda de datos (Tabla 2.4)

Tabla 2.4. Endpoints de búsqueda de datos.

Endpoint	Parámetros(Eg. )	Respuesta	Descripción
<b>GET /incidences</b> Basic-Auth Application /json	ageRange: 1 gender: "M"   "F"   "N" cie10: "A99" aggregation: 1 geogroup: 1 geozones: [] institution: 1 department: 1 city: 1 startDate: '01-01-15' endDate: '01-01-19'	<pre>Array&lt;Object&gt; =&gt; [   {     name: SAUCES,     Polygon&lt;Points&gt;: [blob],     metrics: {       absolute: 100,       relativeToPatients: 10,       relativeToPopulation: 9,       per100Habitants: 5,       per1000Habitants:0.5,       ...     }   },   ... ]</pre>	Retorna un arreglo de geo zonas con un conjunto de valores asociado a cada geo zona dado una serie de filtros, y nivel de privacidad de usuario <b>Caso de Uso: Endpoint principal de consulta de datos</b> para generar las visualizaciones.
<b>GET /geofences</b> Basic-Auth Application /json	granularityLevel: 7 cityId: 1 geogroup: 1 (opcional)	<pre>Array&lt;Object&gt; =&gt; [   {     "id": 1,     "name": "GUAYAQUIL",     "description": "Ciudad de Guayaquil",     "polygon": [Geojson]   }, ... ]</pre>	Retorna un arreglo de las geo zonas disponibles.
<b>GET /age</b> Basic-Auth Application /json	N/A	<pre>Array&lt;Object&gt; =&gt; [   {     "id": 1,     "name": "INFANTE I",     "description": "Infantes en etapa neonatal",   } ]</pre>	Retorna un arreglo de los rangos de edades disponibles.

		<pre> "start_age": 0, "end_age": 1, "period_type": "meses" }, ... ] </pre>	
<b>GET</b> <b>/geogroups</b> Basic-Auth Application /json	N/A	<pre> Array&lt;Object&gt; =&gt; [   {     "id": 13,     "privacy_level": 0,     "name": "Mi zona de interes",     "description": "Esta es una descripcion de prueba",     "polygons": Array&lt;GeoJson&gt;   }, ... ] </pre>	Retorna un arreglo con las diferentes zonas de interés definidas por el usuario autenticado.
<b>POST</b> <b>/geogroups</b> Basic-Auth Application /json	name: Str description: Str geofences: [1, 2, 3]	<pre> {   "code": 'OK',   "message": 'Cargado con exito' } </pre>	Crea un grupo de geo zonas.
<b>GET</b> <b>/diseases</b> Basic-Auth Application /json	q: digestivo (Opcional)	<pre> &lt;Object&gt; =&gt; {   "diseases": [     {       "id": 1051,       "cie10_code": "C26",       "description": "Tumor..",       "name": "Tumor..."     }, ...   ],   "aggregation": [     {       "id": 11,       "name": "Enfermedades del aparato digestivo",       "description": "Capítulo 11: K00-K93",       "numberofdiseases": "479"     }, ...   ] } </pre>	Retorna un objeto con arreglos de las enfermedades disponibles y los grupos de enfermedades disponibles.
<b>GET</b> <b>/institutions</b> Basic-Auth Application /json	cityId: 1	<pre> Array&lt;Object&gt; =&gt; [   {     "id": 1,     "city_id": 1,     "name": "Hospital León Becerra",     "description": "El Hospital León Becerra",     "website": "https://www.bspi.org/hospital_le on_becerra.html",     "location": Object&lt;Geojson&gt;   }, ... ] </pre>	Retorna un arreglo con todas las instituciones disponibles de una ciudad.

<b>GET</b> <b>/departments</b> Basic-Auth Application /json	institutionId: 1	Array<Object> ⇒ [ { "id": 1, "institution_id": 1, "city_id": 1, "name": "HLB General", "website": null, "description": "HLB General", }, ... ]	Retorna un arreglo con todos los departamentos disponibles de una institución.
<b>GET</b> <b>/cities</b> Basic-Auth Application /json	N/A	Array<Object> ⇒ [ { "id": 1, "name": "Guayaquil", "description": "Ciudad de Guayaquil", "geofence_id": 1 }, ... ]	Retorna un arreglo con todas las ciudades disponibles.

## **2.5 Datos externos: datos meteorológicos y datos de tránsito**

La correlación entre factores meteorológicos y el tránsito con la salud ya ha sido estudiada previamente [8, 9]. Por encima de esto, muchos sectores de Guayaquil poseen su propio microclima diferente al de otros sectores de la ciudad. Así como también, diferentes niveles de contaminación de aire durante el transcurso del día debido a las diversas fluctuaciones de tráfico vehicular que existe.

Por lo tanto, la integración de datos meteorológicos y datos tránsito en conjunto con los datos de salud dentro del MEG es parte importante para la detección de enfermedades que guarden correlación con estos factores. Así como también, es relevante para la identificación de sectores que se encuentren gravemente afectados en su estado epidemiológico por estos factores externos.

Es posible obtener los datos meteorológicos directamente de estaciones meteorológicas dentro de Guayaquil. Estas estaciones meteorológicas son propiedad del INAMHI (Instituto Nacional de Meteorología e Hidrología). Las estaciones son capaces de medir los siguientes parámetros meteorológicos con periodicidad horaria: humedad, nubosidad, precipitación, temperatura máxima, temperatura mínima, temperatura promedio, evaporación y velocidad del viento. Sin embargo, para el espectro de este proyecto se realizó una simulación de los datos meteorológicos basado en estos datos.

Los datos de congestionamiento vehicular fueron obtenidos utilizando el API de tránsito de OpenStreetMap. A partir de estos niveles de tránsito vehicular, es posible inferir el nivel de contaminación que produce este volumen de tráfico [44]. Sin embargo, esto no entra en el espectro de este proyecto.

## **2.6 Diseño del visualizador**

El cuarto y último componente del MEG es el visualizador de los datos epidemiológicos. El visualizador es el componente gráfico que transmite la información epidemiológica de la ciudad al usuario final. Por lo tanto, hay que tener especial cuidado con el desarrollo de este componente. El desarrollo del visualizador engloba la creación de buenas experiencias de usuario a través de componentes gráficos que transmiten al usuario la información de manera clara y correcta. Además de presentar la información epidemiológica utilizando diferentes parámetros de filtrado de tal manera que sea útil para la comunidad científica, investigadores y autoridades de salud.

El desarrollo del visualizador se realizó en tres etapas clásicas de la interacción humano computador: (1) prototipado, (2) pruebas de usabilidad y (3) desarrollo. La etapa (1) correspondió al desarrollo de prototipos no funcionales del visualizador. Esta etapa permitió tener una vista preliminar del producto final sobre la cual se pudieron realizar pruebas de usabilidad para reiterar sobre el diseño construido. De tal manera que se desarrolle un diseño que vaya acorde a los objetivos del proyecto y a las necesidades de los usuarios finales.

A continuación, en la etapa (2) se realizó una sesión de pruebas de usabilidad con los usuarios finales. Las pruebas de usabilidad son sesiones en cuales el usuario final interactúa con una aplicación y realiza sugerencias y comentarios sobre esta. Además, se evalúa la usabilidad e intuitividad del diseño desde el punto de vista del usuario final mediante la observación de su interacciones y acciones ejecutadas en el aplicativo. Usualmente el desarrollo de un producto comprende un ciclo de varias iteraciones de pruebas de usabilidad. Debido a que la versión anterior del MEG ya se encuentra validada por los usuarios finales, se realizó solamente una sesión de pruebas de usabilidad con doctores, epidemiólogos y estudiantes universitarios.

Finalmente, la etapa de implementación del visualizador fue ejecutada utilizando React.js [34] en conjunto con Redux [35]. React.js es una librería de JavaScript para construir interfaces de usuario interactivas y reactivas de forma sencilla. React.js es orientado al desarrollo de componentes individuales, que en conjunto construyen una interfaz completa. Cada componente mantiene un estado que puede cambiar dada ciertas acciones de los usuarios. Esto permitió crear componentes que reaccionan de manera rápida a las acciones de los usuarios, pudiendo generar buenas experiencias dentro de la interfaz.

Debido a que el desarrollo en React.js es basado en componentes, existe la posibilidad de reutilizar las mismas en cualquier otro proyecto que utilice la misma tecnología. Como caso de estudio, en un futuro se tiene planeado adaptar estos componentes para que sean parte del proyecto DataCity.



# CAPÍTULO 3

## 3 RESULTADOS Y ANÁLISIS

En la primera sección de este capítulo se describe la implementación y el análisis de resultados de la arquitectura de servidor. Lo cual incluye la información ingresada en el sistema de almacenamiento en disco (PSQL), pruebas de estrés al servidor de API (API de Búsqueda y API de ingreso de datos) y pruebas de carga del sistema de cacheo DataLoader detallado en la sección 2.3.

A continuación, en la sección 3.2. se detalla la implementación del visualizador, componente gráfico que transmite la información epidemiológica de la ciudad al usuario final. En esta sección se describen las múltiples funcionalidades dentro de la interfaz gráfica que nos permite realizar el sistema de almacenamiento de datos. Así como también, se realiza un contraste entre los tiempos de carga de la nueva versión del MEG y la versión anterior del MEG. Además, se discuten los resultados de las pruebas de usabilidad ejecutadas con el mapa de salud.

Finalmente, en la sección 3.3 se presentan los análisis de los resultados obtenidos con la metodología de geolocalización y anonimización propuesta.

### 3.1 Arquitectura de servidor

El sistema de almacenamiento en disco propuesto fue alimentado con la información detallada en la Tabla 3.1. Incluyendo, el ingreso de 72,294 registros de pacientes anonimizados y geolocalizados exitosamente por el API de ingreso de datos. El throughput del API de ingreso de datos es de 10.83 registros por segundo. Es decir, que se han podido geolocalizar y anonimizar 10.83 registros de pacientes por segundo. En la sección 3.3 se expone la validez de los resultados del algoritmo de geolocalización y anonimización.

**Tabla 3.1. Datos almacenados en el esquema de almacenamiento del MEG.**

Entidad	Descripción	Registros Ingresados
País	Geo zona de un país.	1
Ciudad	Geo zona de una ciudad.	1
Parroquias	Geo zonas correspondientes a las parroquias de Guayaquil.	16
Sectores Urbanos	Geo zonas correspondientes a los sectores urbanos de Guayaquil.	177
Barrios Urbanos	Nombre de barrios urbanos de Guayaquil con su geo zona (sector urbano) a la que pertenece.	1,110
Intersecciones de Calles	Punto geográfico de la intersección de dos calles dentro de Guayaquil.	111,134
Enfermedades	Enfermedades basadas en el código CIE10.	14,484
Agrupaciones de Enfermedades	Capítulos (agrupación de enfermedades) basados en los códigos CIE10.	270
Instituciones de Salud	Institución de Salud dentro de Guayaquil (HLB).	1
Subdepartamento de Instituciones de Salud	Subdepartamento de salud de una institución.	1
Pacientes Anonimizados	Información anonimizada y geolocalizada a una geo zona de cada paciente.	72,294

La Tabla 3.2 detalla los resultados del mecanismo de cacheo DataLoader en las consultas realizadas al endpoint más crucial dentro del sistema: */incidences*. Este endpoint es el encargado de computar un arreglo de geo zonas con sus respectivos valores de incidencias de enfermedades de pacientes para poder generar las visualizaciones del mapa de salud.

**Tabla 3.2. Resultados de tiempos de respuesta del servidor de API utilizando y no utilizando el mecanismo de cacheo DataLoader en el endpoint de */incidences*. La celda resaltada representa la consulta en la cual la caché del DataLoader no se encuentra lleno.**

GET <i>/incidences</i> Prueba #	Sin DataLoader (ms)	Utilizando DataLoader (ms)
1	2054	1958
2	1853	253
3	2149	230
4	1943	173
5	2109	220
PROMEDIO	2021.6	566.8 (219.0)

Los tiempos de respuestas del API a este endpoint disminuyeron en promedio un 71.9% luego de la primera consulta gracias al esquema DataLoader. La primera consulta es relativamente lenta dado que se realiza en el almacenamiento en disco (PSQL). Esto es debido a que, en la primera consulta, la caché se encuentra vacía. En las siguientes llamadas al endpoint, el servidor de API encuentra la información dentro de la caché y ejecuta las consultas en ella.

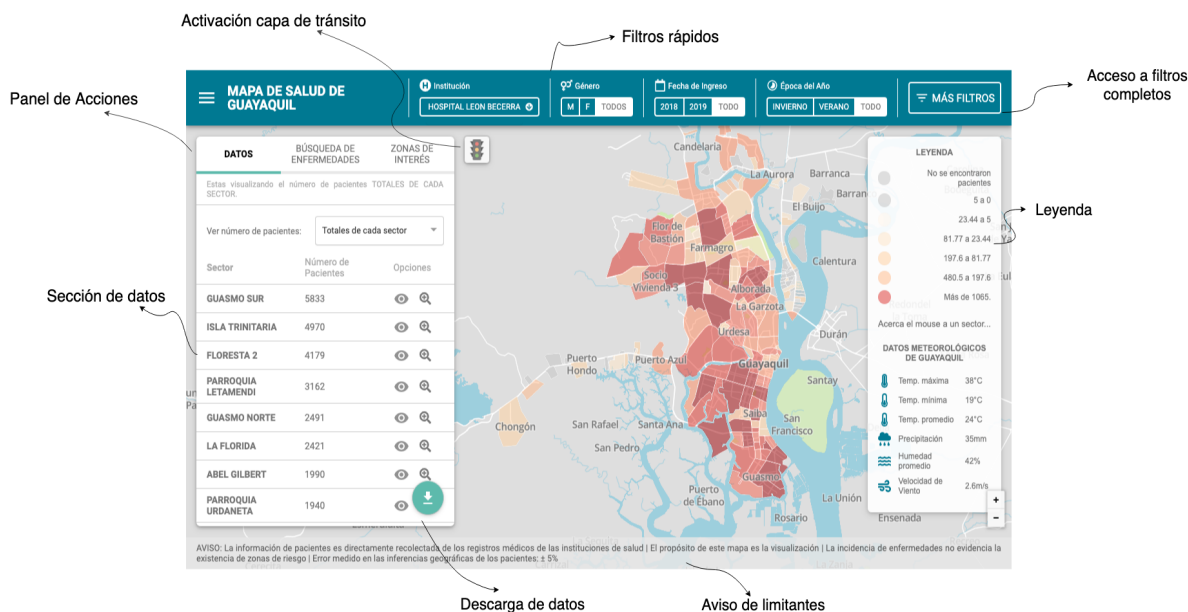
Finalmente, utilizando la herramienta Apache JMeter [45] se llevaron a cabo pruebas de estrés sobre el endpoint más utilizado en el sistema: /incidences. Cada prueba consistió en 20 llamadas paralelas por segundo a este endpoint durante 60 segundos (1200 requerimientos/minuto). Estas pruebas se llevaron a cabo en 2 clusters sobre un servidor 2.7 GHz Intel Core i5 con 8 GB 1867 MHz DDR3 de Memoria RAM. Se alcanzó un tiempo de respuesta promedio de 355ms entre todos los requerimientos. Llegando hasta un máximo tiempo de respuesta de 1597ms en promedio. Además, el servidor no sufrió errores críticos en ninguno de los requerimientos. Los resultados en detalle de las pruebas de estrés conducidas se encuentran en la Tabla 3.3.

**Tabla 3.3. Resultados de Apache JMeter en /incidences. Observación: Experimentos conducidos sobre 2 clusters en un servidor 2.7 GHz Intel Core i5 con 8 GB 1867 MHz DDR3.**

<b>GET /incidences 1200 req/min Prueba #</b>	<b>PROMEDIO (ms)</b>	<b>Max (ms)</b>	<b>Min (ms)</b>	<b>90% (ms)</b>	<b>95% (ms)</b>	<b>99% (ms)</b>	<b>Std. Dev. (ms)</b>	<b>Err %</b>
1	354	1532	54	1488	1512	1530	582	0.0
2	370	1685	72	1621	1638	1651	615	0.0
3	341	1574	41	1492	1534	1569	580	0.0
PROMEDIO	355	1597	55.6	1533.7	1541.7	1583.3	592	0.0

### 3.2 Visualizador

La Figura 3.1 muestra el visualizador del MEG, resaltando cada uno de sus componentes principales. El visualizador del MEG cuenta con cuatro piezas clave: (1) Panel de acciones, (2) filtros rápidos (3) filtros completos y (4) el mapa.



**Figura 3.1. Visualizador del MEG.**

El panel de acciones posee tres secciones principales separadas por pestañas (tabs), las cuales son: Sección de datos, sección de búsqueda de enfermedades y sección de zonas de interés. En la Figura 3.2 se muestran las tres secciones del panel de acciones en detalle.

En la sección de datos se presenta al usuario el listado de sectores con el número de pacientes de cada sector según los filtros que se han aplicado. Además, el usuario puede elegir dentro de un combobox la forma de visualizar el número de pacientes de cada sector. Existen cuatro formas diferentes: el número de pacientes absolutos, el número de pacientes relativos a la población del sector, el número de pacientes filtrados relativos al número de pacientes totales del sector, y el número de pacientes absolutos por cada 1000 habitantes. Ver Figura 3.2.a

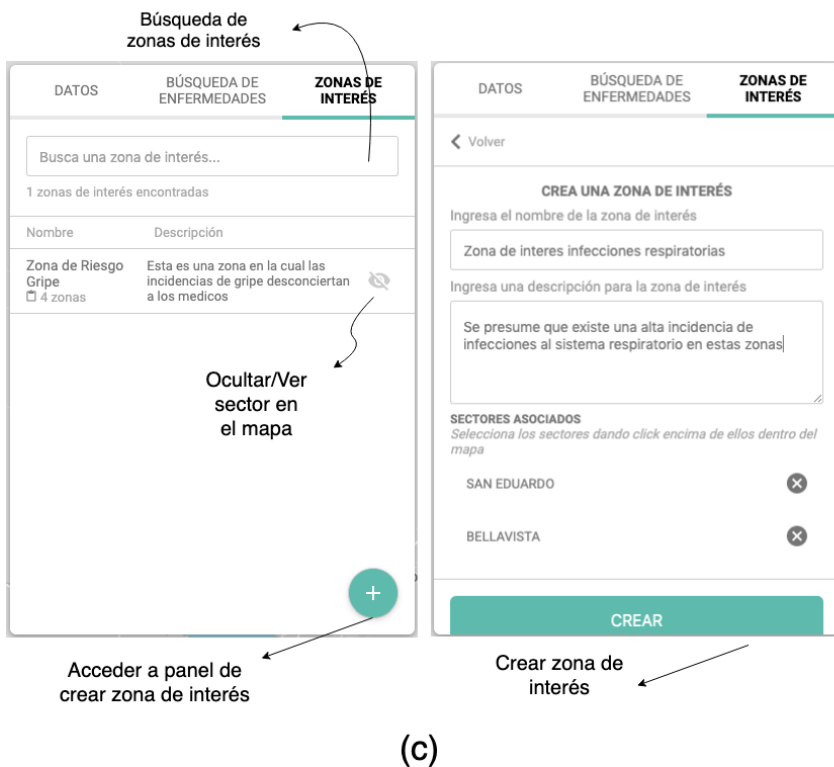
En cada ítem del listado de sectores se da al usuario la posibilidad de ver u ocultar el sector dentro del mapa. Esto se realiza al dar clic en el ícono en forma de un ojo para ocultar el sector, y dar clic nuevamente en el mismo ícono para volver a mostrarlo. Además, al dar clic en el ícono en forma de una lupa, el mapa realiza un acercamiento automático al sector en cuestión. Finalmente, la sección de datos da la posibilidad de descargar los datos mostrados en un archivo con un formato separado por comas (CSV) al dar clic en el botón flotante con el ícono de descarga. Ver Figura 3.2.a

Cabe mencionar que la descarga de este archivo es de utilidad para estudiantes e investigadores que deseen realizar diferentes tipos de análisis o reportes con los datos que hayan obtenido a partir del mapa de salud. Además, al dar clic sobre este botón, los datos descargados por el usuario en paralelo se guardan en la plataforma de CKAN [46]. La cual es el motor detrás del proyecto DataCity. Esto con el propósito de alimentar la base de datos de DataCity con los datos del MEG.

En la sección de búsqueda de enfermedades el usuario puede buscar una enfermedad por palabras clave y aplicarla como filtro del mapa de salud. A partir de esta búsqueda, se le presentan los resultados de enfermedades y agrupaciones de enfermedades que coincidan con los términos buscados. Ver Figura 3.2.b

Finalmente, la sección de zonas de interés le muestra al usuario las zonas de interés disponibles para la ciudad visualizada, y se presenta la opción para poder crear una nueva zona de interés. Una zona de interés se define como un grupo de sectores para los cuales un usuario desea aislar las observaciones del mapa de salud. Por ejemplo: Se define la zona de interés con los sectores: “San Eduardo” y “Bellavista”, debido a que en estos sectores el usuario presume que existe una alta incidencia de infecciones al sistema respiratorio (Ver Figura 3.2.c). De esta manera, se pueden aislar las observaciones, y trabajar solamente visualizando estos sectores específicos en el mapa. Así como los sectores en la sección de datos, las zonas de interés son ocultadas o visualizadas dando clic sobre el ícono en forma de ojo.

Para crear una nueva zona de interés el usuario deberá dar clic en el botón flotante en forma de “+”. Luego, el usuario deberá ingresar un nombre, una descripción y seleccionar los sectores asociados a esa zona de interés dando clic sobre estos dentro del mapa. Ver Figura 3.2.c



**Figura 3.2. Panel de Acciones del MEG. a) Sección de datos. b) Sección de búsqueda de enfermedades. c) Sección de zonas de interés.**



**Figura 3.3. Filtros rápidos para aplicar sobre los pacientes en el MEG.**

A continuación, en la barra de herramientas superior se encuentran los filtros rápidos para aplicar sobre los pacientes visualizados. Con estos filtros se tiene la posibilidad de a través de un solo clic, cambiar las características de los pacientes mostrados. Tales como: la institución a la cual ingresaron los pacientes, el género de los pacientes, la fecha de ingreso de los pacientes y la época del año en la cual ingresaron los pacientes. La Figura 3.3 muestra este componente.

A la totalidad de los filtros disponibles se accede dando clic sobre el botón “MÁS FILTROS”, ubicado en la barra de herramientas en la esquina superior derecha. Al realizar esta acción se nos presenta una ventana flotante de tipo modal para poder definir los valores de todas las variables disponibles por las cuales se pueden filtrar los pacientes. Estos filtros son soportados y ejecutados de forma rápida gracias a el esquema de almacenamiento de datos. La Figura 3.4 muestra la ventana flotante de filtros dentro del MEG.

**FILTROS**

Selecciona las características por las cuales deseas filtrar los pacientes mostrados en el mapa de salud.

Selecciona una ciudad: GUAYAQUIL

Selecciona una institución: HOSPITAL LEON BECER...

Selecciona un departamento: GENERAL

Selecciona un rango de edad: TODOS

Selecciona un género: TODOS

Dividir la ciudad por: SECTORES URBANOS

Selecciona un rango de fechas: 01-01-2015

**FILTRAR**

**Figura 3.4. Ventana flotante de filtros para aplicar sobre los pacientes en el MEG.**

Finalmente, se encuentra el componente del lienzo del mapa. En este se muestran todos los sectores disponibles en la base de datos a la granularidad seleccionada por el usuario. Por defecto, se muestra siempre la granularidad disponible más baja. En este caso, los sectores urbanos.

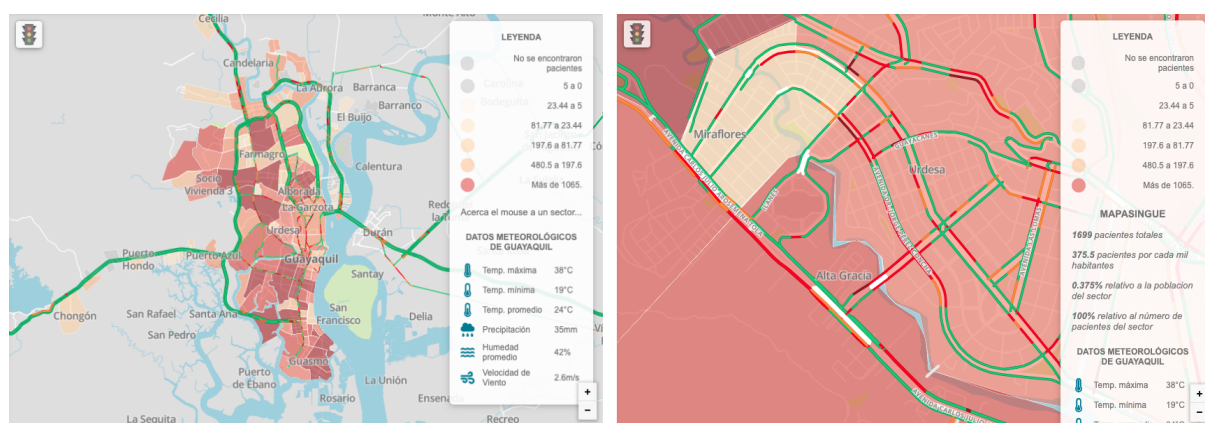
Cada sector es asignado dinámicamente a un color en una escala monocromática de 7 tonalidades de rojo. La intensidad del color asignado al sector depende de los datos de pacientes que se estén visualizando en el momento, y de la forma en que se deseen visualizar estos datos (valores absolutos, relativo a la población del sector, relativo al número de pacientes del sector, y valores absolutos por cada 1000 habitantes). Es decir que, cada filtro aplicado va a alterar la manera en que se visualizan los colores en el mapa. Los rangos de valores que determinan el color asignado son determinados con estadística de cuantiles. Los cuantiles son puntos que dividen las muestras de una distribución en intervalos homogéneos. En este caso, se ha realizado una división de la muestra en 8 cuantiles equidistantes. Adicional a esto, se asigna una tonalidad de gris para las zonas donde existan



menos de dos pacientes. Esto, para salvaguardar la anonimización de los registros únicos existentes por sector.

Además, dentro del lienzo del mapa se ubica la leyenda de este. Esta describe los colores utilizados para cada rango de valores y los datos meteorológicos de la ciudad. En adición a esto, cuando el usuario desliza el ratón sobre un sector, se presenta información detallada sobre los pacientes del sector según los filtros aplicados.

Finalmente, el botón ubicado en la esquina superior izquierda del lienzo del mapa permite activar la capa de transito vehicular de OpenStreetMap. Esta capa es visualizada sobre los sectores del mapa con los datos en tiempo real de tránsito. La Figura 3.5 muestra la capa de tránsito vehicular activada en el MEG. Esta capa funciona por niveles de acercamiento, es decir, mientras más acercamiento se realice al mapa, más detalles de la capa de transito en calles secundarias serán visibles.



**Figura 3.5. Capa de tránsito vehicular activada en el MEG.**

Por último, el visualizador posee un pie de página en donde se exponen las limitantes de este. Estas limitantes son basadas en el porcentaje de error en el algoritmo de anonimización y localización de pacientes. Ver Figura 3.1.

### 3.2.1 Tiempo de carga de interfaz

Se ejecutaron pruebas de carga para medir el tiempo que demora el mapa de salud en ser cargado. De 8 experimentos, se obtuvo un tiempo promedio de carga de 3,441ms. Este tiempo es medido desde que se realiza el requerimiento http a la

URL, hasta que se recibe el último recurso necesario para visualizar y utilizar el mapa de salud. Comparado con la versión anterior del MEG, se logró reducir el tiempo promedio de carga en un 79.89%. Además, al compararlo con otros mapas de salud a nivel mundial analizados en la subsección 1.4.1, el MEG desarrollado obtuvo el menor tiempo promedio de carga. En la Tabla 3.4 se detallan los resultados de los experimentos de tiempo de carga de interfaz.

**Tabla 3.4. Resultados de tiempos de carga del visualizador del Mapa de Salud en la versión anterior del MEG y la versión desarrollada en este proyecto.**

<b>Carga de mapa de salud. Prueba #</b>	<b>MEG 1.0 (ms)</b>	<b>MEG 2.0 (ms)</b>	<b>HealthMap [22] (ms)</b>	<b>GeoSalud [26] (ms)</b>
1	15,840	4,110	3,660	9,940
2	15,850	3,110	4,170	7,030
3	17,460	3,380	3,760	9,880
4	17,040	3,200	4,449	7,340
5	18,310	3,580	4,160	8,930
6	15,760	3,170	3,440	7,940
7	15,540	3,330	3,440	11,050
8	21,100	3,650	3,410	7,840
<b>PROMEDIO</b>	<b>17,112</b>	<b>3,441</b>	<b>3,811.1</b>	<b>8,743.8</b>

### 3.2.2 Pruebas de usabilidad

Se condujeron pruebas de usabilidad con 12 sujetos de prueba distribuidos en 3 grupos diferentes: 4 doctores especializados en epidemiología (epidemiólogos), 4 doctores investigadores en el área de salud y 4 doctores en medicina. De las 12 pruebas, 8 fueron realizadas en la facultad de medicina de la Universidad de Guayaquil, y las 4 pruebas restantes fueron realizadas en los despachos privados de los doctores.

Antes de empezar las pruebas, cada sujeto de prueba firmó un consentimiento informado en el cual se le detalló el propósito de las pruebas y se obtuvo su consentimiento para poder realizar grabaciones de audio y captura de su actividad dentro de una computadora. El consentimiento informado utilizado se encuentra en el apéndice A.

El instrumento utilizado para conducir las pruebas de usabilidad se encuentra descrito en el apéndice B. Cabe destacar que los resultados fueron agregados por grupo de sujetos de prueba. Es decir, doctores, epidemiólogos e investigadores.

Cada sesión de prueba se dividió en cuatro partes. Se empezó cada sesión mencionando una breve introducción al proyecto y se procedió a realizar preguntas para conocer los antecedentes de los sujetos de prueba. Entre estas preguntas determinamos la profesión específica de los sujetos, sus años de experiencia, las actividades de investigación que han realizado, su autopercepción de agilidad navegando por internet y finalmente su experiencia previa con mapas de salud. Los resultados de esta primera parte se encuentran en la Tabla 3.5.

A continuación, se les pidió a los sujetos de prueba ejecutar dos tareas concretas dentro del mapa de salud: Identificar zonas de alta incidencia aplicando un filtro por enfermedad y crear una zona de interés. En la realización de cada tarea, se midió el tiempo tomado en ejecutar la tarea en segundos y el número de clics dados para realizar la tarea. Además, durante la realización de las tareas, se registraron las preguntas que los sujetos realizaron al moderador durante el proceso. Se definió un tiempo máximo de 5 minutos para que los sujetos puedan completar cada tarea. Se consideró que una tarea no fue completada con éxito cuando transcurrieron 5 minutos sin que el sujeto haya completado la tarea. Los resultados de la realización de tareas se encuentran descritos en la Tabla 3.6.

Finalmente, se realizó una serie de preguntas con respecto a la utilidad, intuitividad, agrados y desagradados del mapa de salud. Estas preguntas se dividieron en dos grupos: entrevista y encuesta de usabilidad. Los resultados de este segmento de las pruebas se encuentran detallados en las Tablas 3.7 y 3.8 respectivamente.

Tabla 3.5. Información acerca de los antecedentes de los sujetos de prueba.

ANTECEDENTES	SUJETOS DE PRUEBA		
	DOCTORES (4)	EPIDEMIÓLOGOS (4)	INVESTIGADORES (4)
¿Cuál es su profesión?	<ul style="list-style-type: none"> <li>- Cirujano.</li> <li>- Obstetra.</li> <li>- Gastroenterólogo.</li> <li>- Neumólogo.</li> </ul>	<ul style="list-style-type: none"> <li>- Epidemiólogos.</li> </ul>	<ul style="list-style-type: none"> <li>- PhD. en educación.</li> <li>- Maestría en salud ambiental.</li> <li>- Medicina.</li> <li>- Medico general.</li> </ul>
¿Cuántos años de experiencia posee en su área?	29 en promedio	18 en promedio	28.25 en promedio
¿Ha realizado o realiza actualmente actividades de investigación? Especifique	<b>2 de 4 han realizado investigación en:</b> <ul style="list-style-type: none"> <li>- Infecciones estomacales.</li> <li>- Desprendimiento de placenta.</li> </ul>	<b>4 de 4 han realizado investigación en:</b> <ul style="list-style-type: none"> <li>- Doctorado en medicina.</li> <li>- Escritura de libro de Epidemiología.</li> <li>- Control de vectores para Latinoamérica y el Caribe.</li> <li>- Perfil epidemiológico de enfermedades metaxénicas en Guayaquil.</li> </ul>	<b>4 de 4 han realizado investigación en:</b> <ul style="list-style-type: none"> <li>- Rendimiento académico.</li> <li>- Salud ambiental.</li> <li>- Malaria.</li> <li>- Relación entre hipertensión arterial y la nutrición.</li> </ul>
¿Ha realizado estudios epidemiológicos? Especifique	Ninguno ha realizado estudios epidemiológicos.	<b>4 de 4 han realizado estudios de:</b> <ul style="list-style-type: none"> <li>- Brotes de Influenza.</li> <li>- Brotes de Dengue.</li> <li>- Enfermedades tropicales transmitidas por vectores.</li> <li>- Enfermedades metaxénicas en Guayaquil.</li> </ul>	<b>1 de 4 han realizado estudios de:</b> <ul style="list-style-type: none"> <li>- Malaria en sectores rurales de Guayaquil.</li> </ul>
¿Cuán ágil se considera navegando por el internet? Siendo: 5. Muy ágil. 4. Ágil. 3. Neutral 2. Poco Ágil. 1. Nada Ágil.	3.75 en promedio	4.25 en promedio	3.75 en promedio
¿Ha interactuado con mapas de salud digitales previamente?	<b>1 de 4</b> Ha interactuado.	<b>4 de 4</b> Han interactuado.	<b>2 de 4</b> Han interactuado.

**Tabla 3.6. Resultados de las pruebas de usabilidad. Parte 1: Realización de tareas dentro del mapa por parte de los sujetos de prueba.**

<b>PARTE 1 REALIZACIÓN DE TAREAS</b> Se definió un tiempo máximo de 5 minutos para la realización de cada tarea. En caso de que el sujeto no haya completado la tarea dentro de estos 5 minutos, la tarea se considera como fallida.		<b>SUJETOS DE PRUEBA</b>		
		<b>DOCTORES (4)</b>	<b>EPIDEMIÓLOGOS (4)</b>	<b>INVESTIGADORES (4)</b>
<b>TAREA 1</b>	Tiempo tomado en ejecutar la tarea medido en segundos.	83.25 en promedio	203 en promedio	132,33 en promedio
	Número de clics dados para realizar la tarea.	7.25 en promedio	10.5 en promedio	8,33 en promedio
	Preguntas realizadas durante la ejecución de la tarea.	<b>1 pregunta fue realizada:</b> - “¿No sería bueno poder ver la cantidad (de pacientes) en el mismo mapa?”	<b>1 pregunta fue realizada:</b> - “¿Por qué las enfermedades no se encuentran en los filtros?”	<b>1 pregunta fue realizada:</b> - “¿Dónde veo la información de Guasmo de esta enfermedad?”
	Completada con éxito.	<b>4 de 4 Completaron con éxito la tarea.</b>	<b>3 de 4 Completaron con éxito la tarea.</b>	<b>4 de 4 Completaron con éxito la tarea.</b>
	Observaciones	Los sujetos no pudieron identificar en su primera vez claramente como ver el número total de pacientes. Esto debido a que, al dar clic en el mapa los sujetos esperaban ver los valores sobre el mapa. Los sujetos comentaron que sería imprescindible presentar el número de pacientes asociados a cada sector de manera mas explícita. Además, los sujetos solicitaron que la plataforma informe de manera clara cuando no existan datos disponibles de alguna enfermedad en particular. Esto debido a que, al buscar una enfermedad para la cual no existen pacientes, la interfaz no muestra información en el mapa, ni tampoco muestra algún aviso claro de que no existen datos para la enfermedad buscada. Los sujetos pensaron que era un error de la plataforma. Finalmente, 10 de los 12 sujetos presentaron dificultad para encontrar la sección de “búsqueda de enfermedades”. Los sujetos primero realizaban la búsqueda de este filtro dentro de “Más Filtros”. Sin embargo, 11 de los 12 sujetos completaron la tarea con éxito. El único caso fallido, se dio debido a que el epidemiólogo no logró encontrar la sección de “búsqueda de enfermedades”. Cabe destacar que los epidemiólogos tomaron un tiempo aumentado con respecto a los otros dos grupos debido a que ellos analizaron el MEG con mayor detalle cuando pudieron empezar a utilizarlo. Es decir, se desviaban de su tarea objetivo, para intentar explorar todas las funcionalidades del MEG.		

<b>TAREA 2</b>	Tiempo tomado en ejecutar la tarea medido en segundos.	137.5 en promedio	154.33 en promedio	125,33 en promedio
	Número de clics dados para realizar la tarea.	8 en promedio	14 en promedio	7 en promedio
	Preguntas realizadas durante la ejecución de la tarea.	<b>Ninguna pregunta fue realizada</b>	<b>1 pregunta fue realizada:</b> - “¿Las zonas de interés no se generan automáticamente según la enfermedad que estoy viendo?”	<b>Ninguna pregunta fue realizada</b>
	Completada con éxito	<b>4 de 4 Completaron con éxito la tarea</b>	<b>2 de 3 Completaron con éxito la tarea</b>	<b>3 de 4 Completaron con éxito la tarea</b>
	Observaciones	<p>Gran cantidad de los sujetos tuvo dificultades para encontrar y utilizar el acercamiento (zoom) adecuadamente para ver un área específica. Los sujetos esperaban hacer clic en la zona y que el mapa hiciera el acercamiento por ellos.</p> <p>Además, al seleccionar los sectores para la creación de la zona de interés, no fue claro como identificar rápidamente cuántas áreas los sujetos ya habían agregado a su zona de interés. Esto debido a que el mapa no da una retroalimentación clara sobre los sectores seleccionados.</p> <p>Finalmente, el botón de agregar más geo zonas tapaba el icono para visualizar la última geo zona de la lista, por la cantidad de geo zonas ya creadas.</p> <p>Nota: Uno de los epidemiólogos no accedió a la ejecución de esta tarea por falta de tiempo.</p>		

**Tabla 3.7. Resultados de las pruebas de usabilidad. Parte 2: Entrevista de discusión acerca del mapa de salud.**

<b>PARTE 2 ENTREVISTAS</b>	<b>SUJETOS DE PRUEBA</b>		
	<b>DOCTORES (4)</b>	<b>EPIDEMIÓLOGOS (4)</b>	<b>INVESTIGADORES (4)</b>
<b>¿Qué parte del mapa le agradó en mayor manera?</b>	“La concentración de la información. La buena distribución de colores. El rápido acceso a la información. La información que es descargable”.	“Los filtros y la creación de áreas de interés. La iniciativa. Todo fue agradable. La búsqueda de enfermedades”.	“Información completa y bien estructurada. Una excelente iniciativa”.
<b>¿Qué cambiarías del mapa de salud?</b>	“Más colores por enfermedades”.	“Cambiaría los colores por unos menos intensos.”	“Poder seleccionar las edades por

		Cambiaría los colores por unos que se pueda diferenciar mejor las intensidades. Haría más clara la leyenda del mapa. Y la pondría encima del mapa”.	rangos personalizados. Así como el grupo etario. Cambiaría los filtros, pondría el filtro de enfermedades y el filtro de edades en la barra superior”.
<b>¿Añadiría alguna funcionalidad extra?</b>	“Rango de edades, con enfoque médico. Poder ver más de una enfermedad a la vez”.	“Grupo etario como filtros. Añadiría información de vacunas de pacientes. Medición de radiación. Y poder ver mi ubicación en el mapa. Diferentes colores por parroquias de la ciudad”.	“Rangos de edades más acorde a lo usado por los médicos. Un tutorial para conocer todo el sistema en su primer uso”.
<b>¿Como les gustaría visualizar los datos de clima y de tránsito dentro del mapa para que sea de utilidad?</b>	“El clima con algo lúdico, con personajes o con iconos como nubes. Con algún icono de color sobre el mapa. Iconos más claros, el texto no se ve tan representativo”.	“Es crucial. Se ve bien. Podríamos agregar para ver si hubo incidencia de la corriente del niño. La información meteorológica podría ser con símbolos más representativos y que se vea dentro de la ciudad”.	“Los datos de tránsito se pueden visualizar también como incidencias de accidentes de tránsito. Con respecto a colores, está bien como está”.
<b>Comentarios adicionales o recomendaciones</b>	“Rango de edades más acorde al uso médico, como adulto mayor, adolescente. Que el ministerio lo debería de aplicar para que tenga más datos. Conseguir más información de las enfermedades faltantes. Agregar casos de neumología o relacionados”.	“Es una herramienta muy útil, especialmente para la detección de enfermedades y la generación de hipótesis de zonas de riesgo. Sería interesante desarrollar dentro del mapa un sistema de sugerencias de acciones inteligente basado en la incidencia de enfermedades. Gran iniciativa, muy útil para epidemiólogos de todo el mundo. Sería bueno la implementación de un tutorial al entrar por primera vez. Recomendando tener reuniones periódicas con epidemiólogos para poder iterar nuevamente sobre el diseño ya realizado.”	“Implementación de un tutorial al entrar por primera vez al sistema. Es importante poder visualizar las edades de los pacientes”.

**Tabla 3.8. Resultados de las pruebas de usabilidad. Parte 3: Encuesta de usabilidad del mapa de salud.**

<b>PARTE 3 ENCUESTA DE USABILIDAD</b> Medida en una escala del 1 al 5. Siendo: 5. Totalmente de acuerdo. 4. De acuerdo. 3. Ni de acuerdo ni en desacuerdo. 2. En desacuerdo. 1. Totalmente en desacuerdo.	<b>SUJETOS DE PRUEBA</b>				
	<b>DOCTORES (4)</b>	<b>EPIDEMIÓLOGOS (4)</b>	<b>INVESTIGA- DORES (4)</b>	<b>Media (<math>\mu</math>)</b>	<b>D.E. (<math>\sigma</math>)</b>
<b>Encuentro la aplicación fácil de usar.</b>	4.75	4.75	4.25	4.58	0.51
<b>La información en la aplicación es clara y comprensible.</b>	4.5	4.75	4.25	4.50	0.67
<b>Aprender a usar la aplicación es fácil para mí.</b>	4.5	4.5	4	4.33	0.77
<b>La aplicación cuenta con características útiles.</b>	5	5	4.25	4.75	0.62
<b>La aplicación cuenta con características útiles para mi trabajo.</b>	4.25	4.75	4.5	4.50	0.67
<b>Realizar una búsqueda de enfermedades fue fácil para mí.</b>	4.5	4.25	3.5	4.08	0.99
<b>Crear una zona de interés fue fácil para mí.</b>	4	4.75	4	3.91	0.66
<b>Entender los colores del mapa y su significado fue fácil para mí.</b>	4.75	4.5	3.5	4.25	0.86
<b>En general, mi experiencia con la aplicación ha sido satisfactoria.</b>	5	4.5	4.25	4.58	0.66



Las pruebas de usabilidad dejaron importantes revelaciones hacia el futuro desarrollo del visualizador del MEG. Las pruebas constataron que el MEG es intuitivo y útil. Además, produjo entusiasmo en los sujetos de prueba, debido a que no existe al momento en Ecuador una herramienta similar.

Entre los puntos fuertes de usabilidad del MEG destaca el intuitivo mapeo de las funcionalidades principales del visualizador. Los sujetos de prueba mostraron un rápido entendimiento del significado de la intensidad de colores en el mapa y su relación con el número de pacientes que existe dentro de cada sector. Así como también, la rápida cognición de como los filtros, afectarían a los colores dentro del mapa. Esto se confirma en la encuesta de usabilidad (Tabla 3.8), en donde los sujetos calificaron la facilidad de uso del MEG con un 4.58 de 5 en promedio (siendo 5 totalmente de acuerdo y 1 totalmente en desacuerdo) y la claridad de la información presentada con un 4.50 de 5 (siendo 5 totalmente de acuerdo y 1 totalmente en desacuerdo). Así como también, calificaron con 4.25 de 5 (siendo 5 totalmente de acuerdo y 1 totalmente en desacuerdo) la facilidad para entender el significado de los colores dentro del mapa.

Sin embargo, el aumentado número de funcionalidades y características hace sentir al usuario abrumado al momento de utilizar el MEG por primera vez. Esto se vio reflejado en la dificultad de la mayoría de los sujetos al momento de encontrar en donde se realiza la búsqueda de enfermedades. A partir de las observaciones se pudo determinar que, aparentemente el mapeo visual llama a localizar la búsqueda de enfermedades dentro de los filtros. Este fue el punto de dolor más evidente del MEG. Además, se pudo observar que la visualización de la información detallada de los sectores, la cual se encuentra dentro del contenedor de la leyenda, pasó desapercibida del paneo visual de los sujetos de prueba. Así como la información meteorológica. Finalmente, ninguno de los sujetos de prueba determinó que el ícono en forma de semáforo activaba la capa de tránsito vehicular. Ver Tabla 3.7.

Otro punto de dolor del MEG se centra en la gama de colores utilizada para representar las incidencias de enfermedades dentro de cada sector. Como se puede observar en la Tabla 3.7, los epidemiólogos y doctores mencionaron este aspecto dentro de las características que cambiarían del MEG.

Con respecto a la realización de tareas (Tabla 3.6), 11 de los 12 sujetos de prueba completaron la primera tarea exitosamente (es decir, la búsqueda de enfermedades) y 9 de 11 sujetos de prueba completaron la segunda tarea exitosamente (es decir, crear una geo zona). El número óptimo de clics para completar la primera tarea es 4 clics (medido por los moderadores). A este número se acercaron en mayor manera los doctores (7.25 clics en promedio). Los epidemiólogos tuvieron un número promedio de clics más elevado (10.5 en promedio), y un tiempo de ejecución de la tarea superior al promedio de todos los sujetos de prueba. Esto se debe a que los epidemiólogos interactuaron con otras componentes del MEG durante la realización de la tarea. Esto no es sorpresa, ya que, como expresaron en la encuesta de usabilidad (Tabla 3.8), el MEG representa una herramienta de gran utilidad para su trabajo. Recibiendo 4.75 de 5 en promedio entre los 4 epidemiólogos (siendo 5 totalmente de acuerdo y 1 totalmente en desacuerdo). Por esta razón, los epidemiólogos deseaban explorar todas las funcionalidades del MEG desde un inicio.

Un comportamiento similar de los epidemiólogos ocurrió en la realización de la segunda tarea (es decir, la creación de una zona de interés). La creación de una zona de interés tuvo una tasa de éxito menor a la de la primera tarea, con un promedio de clics y un tiempo de realización de la tarea mayor en comparación con la primera tarea. Esto se debe a que la complejidad de crear una zona de interés es mayor a la de realizar una búsqueda de enfermedades. Los intentos no exitosos de la realización de esta tarea fueron debido a la falta de retroalimentación del mapa al momento de seleccionar los sectores dentro de este. El detalle de las observaciones obtenidas a partir de la realización de ambas tareas se encuentra en la Tabla 3.6.

Entre las recomendaciones dadas, destaca la presentación de un tutorial al momento de ingresar al visualizador por primera vez. De tal manera que los usuarios puedan estar conscientes de la totalidad de funcionalidades que ofrece el visualizador y en que lugar encontrarlas. Así como también la reubicación de los filtros, para que sea más intuitivo. Estas mejoras ayudarán a disminuir el número de clics dados para la realización de las tareas y reducirán el tiempo de ejecución de las mismas.

Los doctores pidieron considerar la posibilidad de poder customizar los colores del mapa, e introducir a través de una diferente tonalidad de color, una tercera variable geográfica. Además, solicitaron añadir el grupo etario a los filtros de pacientes. Así como también, hacer más flexible el filtro por edades. Es decir, poder customizar los rangos de edades de los pacientes.

Finalmente, la integración de datos de clima y tránsito fue esencial en la percepción de la utilidad del mapa para los epidemiólogos e investigadores, ya que reconocen que estos dos factores afectan en gran manera a la salud de la población de Guayaquil y son fuentes de investigación constante. Con respecto a la integración de los datos meteorológicos, los doctores y epidemiólogos recomendaron representarlos dentro del mapa de una forma mas explícita. Es decir, con algún tipo de ícono dibujado dentro de los sectores que represente el clima de este. Los comentarios de los sujetos de prueba sobre la integración de estas dos fuentes externas de datos se describen en detalle en la Tabla 3.7.

### **3.3 Anonimización de datos**

El algoritmo de geolocalización de dos capas propuesto en la subsección 2.2.1 logró geolocalizar el 97.53% de las 74,118 direcciones de pacientes del HLB de Guayaquil. Es decir, 72,294 pacientes anonimizados fueron agregados a la base de datos. El 2.47% de direcciones que no pudieron ser geolocalizados corresponden a direcciones de pacientes conteniendo información insuficiente para poder ser geolocalizadas, o direcciones correspondientes a las afueras de la ciudad de Guayaquil.

De estos 72,294 registros un 97.59% fue geolocalizado en la primera capa del algoritmo propuesto. Mientras que un 2.41% se geolocalizó utilizando la segunda capa de geolocalización (Google Maps Geocoding API). Tomando una muestra aleatoria de 100 registros de este 97.59% de direcciones, se encontró que la precisión (accuracy) de esta capa personalizada fue de un 92%.

El porcentaje de error, correspondiente al 8% de las direcciones geolocalizadas por la primera del algoritmo, se debe a casos en donde las direcciones representan intersecciones de calles dentro de la ciudad cuyos nombres son de origen desconocido. Así como también, direcciones correspondientes a

sectores homónimos. En este último caso, el algoritmo encuentra una ambigüedad semántica entre el nombre de dos lugares geográficamente diferentes que no puede resolver de manera correcta. Esto sucede en gran manera en las direcciones que son dadas como intersecciones de calles.

La Tabla 3.6 muestra en detalle el comportamiento de la primera capa de geolocalización ante diferentes tipos de direcciones, así como también casos de direcciones que no pudieron ser geolocalizadas.

**Tabla 3.9. Resultados de la metodología propuesta para la geolocalización de direcciones.**  
**Observación: La capa 2 del algoritmo no pasó por etapa de validación debido a la alta fiabilidad del método [47].**

Algoritmo de Geolocalización	Registros Procesados	Ejemplos	
		Correctos	Incorrectos
<b>TOTAL DE REGISTROS GEOLOCALIZADOS</b>	72,294 (97.53%)		
CAPA 1: Geolocalización por similaridad de texto	70,554 (97.59%) 92% correctamente geolocalizados de una muestra aleatoria de 100 registros.	<ul style="list-style-type: none"> <li>- SAUCES 3 MZ 537 V 8 ⇒ <b>Sector SAUCES</b></li> <li>- Guasm Sur ⇒ <b>Sector Guasmo Sur</b></li> <li>- Primave 2 ⇒ <b>Sector Primavera</b></li> <li>- 11 y Colombia ⇒ <b>Sector 9 de Octubre</b></li> </ul>	<ul style="list-style-type: none"> <li>- Martha de Roldos ⇒ <b>Ambigüedad sectores homónimos entre Martha de Roldos (avenida) y Cooperativa Martha de Roldos</b></li> <li>- Kennedy Mz 2 ⇒ <b>Ambigüedad sectores homónimos entre Kennedy Nueva y Kennedy Norte</b></li> <li>-17 Y LA I ⇒ <b>Ambigüedad de calles homónimas en el Norte y en el Sur de la ciudad.</b></li> </ul>
CAPA 2: Google Maps Directions API	1,740 (5.2%)		
<b>TOTAL DE REGISTROS NO GEOLOCALIZADOS</b>	1824 (2.41%)	<ul style="list-style-type: none"> <li>- ARGENTIONA Y LA 39AVA ⇒ <b>Error de Escritura</b></li> <li>- COOP. BALLADOLIT ⇒ <b>Error de Escritura</b></li> <li>- MZ.E30 ⇒ <b>Falta de información</b></li> <li>- MZ C14 ⇒ <b>Falta de Información</b></li> <li>- Bolivia ⇒ <b>Falta de Información</b></li> <li>- DURAN ⇒ <b>Sector en las afueras de Guayaquil</b></li> <li>- SALITRE ⇒ <b>Sector en las afueras de Guayaquil</b></li> <li>- BALZAR ⇒ <b>Sector en las afueras de Guayaquil</b></li> </ul>	

# CAPÍTULO 4

## 4 CONCLUSIONES Y RECOMENDACIONES

### 4.1 Conclusiones

Se ha definido exitosamente un esquema para almacenar y administrar datos de salud de una ciudad a diferentes niveles de granularidad geográfica. Este esquema soporta la agrupación de zonas de interés definidas por el usuario y el almacenamiento de las enfermedades en base al código CIE 10 universal y sus diversas agrupaciones.

Además de esto, el esquema propuesto tiene la capacidad de escalar hacia el almacenamiento de datos de diferentes instituciones y departamentos de salud de diversas ciudades. Tomando en consideración niveles de privacidad para cada una de las diferentes entidades (p. ej. instituciones, enfermedades, sectores) con el fin de salvaguardar el acceso a la información de carácter sensible según el criterio de las instituciones o autoridades.

Además, se ha desarrollado exitosamente un algoritmo para la geolocalización de direcciones almacenadas en los sistemas de información de las instituciones de salud. Logrando geolocalizar y anonimizar un 97.53% de las direcciones de pacientes disponibles utilizando la metodología de dos capas propuesta. Las direcciones geolocalizadas por la primera capa personalizada de geolocalización obtuvieron un 92% de precisión al momento de geolocalizar direcciones. Por lo que, con el desarrollo de este proyecto se han superado dificultades tales como la baja calidad de las direcciones ingresadas manualmente en los sistemas de información de las instituciones de salud. Además de, la falta de normalización de las direcciones dentro de una ciudad en vías de desarrollo (es decir, Guayaquil).

Las pruebas de usabilidad realizadas con médicos, epidemiólogos e investigadores revelaron que la nueva versión del MEG es útil (la utilidad del mapa fue calificada con 4.75 de 5 en promedio entre los 12 sujetos de prueba, siendo 5 totalmente de acuerdo y 1 totalmente en desacuerdo) y fácil de usar (la facilidad de uso fue calificada con 4.59 de 5 en promedio entre los 12 sujetos de prueba, siendo

5 totalmente de acuerdo y 1 totalmente en desacuerdo). 11 de 12 sujetos de prueba pudieron realizar una búsqueda de enfermedad dentro del mapa satisfactoriamente, mientras que 9 de 11 sujetos de prueba pudieron crear una zona de interés satisfactoriamente. Todos los sujetos de prueba comprendieron con facilidad el significado de la intensidad de los colores dentro del mapa. En general, la experiencia de los sujetos de prueba con el MEG fue satisfactoria. La experiencia con el mapa fue calificada con 4.58 de 5 en promedio entre los 12 sujetos de prueba, siendo 5 totalmente de acuerdo y 1 totalmente en desacuerdo.

Finalmente, con respecto a la versión anterior del MEG, se logró aumentar en un 26.13% el número de registros que se han podido geolocalizar y anonimizar. Además de, reducir en un 79.89% el tiempo de carga de la interfaz gráfica del visualizador del mapa de salud.

## **4.2 Limitaciones**

A pesar de que se ha podido geolocalizar la gran mayoría de las direcciones disponibles, sigue existiendo un porcentaje de direcciones que no pueden ser geolocalizadas. Esta es una gran limitación latente en este proyecto ya que existe una pérdida relevante de datos que no están siendo visualizados en el mapa de salud. Esto se debe a los registros de direcciones de pacientes conteniendo información insuficiente o errónea para poder ser geolocalizadas.

Otra limitante de este trabajo es el error introducido por el algoritmo de geolocalización propuesto. Este porcentaje de error corresponde a casos en donde las direcciones representan intersecciones de calles dentro de la ciudad cuyos nombres son de origen desconocido, y las direcciones corresponden a sectores homónimos. En este último caso, el algoritmo recae en una ambigüedad semántica al intentar discernir entre dos lugares con un mismo nombre en posiciones diferentes de la ciudad.

Otra de las limitaciones de este proyecto recae en el proceso manual que requiere la actualización de datos de pacientes dentro del sistema a través de la carga de un archivo. Es decir, no existe conexión directa con la institución de salud. Esto puede dar cabida al almacenamiento de información desactualizada con el pasar del tiempo.

Finalmente, se resalta la complejidad relativa del visualizador en su primer uso. Esto es una limitación esperada debido a la gran cantidad de características y funcionalidades que posee el visualizador.

#### **4.3 Recomendaciones y trabajo futuro**

El MEG desarrollado en este proyecto muestra el nivel de incidencia de enfermedades a una granularidad a nivel de sectores urbanos. Se recomienda que en trabajos futuros se implementen niveles geográficos más granulares, en donde se puedan identificar lugares geográficos que representen focos de enfermedades de manera más precisa. Esto permitiría, por ejemplo, realizar hipótesis de correlación entre una enfermedad altamente incidente dentro de una manzana de viviendas, y la existencia de una zanja de aguas servidas cercana a las viviendas.

También se recomienda conseguir acuerdos con diversas instituciones de salud. Con la finalidad de obtener una mayor variedad de fuentes de datos que alimenten el mapa epidemiológico de la ciudad. De esta manera, los resultados de las visualizaciones pueden llegar a ser estadísticamente relevantes con respecto a la población de la ciudad. De igual manera, se recomienda trabajar con las instituciones más a fondo para lograr la recolección de los datos de manera periódica, automática y segura. Esto se podría lograr mediante la instalación de un proceso activo (daemon) dentro de los servidores de la institución que realice la tarea de recolección de nuevos datos de forma automática.

Además, se incita a trabajos futuros a utilizar datos históricos de tránsito que vayan alineados con los filtros de fechas que utilice el usuario. Actualmente se utiliza una capa de tránsito en tiempo real. Esto, limita la validez de los análisis de correlación entre las incidencias de enfermedades y el nivel de congestionamiento vehicular en la ciudad. Ya que, solamente se está visualizando la información en tiempo real de esta variable, mas no un histórico.

Por último, se resalta la importancia de realizar actualizaciones periódicas a la interfaz gráfica del visualizador. De tal manera que, se puedan añadir funcionalidades según las necesidades de los usuarios. Entre las recomendaciones propuestas por los participantes de los estudios de usabilidad se encuentran: poder observar toda la información de salud en detalle a partir de un sector seleccionado.

Así como también, poder seleccionar más de una enfermedad al momento de realizar un filtrado. De tal manera que se pueda contrastar posibles factores que puedan correlacionar las enfermedades seleccionadas. Finalmente, los participantes recomendaron la implementación de un tutorial de bienvenida a los nuevos usuarios del sistema, como mecanismo para guiarlos en su primera experiencia con el visualizador.



# BIBLIOGRAFÍA

- [1] Harbeck, M; Seifert, L; Hänsch, S; Wagner, DM; Birdsell. "Yersinia pestis DNA from Skeletal Remains from the 6th Century AD Reveals Insights into Justinianic Plague" en PLOS Pathogens. 2013. doi: 10.1371
- [2] Redacción Médica (2017, Febrero 3). Síndrome de Laron: Ecuador tiene la mayoría de casos en el mundo [Online]. Disponible en: <https://www.redaccionmedica.ec/secciones/salud-publica/s-ndrome-de-laron-ecuador-tiene-la-mayor-a-de-casos-en-el-mundo-89570>
- [3] Francis Adams. The Genuine Works of Hippocrates, 1849.
- [4] Instituto Nacional de Estadística y Censos (INEC) (2017, Octubre 6). Guayaquil en cifras [Online]. Disponible en: <http://www.ecuadorencifras.gob.ec/guayaquil-en-cifras/>
- [5] Escuela Superior Politécnica del Litoral (2019, Julio). Guayaquil Health Map (MEG) [Online]. Disponible en: <https://gye-health-map.herokuapp.com/>
- [6] Ministerio de Salud Pública del Ecuador (2017). Información estadística y Geográfica de Salud de Tuberculosis y VIH [Online]. Disponible en: <https://www.salud.gob.ec/informacion-estadistica-de-produccion-de-salud/>
- [7] Ministerio de Salud Pública del Ecuador (2013). Boletín Epidemiológico del Dengue en Ecuador [Online]. Disponible en: <https://www.salud.gob.ec/boletin-epidemiologico-del-dengue-en-el-ecuador/>
- [8] Martens, W. J. "Health impacts of climate change and ozone depletion: an eco epidemiological modeling approach" en environmental health perspectives, 106(suppl 1), 1998, pp. 241-251.
- [9] Frank, L. D., & Engelke, P. "Multiple impacts of the built environment on public health: walkable places and the exposure to air pollution" en International regional science review, 28(2), 2005 pp. 193-216.
- [10] Yang, K., Sun, L. P., Huang, Y. X., Yang, G. J., Wu, F., Hang, D. R., ... & Zhou, X. N. "A real-time platform for monitoring schistosomiasis transmission supported by Google Earth and a web-based geographical information system" en Geospatial

health, 2012, pp. 195-203.

[11] Pigott, D. M., Golding, N., Mylne, A., Huang, Z., Henry, A. J., Weiss, D. J., ... & Bhatt, S. Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife*, 3, e04395, 2014.

[12] Noon, C. E., & Hankins, C. T. "Spatial data visualization in healthcare: supporting a facility location decision via GIS-based market analysis" en *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. 2001, pp. 10.

[13] Cinnamon, J., Rinner, C., Cusimano, M. D., Marshall, S., Bekele, T., Hernandez, T., ... & Chipman, M. L. "Evaluating web-based static, animated and interactive maps for injury prevention" en *Geospatial Health*, 2009, pp. 3-16.

[14] Parmanto, B., Paramita, M. V., Sugiantara, W., Pramana, G., Scotch, M., & Burke, D. S. "Spatial and multidimensional visualization of Indonesia's village health statistics" en *International journal of health geographics*, 7(1), 2008, pp. 30.

[15] Al-Rubeaan, K. A., Youssef, A. M., Subhani, S. N., Ahmad, N. A., Al-Sharqawi, A. H., & Ibrahim, H. M. "A Web-based interactive diabetes registry for health care management and planning in Saudi Arabia" en *Journal of medical Internet research*, 15(9), e202, 2013.

[16] Bell, B. S., Hoskins, R. E., Pickle, L. W., & Wartenberg, D. "Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public" en *International journal of health geographics*, 5(1), 2006, pp. 49.

[17] Kushniruk, A. W., & Patel, V. L. "Cognitive and usability engineering methods for the evaluation of clinical information systems" en *Journal of biomedical informatics*, 37(1), 2004, pp. 56-76.

[18] Slocum, T., Sluter, R., Kessler, F., & Yoder, S. "A qualitative evaluation of MapTime, a program for exploring spatiotemporal point data" en *Cartographica: The International Journal for Geographic Information and Geovisualization*, 39(3), 2004, pp. 43-68.

[19] Highfield, L., Arthasarnprasit, J., Ottenweller, C. A., & Dasprez, "A. Interactive web-based mapping: bridging technology and data for health" en *International journal of health geographics*, 10(1), 2011, pp 69.

- [20] Organización Mundial de Salud (OMS) [Online]. Disponible en: <https://www.who.int/es>
- [21] Organización Mundial de Salud (OMS). Ecuador [Online]. Disponible en: <https://www.who.int/countries/ecu/es/>
- [22] Boston Children's Hospital. (2006). HealthMap [Online]. Disponible en: <https://www.healthmap.org/en/>
- [23] Bernal-Acevedo, O., & Forero-Camacho, J. C. (2011). "Sistemas de información en el sector salud en Colombia". Gerencia y Políticas de Salud, 10(21), 2011. [Online]. Disponible en: <https://sig.sispro.gov.co/sigmsp/index.html>
- [24] Williams, G., Duré, I., Dursi, C., & Pereyra, J. L. "Médicos en Argentina: Red Federal de Registros de Profesionales de la Salud. REVISTA ARGENTINA DE MEDICINA", 3(6), 43-46, 2015. [Online] Disponible en: <https://sisa.msal.gov.ar/sisa/>
- [25] Universidad de Chile (2019). Centro Nacional en Sistemas de Información en Salud (CENS). Disponible en: <https://cens.cl/>
- [26] Ministerio de Salud Pública del Ecuador (2018). GeoSalud [Online]. Disponible en: <https://geosalud.msp.gob.ec/geovisualizador/>
- [27] Dayarian, I., & Savelsbergh, M. "Crowdshipping and same-day delivery: employing in-store customers to deliver online orders. Optimization Online". 2017. doi: 07.6142.
- [28] YOU, L., & PENG, Q. Research and implementation of geocoding based on Google Maps API [J]. Journal of Hubei University (Natural Science). 2010, pp. 161-164.
- [29] Momjian, B. PostgreSQL: introduction and concepts (Vol. 192) de New York: Addison-Wesley, 2001.
- [30] GraphQL (2015). graphql/dataloader [Online]. Disponible en: <https://github.com/graphql/dataloader>
- [31] Casalicchio, E., & Perciballi, V. "Measuring docker performance: What a mess!!!" en Proceedings of the 8th ACM/SPEC on International Conference on Performance Engineering Companion, 2017, pp. 11-16.

- [32] Docker (2014). docker/compose [Online], Disponible en: <https://github.com/docker/compose>.
- [33] Chi, X., Liu, B., Niu, Q., & Wu, Q. "Web load balance and cache optimization design based Nginx under high-concurrency environment" en Third International Conference on Digital Manufacturing & Automation, 2012, pp. 1029-1032.
- [34] Fedosejev, A. (2015). React.js Essentials. Packt Publishing Ltd, 2015.
- [35] Abramov, D. (2016). Redux: predictable state container for javascript apps [Online]. Disponible en: <https://redux.js.org>
- [36] Secretaría nacional de Planificación Y Desarrollo (2015) [Online]. Agenda Zonal. Zona 8 Guayaquil – Samborondón - Duran. pp. 130 - 131. Disponible en: <https://www.planificacion.gob.ec/wp-content/uploads/downloads/2015/11/Agenda-zona-8.pdf>.
- [37] Ministerio de Educación (2019). Distritos Educativos Zona 8. Disponible en: <https://educacion.gob.ec/distritos-educativos-zona-8/>
- [38] Ministerio del Interior (2013). Nuevo ordenamiento territorial le cambia la cara a las comisarías urbanas en Guayaquil [Online]. Disponible en: <https://www.ministeriodegobierno.gob.ec/nuevo-ordenamiento-territorial-le-cambia-la-cara-a-las-comisarias-urbanas-en-guayaquil/>
- [39] Alcaldía de Guayaquil (2019). División de la ciudad [Online]. Disponible en: <https://guayaquil.gob.ec/divisi%C3%B3n>
- [40] Municipalidad de Guayaquil (2019). División de Barrios de Guayaquil [Online]. Disponible en: <https://guayaquil.gob.ec/divisi%C3%B3n>
- [41] Yujian, L., & Bo, L. A normalized Levenshtein distance metric. IEEE transactions on pattern analysis and machine intelligence, 29(6), 2007, pp. 1091-1095.
- [42] Su, Z., Ahn, B. R., Eom, K. Y., Kang, M. K., Kim, J. P., & Kim, M. K. "Plagiarism detection using the Levenshtein distance and Smith-Waterman algorithm" en 3rd International Conference on Innovative Computing Information and Control. 2008, pp. 569-569.
- [43] Heeringa, W. J. "Measuring dialect pronunciation differences using Levenshtein

distance”, Tesis de Doctorado, University Library Groningen. 2004.

[44] Shang, J., Zheng, Y., Tong, W., Chang, E., & Yu, Y. “Inferring gas consumption and pollution emission of vehicles throughout a city” en Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. 2014, pp. 1027-1036.

[45] Nevedrov, D. “Using JMeter to Performance Test Web Services” en dev2dev. 2016.

[46] Open Knowledge International(2011). CKAN [Online]. Disponible en:  
<https://docs.ckan.org>

[47] Roongpiboonsopit, D., & Karimi, H. A. “Comparative evaluation and analysis of online geocoding services” en International Journal of Geographical Information Science, 24(7), 2010, 1081-1100.

# APÉNDICES

## Apéndice A: Consentimiento informado

### **Consentimiento Informado para Participantes de Investigación**

El propósito de este documento de consentimiento es proveer a los participantes en esta investigación una clara explicación sobre el objetivo de esta.

La presente investigación es conducida por estudiantes de la Escuela Superior Politécnica del Litoral que desarrollan una aplicación web para la visualización de datos epidemiológicos de Guayaquil en un mapa de salud bajo la dirección de la Dra. Lorena Carlo Unda. La meta de este estudio es recolectar parámetros, sugerencias y recomendaciones que ayuden a definir el nivel de usabilidad (facilidad de uso, utilidad) que posee nuestra aplicación con el objetivo de mejorarla a futuro.

Si usted accede a participar en este estudio, se le pedirá responder preguntas en una encuesta al inicio y al final de la evaluación de la aplicación. Lo que conversemos durante esta sesión se grabará, de modo que el investigador pueda transcribir después las ideas que usted haya expresado. Además, se grabará toda la actividad que se realice dentro del computador con el fin de medir el nivel de usabilidad de la aplicación. Esta grabación de actividad incluye la captura de: movimientos del ratón y clics.

La participación en este estudio es estrictamente voluntaria. La información que se recoja será confidencial y no se usará para ningún otro propósito fuera de los de esta investigación.

Si tiene alguna duda sobre este proyecto, puede hacer preguntas en cualquier momento durante su participación en él. Igualmente, puede retirarse del proyecto en cualquier momento sin que eso lo perjudique en ninguna forma. Si alguna de las preguntas durante la entrevista le parece incómodas, tiene usted el derecho de hacérselo saber al investigador o de no responderlas.

Desde ya le agradecemos su participación.

Acepto participar voluntariamente en esta investigación, conducida por estudiantes de la Escuela Superior Politécnica del Litoral que desarrollan una aplicación web para la visualización de datos epidemiológicos de Guayaquil en un mapa de salud bajo la dirección de la Dra. Lorena Carlo Unda.

He sido informado (a) de que la meta de este estudio es recolectar parámetros, sugerencias y recomendaciones que ayuden a definir el nivel de usabilidad que posee la aplicación. Me han indicado también que tendré que responder una encuesta.

Reconozco que la información que yo provea en el curso de esta investigación es estrictamente confidencial y no será usada para ningún otro propósito fuera de los de este estudio sin mi consentimiento. He sido informado de que puedo hacer preguntas sobre el proyecto en cualquier momento y que puedo retirarme del mismo cuando así lo decida, sin que esto acarree perjuicio alguno para mi persona. De tener preguntas sobre mi participación en este estudio, puedo contactar a la Dra. Lorena Carlo Unda, directora del proyecto, al correo electrónico: lcarlo@espol.edu.ec.

<p><b>Permiso para grabar la entrevista:</b>  Acepto que se use una grabadora de voz durante la evaluación de la aplicación:</p> <p>Si _____ No _____</p>	<p><b>Permiso para grabar actividad en la PC:</b>  Acepto que se graben los movimientos del mouse y los clics dados durante la evaluación de la aplicación:</p> <p>Si _____ No _____</p>
<p>Nombre del Participante:</p> <p>_____</p> <p>Firma:</p> <p>Fecha:</p> <p>_____</p>	<p>Nombre de la persona que obtiene el consentimiento:</p> <p>_____</p> <p>Firma:</p> <p>Fecha:</p> <p>_____</p>

## Apéndice B: Instrumento de usabilidad

### ANTECEDENTES. -

Se utilizarán computadoras propias para conducir las pruebas de usabilidad. En las cuales se podrá monitorear todas las actividades de los sujetos de prueba mediante la grabación de la pantalla y audio. **Toda la sesión será grabada mediante una grabadora de pantalla y una grabadora de audio.**

### PRIMERO. - INTRODUCCIÓN (1 minuto)

Se realiza una introducción del proyecto hacia los sujetos de prueba. Se menciona brevemente cuales son las funcionalidades del mapa de salud desarrollado, su objetivo principal y su alcance. En todo momento visualizan una captura de la pantalla principal del visualizador. No se realizan inferencias sobre que realiza cada componente.

*En este momento se realiza la firma del consentimiento informado.*

### SEGUNDO. - PRUEBA DE USABILIDAD (12 minutos)

Se empieza con una serie de preguntas que nos ayudarán conociendo los antecedentes de los sujetos de prueba.

1. ¿Cuál es su profesión?
2. ¿Cuántos años de experiencia posee en su área?
3. ¿Ha realizado o realiza actualmente actividades de investigación?
4. En caso de haber respondido sí a la pregunta anterior, mencione brevemente su área de investigación.
5. ¿Cuán ágil se considera navegando por el internet? Siendo: 5. Muy ágil, 4. Ágil, 3. Neutral, 2. Poco Ágil, 1. Nada Ágil.
6. ¿Ha interactuado con mapas de salud digitales previamente?

Se les pide a los sujetos de prueba ejecutar dos tareas en concreto. Se proporciona un tiempo máximo de 5 minutos por cada tarea. **Las variables serán medidas una vez que terminen las pruebas a través de las grabaciones.**

TAREAS	VARIABLES MEDIDAS
1. Realizar una búsqueda de la siguiente enfermedad: "Diarrea y Gastroenteritis de Presunto Origen Infeccioso". Luego, visualizar e identificar las zonas con un numero de pacientes más alto de esta enfermedad.	<ul style="list-style-type: none"> <li>- Tiempo tomado en ejecutar la tarea (Medido en segundos)</li> <li>- Número de clics dados para realizar la tarea.</li> <li>- Preguntas realizadas durante la ejecución de la tarea.</li> <li>- Pudo completar la tarea (SI O NO)</li> </ul>
2. Crear una zona de interés. (Para esto, se da un breve contexto de lo que representa una zona de interés).	<ul style="list-style-type: none"> <li>- Tiempo tomado en ejecutar la tarea (Medido en segundos)</li> <li>- Número de clics dados para realizar la tarea.</li> <li>- Preguntas realizadas durante la ejecución de la tarea.</li> <li>- Pudo completar la tarea (SI O NO)</li> </ul>

### TERCERO. – ENTREVISTA (5 minutos)

Finalmente, se realiza una entrevista, seguida de una encuesta con respecto al mapa de salud a los sujetos de prueba. En las preguntas objetivas las respuestas se miden en una escala del 1 al 5,



siendo: 5. Totalmente de acuerdo, 4. De acuerdo, 3. Ni de acuerdo ni en desacuerdo, 2. En desacuerdo, 1. Totalmente en desacuerdo.

**Entrevista:**

- |   |
|---|
| 1. ¿Qué parte del mapa le agradó en mayor manera?   |
| 2. ¿Qué cambiarías del mapa de salud?   |
| 3. ¿Añadiría alguna funcionalidad extra?  |
| 4. ¿De qué manera le gustaría visualizar los datos de clima y de tránsito dentro del mapa para que sea de utilidad? |
| 4. Comentarios adicionales o recomendaciones.   |

**Encuesta de usabilidad:**

- |  |
|--|
| 1. Encuentro la aplicación fácil de usar.                              |
| 2. La información en la aplicación es clara y comprensible.            |
| 3. Aprender a usar la aplicación es fácil para mí.                     |
| 4. La aplicación cuenta con características útiles                     |
| 5. La aplicación cuenta con características útiles para mi trabajo     |
| 6. En general, mi experiencia con la aplicación ha sido satisfactoria. |

**CUARTO. - FINALIZACIÓN (1 minuto)**

Se da por finalizada las pruebas de usabilidad y se agradece a los asistentes.

**DURACIÓN TOTAL APROXIMADA: 15 - 17 minutos**

**QUINTO. - POST-OBSERVACIÓN**

Se realiza una reunión del grupo que condujo las pruebas para discutir los resultados de cada individuo y se responde a las siguientes conclusiones:

- |                                    |
|------------------------------------|
| 13. Puntos de dolor de usabilidad. |
| 14. Puntos fuertes de usabilidad.  |