# BASES DE DATOS GEO-TEMPORALES Y SISTEMAS DE INFORMACIÓN GEOGRÁFICA PARA EL ALMACENAMIENTO Y EXPLOTACIÓN DE DATOS PARA EL MONITOREO DE POSIBLES FOCOS DE INFECCIÓN DE ENFERMEDADES TRANSMITIDAS POR VECTORES BIOLÓGICOS

Enoc L. García V. – Licenciatura en Inteligencia Artificial
Profesor: Ana Darcacha
Fecha de entrega: 19/06/2024

***Abstract*— Las enfermedades transmitidas por vectores biológicos son una problemática global que pone en riesgo potencial de infección a 2.5 billones de personas cada año. En la Argentina y en los países limítrofes casos confirmados de Dengue llegaron a máximos históricos en el año 2024 confirmando que las enfermedades transmitidas por vectores como el Dengue son una problemática local y significativo en la actualidad. En este contexto, se presenta necesario generar la mayor cantidad de datos relevantes que puedan ser procesados y utilizados por los actores correspondientes para tomar acciones preventivas y correctivas sobre los focos de infección tomando en cuenta que el control en la proliferación de vectores es la estrategia principal para la prevención de este tipo de enfermedades. Los sistemas de información geográficos y bases de datos para series temporales con capacidad de consulta para datos geográficos se muestran como herramientas con el potencial de dar el soporte necesario para el análisis de los datos generados sobre casos de infección.**

***Término Índice*—Vectores biológicos, bases de datos de series temporales, bases de datos geográficas, sistemas de información geográficos.**

## I. Introducción

Las enfermedades transmitidas por vectores biológicos representan una problemática creciente a nivel global donde, con el pasar de los años, cada vez más personas corren el riesgo de contraer este tipo de afecciones. A pesar de los avances en la medicina, el método más efectivo para la radicación de este tipo de enfermedades sigue siendo la eliminación del vector biológico, por lo que el tiempo de respuesta de las instituciones -sean públicas o privadas- al actuar sobre focos de infección es crítico para evitar la proliferación del vector y en consecuencia, la transmisión de la enfermedad.

En este sentido, se presenta necesario recolectar la mayor cantidad de datos, que luego de ser procesados, puedan proveer información útil y oportuna para la toma de decisiones. En este contexto, se presenta la utilización de bases de datos geo-temporales y los sistemas de información geográfica como una alternativa para el almacenamiento y explotación de datos con estas características.

## II. Justificación

En esta sección se discuten los desafíos asociados a las enfermedades transmitidas por vectores biológicos así como problemáticas asociadas al almacenamiento y procesamiento de información con detalle geo-temporal.

### *A. Desafíos asociados a las enfermedades de transmisión por vectores biológicos*

Cada año más de un billón de personas son infectadas y más de un millón de personas mueren por enfermedades transmitidas por vectores biológicos incluyendo malaria, dengue, esquistosomiasis, leishmaniasis, mal de Chagas, fiebre amarilla, filariasis linfática y oncocercosis [1]. Se estima que más de la mitad de la población global está en riesgo de contraer alguna de estas enfermedades.

A pesar de los avances en la medicina, el control de vectores, históricamente y en la actualidad, se presenta como un método altamente efectivo para el control de transmisión de este tipo de enfermedades, pero la complacencia y la falta de inversión económica ha resultado en el resurgimiento de estas enfermedades [2].

En los últimos años se han visto aumentos de casos reportados de Dengue a nivel nacional en Argentina (la enfermedad transmitida por vector con el crecimiento más

rápido global, con más de 2.5 billones de personas en riesgo de contraer la enfermedad), y en el año 2024 se obtuvo el récord histórico de casos confirmados de Dengue. Además, en otros países de la región como Brasil y Uruguay se registraron simultáneamente brotes importantes de la misma enfermedad [3][4][5].

### *B. Desafíos asociados a la recolección de datos sobre la propagación de enfermedades de transmisión por vectores biológicos*

El desafío más significativo en el modelaje de riesgo de enfermedades transmitidas por vectores biológicos es la naturaleza diversa, la calidad y la accesibilidad a set de datos observacionales, particularmente en países de bajos ingresos los cuales tienen la mayor carga de enfermedades transmitidas por vectores biológicos [6].

Si bien en la Argentina por ley (Ley˚ N 15465), es obligatoria para los prestadores de servicios médicos la notificación de enfermedades como Dengue, Zika, Fiebre Amarilla, Chikungunya, Mal de Chagas y Leishmaniasis (todas enfermedades transmitidas por vectores biológicos), globalmente están sub-reportados [7][8].

Además, diferentes aspectos como la supervivencia, desarrollo y maduración de los vectores en la dinámica de transmisión del dengue (y de otras enfermedades que se transmiten por vector) dependen de las condiciones climatológicas [9], lo que se traduce en set de datos espacio-temporales complejos.

Otro problema asociado a la recolección y el análisis de estos datos está en considerar la dirección del paciente como el lugar de riesgo o contagio asume una inmovilidad de los individuos e ignora la posibilidad de contagio en otros lugares como la dirección de trabajo u otros lugares de ocio/entretenimiento [10].

### *C. Desafíos asociados a los sistemas de información geográfica y a los datos espacio-temporales*

Las bases de datos espaciales convencionales soportan un manejo eficiente de objetos espaciales, sin embargo, manejan solamente los datos para un tiempo específico dado, por lo que les es muy difícil manejar de manera eficiente objetos con información geográfica que varía con el tiempo [9]. Por ese motivo, se han propuesto distintos modelos espacio-temporales y lenguajes de consulta para hacer un uso efectivo de ese tipo de datos, a pesar de eso, la complejidad inherente a estos tipos de datos ha resultado en pocos sistemas completos que soporten eficientemente los cambios de datos espaciales que ocurren con el tiempo [11].

En lo que respecta al desarrollo de las bases de datos geo-espacio-temporales existen múltiples desafíos que se traducen en la cobertura que pueda ofrecerse a datos espaciales y datos no espaciales que se modifican con el tiempo, capacidades de consultas (algoritmos de uniones, algebra de consulta, optimizaciones, etc.), la no practicidad de funcionalidades y dificultades de implementación de los modelos [12].

## III. Marco Teórico

En esta sección, se tratarán conceptos claves relacionados con la transmisión de enfermedades por vectores, las bases de datos geo-temporales y los sistemas de información geográficos (SIG); así como las ventajas de estos sistemas para la exposición y análisis de datos con información geográfica.

### *A. Enfermedades transmitidas por vectores biológicos*

Las enfermedades transmitidas por vectores son todas aquellas enfermedades cuya transmisión está asociada a un vector biológico, es decir, a un organismo vivo que transmite patógenos infecciosos entre humanos y desde animales a humanos. Los vectores usualmente son invertebrados, comúnmente insectos y la mayoría de las enfermedades transmitidos por vector son transmitidos por mosquitos y garrapatas. Generalmente, una vez que un vector se hace infeccioso, es capaz de transmitir el patógeno por el resto de su vida con cada mordida o ingesta de sangre [13].

Es una realidad que enfermedades como el Dengue, Chikungunya y Zika están presentes en Argentina y todas son transmitidas por el mosquito Aedes. Estas enfermedades presentan algunos síntomas comunes como presencia de fiebre, dolores musculares, dolores de cabeza, entre otros y en todos los casos o no existen vacunas disponibles, o son pocas las opciones y con efectividad limitada.

En general, las opciones de tratamiento de estas enfermedades son limitados, por lo que la prevención es clave. El riesgo de transmisión de enfermedades transmitidas por mosquitos es más alto entre el anochecer y el amanecer (a excepción del mosquito Aedes transmisor del dengue, Chikungunya, zika y fiebre amarilla que es más activo durante el día) y se ha probado como estrategias efectivas el uso de insecticidas, repelentes y barreras mecánicas para evitar las mordidas de estos insectos [14].

Por otro lado, el cambio climático representa un aumento en el riesgo potencial de transmisión de estas enfermedades, ya que la reproducción y desarrollo de los vectores está altamente influenciado por las condiciones climáticas. En las regiones tropicales y subtropicales existen condiciones favorables para el desarrollo de vectores transmisores de otras enfermedades

como la malaria y la esquistosomiasis la mayor parte del año, pero a medida que se acentúan los efectos del cambio climático (aumento global de la temperatura y cambio en los ciclos de lluvias) se puede esperar una expansión de las áreas geográficas susceptibles a la transmisión de estas enfermedades [15].

Debido a que la acción de los vectores está acotado a un espacio geográfico limitado, la estrategia principal para el control de la transmisión de estas enfermedades es el control sobre vector e históricamente se ha presentado como una de las estrategias más efectivas. En ese sentido, se hace vital el análisis geográfico y temporal de los focos de infección para que las campañas de control de vectores sean efectivas.

### *B. Sistemas de información geográfica*

En términos generales, los sistemas de información geográfica (SIG) son una clase especial de sistemas informáticos que no solo llevan el rastro de eventos, cosas y actividades sino también su ubicación. Los sistemas de información geográfica son reconocidos como herramientas analíticas y de toma de decisiones en muchos campos distintos. Es común su uso en administración de la tierra, mapeo estadístico, transporte, redes y administración ambiental, imágenes de sensores remotos, administración de agua y desperdicios, mantenimiento y administración de alumbrado público, planeamiento urbano y rural, planeamiento de turismo, planeamiento de servicios de salud y administración de seguridad y crimen [16].

En los sistemas de información geográfica, el mundo se representa a través de archivos raster (donde cada pixel se utiliza para representar el valor de una variable en distintas ubicaciones), o a través de modelos de vectores (donde la descripción geográfica se hace a través de puntos, líneas y polígonos), generando capas que describen distintas variables o características y que se acumulan para hacer una representación más completa del objeto que modelan [17].

En la actualidad existen muchas opciones distintas de sistemas de información geográfica. Existen software licenciados con larga trayectoria como ArcGIS propiedad de la empresa norteamericana ESRI con capacidades avanzadas para la generación de mapas, trabajo colaborativo, visualización y análisis de datos y servicios en la nube [18]. El principal sistema de información geográfica gratuito es QGIS, un software multiplataforma de código abierto y licenciado bajo GNU también disponible en aplicación de escritorio, online y móvil [19]. Otras opciones disponibles en el mercado de código abierto son GRASS GIS [20] y gvSIG [21]. GRASS GIS siendo un miembro fundador de la Fundación Geoespacial de código abierto (OSGeo Project en inglés) y gvSIG un proyecto originalmente impulsado por el gobierno regional de la Comunidad Valenciana pero que luego pasó a ser sostenido por la asociación gvSIG compuesta tanto por empresas como por instituciones sin fines de lucro y educativas.

En muchas áreas del conocimiento, las investigaciones con datos geográficos se hacen en términos de espacios estáticos ignorando la riqueza que se obtiene de evaluar la dimensión temporal y la movilidad humana. Muchos fenómenos no pueden ser entendidos en su totalidad sin tomar en cuenta la movilidad como un elemento esencial en los modelos espacio-temporales [22]. Es por esto que los sistemas de información geográfica en tiempo real y las bases de datos espacio-temporales se muestran como potentes herramientas para el análisis y visualización de estos modelos.

Desde los años 80 se viene trabajando en sistemas de información geográfica en tiempo-real y en la década de los 90 ya existían sistemas comerciales con funcionalidades para el análisis en tiempo real de datos espacio-temporales en rubros gubernamentales, de negocios y educativos. Los sistemas geográficos de tiempo real abren nuevas posibilidades para interactuar, representar y tomar decisiones sobre el mundo que nos rodea. A pequeña escala permiten aproximar como los individuos interactúan con su ambiente en su vida diaria y en gran escala permite entender y monitorear dinámicas poblacionales complejas en tiempo real así como lo fue con la evacuación de los afectados por el huracán Katrina o las migraciones de refugiados en zonas de conflicto en África y Medio Oriente [23].

Si bien existen avances en el desarrollo de sistemas de información geográfica de tiempo real, el manejo de datos espacio-temporales sigue representando un desafío.

### *C. Bases de datos espacio-temporales*

Las bases de datos espacio-temporales trabajan con aplicaciones donde los tipos de datos se caracterizan por semánticas espaciales y temporales simultáneamente, sin embargo, el tratamiento de este tipo de datos no es tan directa debido a la complejidad de las estructuras de datos sumado a la dificultad de representar y manipular de los datos involucrados. Las bases de datos relacionales tradicionales no son eficientes en el manejo de este tipo de datos multi-dimensionales con estructuras y comportamientos complejos [24]. En el mercado actual existen soluciones especializadas en el manejo de datos espacio-temporales como InfluxDB [25] y Kinetica [26] y extensiones sobre bases de datos relacionales existentes como el caso de TerraLib [27] y Geomesa [28] para trabajar con datos espacio-temporales.

Los modelos de datos espacio-temporales buscan modelar, consultar y grabar la evolución temporal de hechos bajo diferentes ópticas del tiempo como tiempo válido, tiempo de transacción o tiempo del evento, etc. Los modelos de datos espacio-temporales buscan ampliar el conocimiento guardado en las bases de datos sobre hechos actuales y hechos pasados [29].

## IV. Estado del arte

En esta sección se destacarán otras investigaciones en la misma área de conocimiento y que definen un fundamento o un estado del arte para la presente investigación.

### *A. Uso de sistemas colaborativos para la recolección de datos*

En Managua, Nicaragua, entre los años 2014 y 2016, se llevó a cabo una prueba piloto de "DengueChat", una plataforma que promovía las intervenciones de la comunidad en el control del vector del dengue [30]. Esta plataforma impulsaba la participación de la comunidad en la recolección de evidencias, reportes y análisis e incorporaba contenido pedagógico, información clave y dinámicas de juegos para motivar a las comunidades a aplicar estrategias de reducción del vector biológico. La idea fundamental detrás de la aplicación de software era motivar a las comunidades a identificar y eliminar los criaderos de mosquitos.

Las premisas en esta investigación fueron que: reducir los índices entomológicos a través del compromiso sostenido de la comunidad se traduciría en una reducción significativa del riesgo de Dengue en la comunidad, por otro lado, las comunidades afectadas por Dengue eran la mejor fuente de información sobre posibles (y reales) lugares de proliferación de mosquitos y por lo tanto, las comunidades eran las indicadas para llevar a cabo estrategias de eliminación de vectores, y por último, el uso de software junto a una organización de persona a persona fomentaría la movilización de la comunidad a eliminar los mosquitos.

La conclusión que se obtuvo luego de los diecinueve (19) meses de la prueba piloto en Managua fue que la combinación del software social y software movilizó a los residentes a producir datos generados comunitariamente que son científicamente válidos, abundantes, con profundidad históricos y cruciales tanto para las políticas de prevención como para políticas de predictivas en el control de vectores. La prueba mostró que durante el tiempo que estuvo activa, los índices etimológicos (el índice que media la presencia del insecto) del mosquito Aedes disminuyeron y por lo tanto el riesgo de contagio de Dengue disminuyó, mientras que en las comunidades de control se mantuvieron altos.

El principal desafío hacia el futuro que tenía la aplicación era su efectividad en comunidades más grandes. En caso de que la aplicación no fuese capaz de movilizar y mantener comprometida a comunidades más grandes con costos aceptables, entonces quedaban limitados los méritos científicos y su importancia política. Por el contrario, si se mostraba efectiva en poblaciones considerablemente más grandes a costos aceptables, entonces su importancia en la prevención de enfermedades, investigación científica y acción civil es más que considerable.

### *B. Uso de sistemas clásicos de información geográfica para el análisis estático de casos de Dengue*

Existen investigaciones previas donde se utilizaron sistemas de información geográfica para analizar y representar los focos de infección en brotes de Dengue tanto en Argentina como en otros países de la región.

En el año 2011, en el marco de un convenio entre CONAE (Comisión Nacional de Actividades Espaciales) y el Ministerio de Salud de la Nación (Argentina) se desarrolló un proyecto para implementar un sistema informático de alerta temprana de Dengue. En este contexto, se desarrollaron algoritmos para cálculo del riesgo ambiental de Dengue basado en factores ambientales [31].

Estos algoritmos se basaron en una combinación de dos componentes, en primer lugar, un componente estático relacionado con las condiciones ambientales históricas que determinan la probabilidad de que el vector se desarrolle y en segundo lugar, un componente dinámico relacionado con las condiciones actuales que posibiliten la transmisión viral. El resultado es una estratificación de localidades de Argentina en función de su riesgo ambiental de Dengue disponible a través de un visualizador web del sistema de información geográfica basada en las tecnologías Geomajas y GeoServer.

En el año 2017 se publicó el reporte analítico que un grupo de investigadores realizaron en el año 2016 durante el brote histórico de Dengue en la ciudad de Córdoba (Argentina) [32].

Para realizar este reporte, los investigadores se basaron en datos oficiales del sistema de monitoreo de Dengue publicados entre inicios de enero y mediados de mayo cubriendo así 19 semanas. Los datos se analizaron de manera global usando el test de Knox, un test estadístico para comprobar la existencia de "clusters" en un contexto espacio-temporal. En este sentido el test de Knox se utilizó para encontrar los "clusters" que corresponderían con las zonas de alta incidencia o alto riesgo de infección epidemiológica, que luego serían los principales focos de las estrategias de fumigación para la eliminación del vector biológico. Es importante destacar que la ubicación geográfica de los casos utilizados fue su ubicación de residencia, por lo que las zonas de alta incidencia se definieron en función de la dirección de vivienda de los casos utilizados en el análisis. El sistema de información geográfica utilizado en este caso fue QGIS.

En el año 2018 en el mismo marco de cooperación entre CONAE y el Ministerio de Salud de la Nación, se publicó una

investigación sobre el uso de herramientas geoespaciales para la toma de decisiones en la prevención y control de arbovirus [33]. Este análisis se realizó a partir de imágenes satelitales de la ciudad de Córdoba (Argentina) y datos tomados en el campo generando así mapas de probabilidad de la presencia de presencia, en este caso, del vector que transmite los arbovirus. Estos mapas proporcionan a los interesados de lugares de foco donde aplicar estrategias de control de vectores.

En el año 2013 se realizó una revisión sistemática de los análisis geoespaciales en estudios epidemiológicos de dengue conjunta entre la Universidade de São Paulo (Brasil) y Johns Hopkins University (USA) [34]. Esta revisión del estado del arte de ese momento, arrojó como resultado luego de analizar distintas investigaciones publicadas entre 2001 y 2011 que no eran muchas las investigaciones que realizaban análisis espaciales, sin embargo, en las investigaciones que si realizaron análisis espaciales se pudo comprobar la utilidad y la necesidad de estos para  para el desarrollo de análisis espaciales con complejidad que se extiende más allá de la simple visualización afirmando que la información que se obtiene de herramientas como los sistemas de información geográfica pueden ser herramientas poderosas para el monitoreo de la transmisión de dengue a un nivel local.

### *C. Otros análisis estadísticos para la predicción de brotes*

En el año 2015 dos investigadores de la facultad de computación e informática de Multimedia University en Malasia hicieron un trabajo para el análisis de series temporales y predicción de Dengue a partir de datos obtenidos del gobierno de Malasia [35].

En este caso, el enfoque fue estadístico y se apoyó en datos históricos entre febrero de 2010 hasta abril de 2015, cubriendo así un periodo de 269 semanas, publicados por el gobierno de Malasia. La metodología de análisis consistió en la descomposición de la serie temporal y aplicar ARIMA para realizar predicciones.

En la descomposición, la serie temporal se separa en componentes de tendencia, estacionales y cíclicos. El componente de tendencia indica si en el largo plazo hay tendencias crecientes o decrecientes, el componente estacional indica si la serie temporal está influenciada por factores estacionales (ciertos meses o semanas del año, días de la semana, etc.) y el componente cíclico está presente cuando en la serie temporal hay altas y bajas que no son fijas en el tiempo por periodos de al menos dos (2) años. ARIMA es un modelo estadístico altamente usado para realizar predicciones sobre modelos de series temporales al obtener la autocorrelación en series temporales estacionarias. La conclusión de esta investigación fue que la serie temporal de casos acumulados de dengue presentaba un patrón estacional con tendencias a largo plazo por lo que era factible utilizar métodos de descomposición para realizar predicciones.

En el año 2019 se realizó un estudio similar en la India [36] en el que se analizaron las tasas de admisión de pacientes con Dengue obteniendo resultados similares.

En la región, en el año 2008 se realizó una investigación para analizar a través de series temporales la incidencia del Dengue en Rio de Janeiro, Brasil [37]. En esta investigación se utilizaron datos históricos entre 1997 y 2004. Utilizando el modelo ARIMA en la serie temporal concluyeron que el número de casos de Dengue en un mes se pueden estimar por la cantidad de casos de Dengue, un, dos y doce meses antes del mes a estimar.

## V. Desarrollo

Basado en todo lo anterior, se presenta necesario en el contexto argentino desarrollar y mejorar las plataformas de recolección de datos, así como las herramientas para el procesamiento y análisis de esos datos con la finalidad de que las instituciones públicas y/o privadas puedan tomar acciones preventivas o acciones correctivas sobre los focos o posibles focos de infección no solo del dengue, sino también de otras enfermedades transmitidas por vectores como Zika, Chikungunya, fiebre amarilla, mal de Chagas, etc., tomando en cuenta que el control de vector se presenta como la principal estrategia para el control de este tipo de enfermedades.

La solución que se presenta está basada en un sistema de recolección de datos anónima donde los usuarios harán un registro de los síntomas que tienen y además registran no solo una dirección aproximada de su lugar de residencia, sino incluyen otras direcciones relevantes como dirección de trabajo/estudio. Inicialmente se registrará en la base de datos la fecha del registro, fecha de aparición de síntomas y dependiendo de los síntomas se inferirá una posible fecha de contagio y se solicitará a los usuarios confirmación y diagnostico en caso de que acudan a un servicio médico autorizado.

Al agregar no solo la dirección de residencia sino otras direcciones relevantes al usuario se incluye el concepto de movilidad dentro los datos a analizar, un concepto que usualmente se ignora en otros estudios y soluciones similares.

Por otra parte, al inferir la enfermedad y la fecha de infección desliga del usuario el diagnosticar la enfermedad evitando el autodiagnóstico, y al permitir al usuario verificar la enfermedad luego de ser evaluado por un prestador de servicios médicos autorizado, se busca corregir errores de diagnóstico realizados por la aplicación.

Cabe destacar que el objetivo de la aplicación no es realizar diagnósticos médicos, por lo que la aplicación no devolvería al usuario un posible diagnóstico, sino que a partir de los síntomas y de las ubicaciones relevantes recopilados de manera anónima se puedan inferir o ubicar los focos de infección de las distintas enfermedades.

Para que el diagnóstico sea el más acertado posible se necesita de algún servicio que pueda realizar la predicción. El objetivo de la aplicación no es realizar diagnósticos a los usuarios, por lo que en esta propuesta no se hace foco en el algoritmo o herramienta que permita inferir la enfermedad, sino que se asume se utilizará un servicio de terceros, que basados en los síntomas declarados y otros factores como, por ejemplo, la época del año ubicación geográfica, otros casos de enfermedades confirmadas actuales y la historia de brotes de enfermedades en el pasado, genere una predicción de la enfermedad que tiene el usuario.

Los set de datos generados a través de esta plataforma tienen varios datos importantes. El primero es la enfermedad asociada, luego se tiene la fecha de registración del evento, fecha de aparición de síntomas y la fecha inferida de contagio basada en la enfermedad diagnosticada, por otro lado se tienen varias direcciones o referencias geoespaciales las cuales analizar. La identidad del usuario no es relevante, por lo que simplemente existirá algún identificador que permita luego al usuario reingresar al sistema y actualizar el diagnostico real en caso haya sido diagnosticado en un centro médico, como por ejemplo, número de teléfono.

En este sentido, la base de datos sobre la cual guardar toda esta información se presta de vital importancia. Dadas las limitaciones de la mayoría de los sistemas de información geográfica para procesar datos en tiempo real, o hacer análisis espacio-temporales, sobre todo en las aplicaciones de acceso gratuito y de licencia abierta, las capacidades que pueda tener la base de datos para almacenar, y procesar datos con características espacio-temporales puede ser crítica para salvar las limitaciones del sistema de información geográfica.

La base de datos que almacene y soporte el análisis de estos datos debe tener la capacidad de procesar en tiempos razonables preguntas en el dominio espacio-temporal como las siguientes: para un periodo de tiempo dado, cuáles fueron las zonas asociadas a la mayor cantidad de casos de X enfermedad, o para una zona geográfica delimitada, cuál o cuáles fueron los periodos donde la ocurrencia de X enfermedad cumple cierto criterio. Es importante que preguntas de este tipo puedan ser respondidas con cierta facilidad y que luego los resultados de las consultas puedan ser exportados y utilizados en otras aplicaciones como sistemas de información geográfica u otras aplicaciones de terceros.

Si bien se pueden utilizar bases de datos relacionales de libre acceso clásicas como MySQL y PostgreSQL que además incluyen extensiones para el tratamiento de datos geo-referenciados, una alternativa que se presenta interesante es la de utilizar bases de datos específicamente creadas para series temporales como pueden ser Redis Time Series, Prometheus, o InfluxDB para guardar y analizar los datos como series temporales que tienen referencia geográfica y no al revés, es decir, datos geográficos con referencia temporal.

Esto es así debido a que en este caso de estudio específico, nuestro objeto sujeto a temporalidad no es la geografía u objetos que representan geografía como tal (como en casos donde la geografía sufre variaciones en el tiempo) sino eventos que interactúan y se acumulan en el tiempo pero que ocurren sobre una geografía estática. Es importante recalcar esto porque el objetivo de la solución es, sobre una geografía estática, localizar en tiempo real los posibles focos de infección y utilizar los sistemas de información geográfica para dar soporte a otro tipo de análisis estáticos que puedan estar asociados a datos geográficos más puros como pueden ser el nivel de urbanización, presencia de cuerpos de agua, obra pública, etc.

En la actualidad existen múltiples soluciones para el almacenamiento y procesamiento de series temporales. Entre las soluciones actuales están: Prometheus [38] un proyecto de código abierto alojado en la CNCF (Cloud Native Computing Foundation, que también aloja el proyecto Kubernetes), Redis Time Series [39] un módulo especializado para el manejo de series temporales sobre la plataforma de datos en tiempo real Redis. Otras soluciones disponibles son InfluxDB, Kinetica y TimescaleDB [40] que se presentan como soluciones más integrales al problema ya que todas son capaces de realizar análisis espacio-temporales sobre datos en series temporales, pero que se ofrecen como soluciones en la nube con costos extras asociados.

Luego de una amplia búsqueda, CrateDB [41] se presenta como la mejor opción de bases de datos para series temporales disponible para la solución planteada. Esto es así debido a que CrateDB es una base de datos de código abierto, cuya implementación en servidores propios es gratuita que además ofrece soporte nativo para datos espaciales y capacidad de hacer consultas sobre datos geoespaciales y temporales.

Entonces, los datos recolectados pueden ser utilizados para analizar tanto históricamente como en tiempo real los patrones de las enfermedades, individualmente, o en conjunto, en la población a través de las referencias geoespaciales. A partir de los datos se pueden responder preguntas como por ejemplo: ¿cuáles son para un periodo de tiempo dado los “focos” de infección de X enfermedad? ¿cómo se desarrolla/propaga la enfermedad a lo largo del tiempo? ¿cuál es la predicción de número de casos basados en el comportamiento actual e histórico de la serie temporal de X enfermedad?

Algoritmos de inteligencia artificial y modelos de aprendizaje para series temporales pueden ser útiles para modelar los datos históricos y obtener "insights" sobre los datos en tiempo real.

Por último, una vez procesada o preprocesada la información, esta puede trasladarse a sistemas de información geográfica de acceso gratuito y de código abierto como QGIS para hacer otros análisis espaciales como distintas técnicas de análisis de "clusters", autocorrelación espacial de los casos, generación de mapas temáticos, entre otros.




*Fig. 1 Propuesta de Arquitectura*

## VI. Conclusiones

Las enfermedades transmitidas por vectores biológicos son una problemática global que pone en riesgo potencial de infección a 2.5 billones de personas cada año. En la Argentina y en los países limítrofes casos confirmados de Dengue llegaron a máximos históricos en el año 2024 confirmando que las enfermedades transmitidas por vectores como el Dengue son una problemática local y significativo en la actualidad. En este contexto, se presenta necesario generar la mayor cantidad de datos relevantes que puedan ser procesados y utilizados por los actores correspondientes para tomar acciones preventivas y correctivas sobre los focos de infección tomando en cuenta que el control en la proliferación de vectores es la estrategia principal para la prevención de este tipo de enfermedades. Los sistemas de información geográficos y bases de datos para series temporales con capacidad de consulta para datos geográficos se muestran como herramientas con el potencial de dar el soporte necesario para el análisis de los datos generados sobre casos de infección.

Si bien la implementación de una solución como la propuesta en el desarrollo puede tener un impacto significativo basado en casos similares, aun utilizando herramientas de código abierto con licencia de libre uso, los costos de implementación no son fácilmente dimensionables y encontrar los patrocinadores que quieran financiar un proyecto para recolectar información sobre posibles casos de infección abierto a la comunidad basado en bases de datos de series temporales se presenta como una dificultad considerable.

Por otro lado al momento de la realización de esta investigación no se encontró un servicio de terceros que pueda ofrecer un diagnóstico y tiempos de incubación de la enfermedad basado en los síntomas reportados, por lo que un elemento importante de la solución no existe en este momento, mostrándose como una posible línea de investigación futura.

## VII. Referencias

[1] World Health Organization. (2014). A global brief on vector-borne diseases. World Health Organization. https://iris.who.int/handle/10665/111008.

[2] Wilson AL, Courtenay O, Kelly-Hope LA, Scott TW, Takken W, Torr SJ, et al. (2020) The importance of vector control for the control and elimination of vector-borne diseases. PLoS Negl Trop Dis 14(1): e0007831. https://doi.org/10.1371/journal.pntd.000783.

[3] (2024, Abril 7). Brote histórico de dengue en Argentina: Ya son 161 los muertos y más de 230 mil casos registrados. Infobae. Recuperado en Junio de 2024 de https://www.infobae.com/salud/2024/04/07/brote-historico-de-dengue-en-argentina-ya-son-161-los-muertos-y-mas-de-230-mil-casos-registrados/

[4] (2024, Abril 9). Brasil ya registró un nuevo récord anual de muertes por dengue en poco más de tres meses. *Infobae*. Recuperado en Junio de 2024 de https://www.infobae.com/america/america-latina/2024/04/09/brasil-ya-registro-un-nuevo-record-anual-de-muertes-por-dengue-en-poco-mas-de-tres-meses/

[5] (2024, Abril 8). Uruguay atraviesa un récord de pacientes graves por dengue pero descarta la vacunación. *Infobae*. Recuperado en Junio de 2024 de https://www.infobae.com/america/america-latina/2024/04/08/uruguay-atraviesa-un-record-de-pacientes-graves-por-dengue-pero-descarta-la-vacunacion/

[6] Caminade, C., McIntyre, K.M. and Jones, A.E. (2019), Impact of recent and future climate change on vector-borne diseases. Ann. N.Y. Acad. Sci., 1436: 157-173. https://doi.org/10.1111/nyas.13950

[7] Brady, O.J.; Gething, P.W.; Bhatt, S.; Messina, J.P.; Brownstein, J.S.; Hoen, A.G.; Moyes, C.L.; Farlow, A.W.; Scott, T.W.; Hay, S.I. Refining the Global Spatial Limits of Dengue Virus Transmission by Evidence-Based Consensus. PLoS Negl. Trop. Dis. 2012, 6, e1760.

[8] Angelo, M., Ramalho, W. M., Gurgel, H., Belle, N., & Pilot, E. (2020). Dengue Surveillance System in Brazil: A Qualitative Study in the Federal District. International journal of environmental research and public health, 17(6), 2062. https://doi.org/10.3390/ijerph17062062

[9] Lopez-Cruz, R. (2008). Spatial and temporal dynamics of dengue fever in Peru: 1994-2006. Epidemiology and Infection.

[10] de Carvalho, R. M., & Nascimento, L. F. C. (2012). Spatial distribution of dengue in the city of Cruzeiro, São Paulo State, Brazil: use of geoprocessing tools. Revista Do Instituto De Medicina Tropical De São Paulo, 54(5), 261–266. https://doi.org/10.1590/S0036-46652012000500005

[11] Dong Ho Kim, Keun Ho Ryu, Hong Soo Kim. A spatiotemporal database model and query language, Journal of Systems and Software, Volume 55, Issue 2, 2000, Pages 129-149, ISSN 0164-1212, https://doi.org/10.1016/S0164-1212(00)00066-2. (https://www.sciencedirect.com/science/article/pii/S0164121200000662)

[12] Paton, Norman & Fernandes, Alvaro & Griffiths, Tony. (2000). Spatio-temporal databases: contentions, components and consolidation. 851 - 855. 10.1109/DEXA.2000.875125.

[13] World Health Organization. Vector-borne diseases. Mayo 2024. http://www.who.int/mediacentre/factsheets/fs387/en/. Consultada el 18 de mayo de 2024.

[14] Huntington, M. K. (2016). Emerging Vector-Borne Diseases. American Family Physician.

[15] Martens, W. J. M., Jetten, T. H., Rotmans, J., & Niessen, L. W. (1995). Climate change and vector-borne diseases. Global Environmental Change, 5(3), 195–209. https://doi.org/10.1016/0959-3780(95)00051-O
[16] Remoaldo, P. C., Ribeiro, V. P., Lopes, H. S., & Silva, S. C. G. (2018). Geographic Information Systems. Encyclopedia of Information Science and Technology, Fourth Edition, 3460–3472. https://doi.org/10.4018/978-1-5225-2255-3.CH301
[17] Waters, N. (2003). Geographic Information Systems. Aters, N. M. 2003. Geographic Information Systems. Pp. 1106-1115 in the Encyclopedia of Library and Information Science (Second Edition), Edited by Drake, M., Marcel Dekker, Inc., New York.
[18] ArcGIS Inc. (2024). ArcGIS. Recuperado en Junio 2024 de https://www.arcgis.com/index.html
[19] GRASS Development Team (2024) GRASS GIS. Recuperado en Junio 2024 de https://grass.osgeo.org
[20] Asociación gvSIG (2024). gvSIG. Recuperado en Junio 2024 de http://www.gvsig.com
[21] Kwan, M. P. (2013). Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility: Space–Time Integration in Geography and GIScience. Annals of the Association of American Geographers, 103(5), 1078–1086. https://doi.org/10.1080/00045608.2013.792177
[22] Richardson, D. B. (2013). Real-Time Space–Time Integration in GIScience and Geography: Space–Time Integration in Geography and GIScience. Annals of the Association of American Geographers, 103(5), 1062–1071. https://doi.org/10.1080/00045608.2013.792172
[23] Brent Hall; Michael G. Leahy (2008). Open Source Approaches in Spatial Data Handling. Springer. pp. 126–128. ISBN 978-3-540-74830-4.
[24] Jin, Peiquan & Yue, Lihua & Gong, Yuchang. (2005). RESEARCH ON UNIFIED SPATIOTEMPORAL DATA MODEL.
[25] InfluxData Inc. (2024). Recuperado en Junio 2024 de https://www.influxdata.com/blog/tldr-geotemporal-data-influxdb/#:~:text=Geo%2Dtemporal%20data%20has%20both,connect%20them%20with%20other%20events.
[26] Kinetica Inc. (2024) Recuperado en Junio 2024 de https://www.kinetica.com/blog/what-database-for-spatio-temporal-analysis/
[27] Recuperado en Junio 2024 de http://www.dpi.inpe.br/terralib5/wiki/doku.php
[28] The GeoMesa Project (2024). Recuperado en Junio 2024 de http://www.geomesa.org
[29] Erwig, M., Schneider, M., & Güting, R. H. (1997). Temporal and Spatio-Temporal Data Models and Their Expressive Power (Vol. 225). FernUniversität in Hagen. https://nbn-resolving.org/urn:nbn:de:hbz:708-dh5129
[30] Holston, J. (2021). DengueChat: A Social and Software Platform for Community-based Arbovirus Vector Control. The American Journal of Tropical Medicine and Hygiene.
[31] Lanfri, Sofia & Frutos, Nicolás & Rotela, Camilo & Peralta, Gonzalo & Elia, Estefanía & Lanfri, Mario & Scavuzzo, Carlos. (2011). Algoritmos para la alerta temprana de Dengue en un ambiente geomático.
[32] Gorla, D. (2017). Analytical report of the 2016 dengue outbreak in Córdoba city, Argentina. Geospatial Health.
[33] Nazarena, María & Rojas, R & Lopez, Laura & Guzman, Claudio & Lanfri, Mario & Albornoz, Carlos & Scavuzzo, Carlos & Garcia, Fernanda & Porcasi, Ximena. (2018). Use of geospatial tools for decision-making in the arboviruses prevention and control, in the Cordoba city, Argentina Ximena Porcasi Exequiel Aguirre Andrés Lighezzolo.
[34] Castillo-Salgado, C., & Lenharo, A. (2013). Geospatial analysis applied to epidemiological studies of dengue: a systematic review. Revista Brasileira De Epidemiologia.
[35] Ho, C. C., & Ting, C.-Y. (2015). Time Series Analysis and Forecasting of Dengue Using Open Data. Lecture Notes in Computer Science, 51–63. https://doi.org/10.1007/978-3-319-25939-0_5
[36] Abhinandithe, S.K. (2019). Time Series Analysis on Admission Rates of Dengue In Medical College Hospital.
[37] Luz, Paula & Mendes, Beatriz & Codeço, Cláudia & Struchiner, Claudio & Galvani, Alison. (2009). Time Series Analysis of Dengue Incidence in Rio de Janeiro, Brazil. The American journal of tropical medicine and hygiene. 79. 933-9. 10.4269/ajtmh.2008.79.933.
[38] Prometheus Authors (2024). Recuperado en Junio 2024 de https://prometheus.io
[39] Redis (2024). Recuperado en Junio 2024 de https://redis.io/timeseries/
[40] Timescale Inc. Recuperado en Junio 2024 de https://www.timescale.com
[41] CrateDB (2024). Recuperado en Junio 2024 de https://cratedb.com

## VIII. Biografía del autor



**Enoc Leonhardo García Veliz**. Nacido en Caracas, Venezuela el 9 de octubre de 1996. Viviendo en Buenos Aires desde el año 2017. Estudiante de Licenciatura en Inteligencia Artificial en la Universidad de Palermo, Argentina. Músico de hobby, aficionado a las series y a las películas. Con experiencia laboral principalmente en el ecosistema de SAP Business One, como consultor y desarrollador.