



Patrones evolutivos de infraestructuras de datos espaciales para el modelado de ecosistemas de datos

Tatiana Delgado Fernández ^{1*} <https://orcid.org/0000-0002-4323-9674>
José Luis Capote Fernández ² <https://orcid.org/0000-0002-3733-0893>
Joep Crompvoets ³ <https://orcid.org/0000-0003-1077-597X>
Rafael Cruz Iglesias ² <https://orcid.org/0000-0002-7564-7779>
Libia Denise Cangrejo Aljure ⁴ <https://orcid.org/0000-0001-8653-4393>
Guillermo González Suárez ² <https://orcid.org/0000-0003-0040-6062>

¹ Universidad Tecnológica de La Habana José Antonio Echeverría. La Habana, Cuba

² Empresa GEOCUBA Investigación y Consultoría, Cuba

³ Universidad Católica de Lovaina. Valonia, Bélgica

⁴ Universidad Nacional de Colombia. Bogotá, Colombia

*Autor para la correspondencia: tatiana.delgado@uic.cu

Editor

Lisset González Navarro
Academia de Ciencias de Cuba.
La Habana, Cuba

Traductor

Darwin A. Arduengo García
Academia de Ciencias de Cuba.
La Habana, Cuba

RESUMEN

Introducción: Las infraestructuras de datos espaciales no están diseñadas para convivir de forma colaborativa con los nuevos modelos de datos provenientes de fuentes heterogéneas. **Objetivo:** Analizar la evolución de las interrelaciones de estas infraestructuras con tecnologías digitales, y cómo esto conduce a modelar los ecosistemas de datos. **Métodos:** Se siguen 3 marcos de trabajo. El primero es el modelo de referencia de procesamiento abierto distribuido usado para describir las complejidades de las infraestructuras de datos espaciales. El segundo se relaciona con patrones evolutivos que contemplan modelos y arquitecturas representativas de las interrelaciones entre tales infraestructuras con tecnologías digitales emergentes. El último patrón sustenta la arquitectura de los ecosistemas de datos propuesta. **Resultados:** El primer patrón dio como resultado una arquitectura de infraestructuras de datos espaciales colaborativa en la nube y constituyó un hito por su escalabilidad. Con el segundo, se obtiene un modelo de servicios semánticos con mayor eficacia en la recuperación de información. La sensibilidad al contexto y la gestión de datos en entornos de Internet de las cosas para datos geoespaciales es el resultado del tercer patrón, mientras el cuarto, añade valor a los anteriores y ofrece arquitecturas para gestionar datos complejos. La experimentación de la gobernanza de estas infraestructuras se realiza a partir de una simulación basada en agentes. **Conclusiones:** Los patrones evolutivos experimentados y algunos referentes de arquitecturas de datos permitieron modelar una arquitectura de ecosistemas como base de nuevos ecosistemas digitales flexibles, ágiles y autosostenibles.

Palabras clave: infraestructura de datos espaciales; arquitectura; ecosistemas de datos; transformaciones digitales

Evolving patterns of spatial data infrastructures for modeling of data ecosystems

ABSTRACT

Introduction: Spatial data infrastructures are not designed to coexist collaboratively with new data models from heterogeneous sources. **Objective:** To analyze the evolution of the interrelationships of these infrastructures with digital technologies, and how this leads to modeling data ecosystems. **Methods:** Three frameworks are followed. The first is the open distributed processing reference model, used to describe the complexities of spatial data infrastructures. The second is related to evolutionary patterns that contemplate models and architectures representative of the interrelationships between such infrastructures with emerging digital technologies. The latter pattern underpins the proposed data ecosystem architecture. **Results:** The first pattern resulted in a collaborative spatial data infrastructure architecture in the cloud and was a milestone for its scalability. With the second, it was obtained a semantic service model with greater efficiency in information retrieval. Context sensitivity and data management in Internet of Things environments for geospatial data is the result of the third pattern, while the fourth adds value to the previous ones and offers architectures to manage complex data. The experimentation of the governance of these infrastructures is carried out from an agent-based simulation. **Conclusions:** The evolutionary patterns experienced and some references of data architectures made it possible to model an ecosystem architecture as the basis for new flexible, agile and self-sustaining digital ecosystems.

Keywords: spatial data infrastructure; architecture; data ecosystems; digital transformations.

INTRODUCCIÓN

Surgidas a finales del siglo XX las infraestructuras de datos espaciales (IDE) se definen por un conjunto de políticas, arreglos institucionales y tecnologías que aprovechan la provisión y el uso de datos y servicios espaciales estandarizados.

⁽¹⁾ Las IDE se conciben como un tipo particular de infraestructura de información emergida sobre la web, especializadas en el dominio de la información geoespacial.⁽²⁾ Como redes de redes estas infraestructuras de datos están soportadas por una relación de coordinación entre sus componentes, que se apoya en estándares y políticas, más que en una funcionalidad puramente de control.⁽³⁾

En la evolución del concepto de IDE, Craglia ⁽⁴⁾ reconoció que las mismas resultaron ser más complejas de lo previsto originalmente y que se trata más de personas y redes sociales que de tecnología y estándares. A partir de la evolución de la web social o web 2.0, comienza un movimiento donde los propios usuarios se convierten en proveedores de datos, lo cual trajo consigo una oportunidad en el dominio geoespacial, particularmente en aquellos territorios donde la cartografía oficial era escasa o nula. El término información geográfica voluntaria (VGI) es universalmente aceptado para denominar a aquella información geográfica generada por usuarios no

especializados, generalmente utilizando una plataforma colaborativa como OpenStreetMap o Wikimapia.

El proyecto más conocido de datos generados por ciudadanos en el dominio geoespacial es OpenStreetMap (OSM), que involucra a voluntarios de todo el mundo en la creación y actualización de un mapa global, disponible bajo una licencia de acceso abierto, y es actualmente utilizado por un sinnúmero de actores, incluidas organizaciones humanitarias, empresas y gobiernos.⁽⁵⁾ El término VGI se inserta en un movimiento de la web social donde los datos geográficos son proveídos, modificados y compartidos por los usuarios utilizando servicios online interactivos.⁽⁶⁾

Uno de los impactos más importantes de los datos abiertos y la información geográfica voluntaria en la evolución de las IDE se manifiesta en su gobernanza. Estructuradas para ser accedidas a través de metadatos que describen los conjuntos de datos espaciales disponibles, las IDE tradicionales se suelen construir desde las agencias cartográficas y otras entidades gubernamentales, las cuales publican sus datos siguiendo estándares abiertos. Sin embargo, las IDE desde sus conceptualizaciones básicas no están suficientemente equipadas para admitir datos de las llamadas multitudes o crowdsourcing, que suelen recopilarse de forma simple, de acuerdo a la experiencia del usuario.

Particularmente relevantes para el aprendizaje sobre IDE y su perfeccionamiento en las 2 últimas décadas, resultaron ser las evaluaciones que desde múltiples perspectivas se hicieron a nivel nacional, regional y global. Entre las investigaciones más importantes al respecto se encuentran las contenidas en un marco de evaluación de las IDE, que abarca los modelos más consensuados en el mundo para evaluar las mismas. ^(7,8) Es de destacar el relevante papel que jugó Cuba en este movimiento al proponer un modelo holístico de evaluación multicriterio que ha servido para evaluar no solo el progreso de la infraestructura de datos espaciales de la República de Cuba, ⁽⁹⁾ sino también en el Caribe, ⁽¹⁰⁾ en la región de Latinoamérica y el Caribe, ⁽¹¹⁾ y a nivel global. ⁽⁷⁾

Recientemente la gobernanza de las IDE es abordada como sistema adaptativo complejo ⁽¹²⁾ y dinámico. ⁽¹³⁾ Cada vez más la gobernanza de lo espacial es menos "especial", debido a la convergencia tecnológica que facilita la propia transición hacia ecosistemas de datos geoespaciales. Esta transición ha venido ocurriendo paulatinamente, a medida que se desarrollan las tecnologías habilitadoras de la transformación digital. La complejidad de los datos modernos, que suele caracterizarse por 5 V (volumen, velocidad, variedad, veracidad y valor), impone la emergencia de nuevos modelos computacionales y organizacionales. Big data, Internet de las cosas, la computación en la nube, la inteligencia artificial, las cadenas de bloque o blockchain, y la hiperconectividad son, entre otras, las tecnologías habilitadoras que se asocian a este nuevo paradigma. ⁽¹⁴⁾

Este ecosistema emergente requiere un cambio de paradigma en la estrategia digital de las organizaciones y a nivel de toda la sociedad para ofrecer experiencia digital ambiental a usuarios, a través de múltiples canales y conectividad en tiempo real entre personas, procesos, datos, cosas, negocios y servicios para un mayor valor de la organización. ⁽¹⁵⁾

Nuevas arquitecturas de datos están emergiendo en los últimos años que pueden sustentar los emergentes ecosistemas de datos. Entre estos enfoques se encuentran el de malla de datos (data mesh) y tejido de datos (data fabric). Data mesh es un enfoque sociotécnico descentralizado para compartir, acceder y gestionar datos analíticos en entornos complejos y de gran escala, dentro o entre organizaciones. ⁽¹⁶⁾ Se basa en 4 principios fundamentales: propiedad del dominio, datos como producto, plataforma de autoservicio de datos y gobernanza computacional federada.

Por su parte, data fabric es un conjunto de principios de gestión de datos, prácticas rectoras, comunidades y estándares que pueden optimizar el acceso a los datos distribuidos de una organización y curarlos y organizarlos de manera inteligente para la entrega de autoservicio. ⁽¹⁷⁾ Constituye una ar-

quitectura de información y una plataforma para la gestión y la integración a nivel de datos, y proporciona interfaces, API y servicios para la integración y comunicación entre los datos. ⁽¹⁸⁾

La arquitectura data fabric fusiona la gestión y analíticas de datos con el poder integrador y semántico de los grafos de conocimientos. Un grafo de conocimiento (KG, por sus siglas en inglés) es entendido como una estructura en forma de gráfico de datos, también conocido como grafo de datos, el cual está destinado a acumular y transmitir conocimiento del mundo real, donde los nodos del grafo representan entidades de interés y sus aristas representan relaciones entre estas entidades. ⁽¹⁹⁾

Por su naturaleza distribuida y el enfoque basado en datos como producto, las IDE están preparadas para insertarse en modelos de ecosistemas de datos soportados por los principios de data mesh. La experiencia acumulada en más de 3 lustros investigando y desarrollando extensiones semánticas de IDE ha facilitado la inclusión de los grafos de conocimiento, como nodo integrador, bajo los principios de data fabric.

Los ecosistemas de datos se definen como un conjunto de redes compuestas por actores autónomos que, directa o indirectamente, consumen, producen o proporcionan datos y otros recursos relacionados (por ejemplo, software, servicios e infraestructura). Cada actor desempeña uno o más roles y está conectado a otros actores a través de relaciones, de tal manera que la colaboración y competencia de los actores promueve la autorregulación del ecosistema de datos. ⁽²⁰⁾

Este trabajo tiene como objetivo describir un grupo de patrones evolutivos de IDE, expresados en arquitecturas de tecnología de información y dinámicas de gobernanza, que alimentan el diseño de una nueva arquitectura de ecosistema de datos geoespaciales, soportada en los principios de data mesh y grafos de conocimiento. Por su naturaleza, esta arquitectura debe contribuir a la autosostenibilidad, flexibilidad y agilidad de los ecosistemas digitales que emergen en torno a la transformación digital de la sociedad.

MÉTODOS

Para respaldar la construcción de una arquitectura de ecosistema de datos, se sigue una metodología que descansa en 3 marcos de trabajo: a) modelado de IDE según el modelo de referencia de sistemas de procesamiento abierto distribuido (RM-ODP) ISO/IEC. ⁽²¹⁾ b) patrones evolutivos de IDE soportados en las tecnologías habilitadoras de la transformación digital. c) nuevas arquitecturas de datos con enfoque distribuido (*data mesh*, *data fabric* y grafos de conocimiento).

El modelo de referencia de los sistemas de procesamiento abierto distribuido (RM-ODP) permite desarrollar modelos especificando componentes que son mutuamente consisten-

tes y que pueden ser combinados para construir infraestructuras conectadas con los requerimientos de usuarios. ⁽²⁾ En RM-ODP se establecen 5 puntos de vista, cada uno de los cuales tiene asociado un lenguaje de punto de vista, el cual expresa conceptos y reglas relevantes a un área particular. Estos 5 puntos de vista o perspectivas son: a) perspectiva de negocio o empresa: relacionada con las políticas, actividades y roles presentes en la organización. b) perspectiva de información: relacionada con la información que necesita ser guardada y procesada en el sistema. c) perspectiva computacional: relacionada con la descripción de un sistema como un conjunto de objetos que interactúan mediante interfaces permitiendo la distribución. d) perspectiva de ingeniería: relacionada con los mecanismos que soportan la distribución. e) perspectiva tecnológica: relacionada con el detalle de los componentes con los que el sistema distribuido es construido. ⁽²⁾

Los patrones de interrelación de tecnologías, capacidades y modelos emergentes de las IDE permiten ir descubriendo incrementalmente los componentes arquitectónicos principales de los ecosistemas de datos de próxima generación. Adicionalmente, se ofrece un patrón evolutivo enfocado a la gobernanza de IDE en el contexto cambiante y adaptativo en el que se desarrollan los ecosistemas digitales actuales.

Los principios de *data mesh*, referidos como propiedad del dominio, datos como producto, plataforma de datos de auto-servicio y gobernanza computacional federada ⁽¹⁶⁾ junto con la arquitectura *data fabric*, soportada en Grafos de conocimiento, sustentan el diseño de una nueva arquitectura de ecosistema de datos geoespaciales que responde a las dinámicas y megatendencias asociadas a la transformación digital.

RESULTADOS

Modelado de infraestructuras de datos espaciales basado en modelo de referencia de sistemas de procesamiento abierto distribuido

Como un fenómeno sociotécnico de naturaleza compleja las infraestructuras de datos espaciales deben ser modeladas desde varios puntos de vista. Para conceptualizar las IDE se modelaron inicialmente las 5 perspectivas de RM-ODP. ⁽²⁾ Más tarde se refinaron las perspectivas de negocio e información ⁽²²⁾ y la perspectiva computacional. ⁽²³⁾ Se usó también RM-ODP como base referencial para modelar la IDE extendida con *web semántica* ^(24,25) y en la nube, ⁽²⁶⁾ facilitando un marco de trabajo para la implementación de nuevas extensiones a la infraestructura de datos espaciales de la República de Cuba.

Patrón de interrelación de computación en la nube e infraestructuras de datos espaciales

La Web 2.0 y la computación en la nube tienen una influencia directa sobre la forma en que se consume una IDE. La infraestructura de datos espaciales de la República de Cuba (IDERC) ha sido beneficiada con las contribuciones de varias generaciones de investigadores. Uno de estos aportes fue resultado de un estudio doctoral ⁽²⁶⁾ que modeló la IDERC en la nube, ofrecida como un servicio, en sus 3 niveles: IaaS-infraestructura como un servicio, PaaS-plataforma como un servicio y SaaS-*software* como un servicio. En este trabajo se destaca la capa de plataforma como un servicio (PaaS) que facilita la provisión de datos espaciales mediante servicios *web*, lo que permite que los proveedores no tienen que asimilar ni implementar ningún *software*, mejorando la disponibilidad de la información necesaria para la toma de decisiones a todos los niveles.

IDEaaS fue un modelo pionero en el mundo al migrar la IDE a la nube, particularmente en su dimensión PaaS (plataforma como servicio) que habilitó capacidades colaborativas prácticamente ilimitadas para este tipo de infraestructuras. Significó un hito en el desarrollo de la IDERC, por cuanto permitió desplegar los servicios de las IDE como servicios virtualizados y escalables elásticamente en el Centro de Datos del Ministerio de Transporte en Cuba (Transnet), para soportar otras aplicaciones, siendo una de las más importantes el sistema de control de flotas, MovilWeb.

Patrón de interrelación de la web semántica y las infraestructuras de datos espaciales

Ante la ineficacia en las consultas tradicionales de una IDE y aprovechando el surgimiento de la Web 3.0 o Web semántica, se llevó a cabo una investigación doctoral que propuso un modelo de servicios semánticos para habilitar el uso de restricciones basadas en los atributos de las fuentes de datos y, con ello, aumentar la eficacia de la recuperación de información en infraestructuras de datos espaciales. ⁽²⁴⁾ Como resultado se creó una base de conocimientos para la IDERC con las ontologías siguientes:

- Ontología geoespacial base (derivada de ISO)
 - Elemento,
 - tipo de función,
 - relación espacial.
- Ontologías de IDE
 - Ontología de los nombres geográficos,
 - límites de administración,
 - ontología de servicios geoespaciales basada en OWL-S.
 - Ontología de aplicación (propiedad del modelo)
 - Ontología de traducción de filtros,
 - ontología de anotación.

Este modelo ofreció una solución de consultas espaciales sin almacenar información de geometrías en la base de conocimiento, dado que el procesamiento de las consultas se realiza en los servicios de objetos geográficos (WFS- *web feature service*) de OGC (*open geospatial consortium*). De esta forma, la base de conocimiento queda más pequeña y no es necesaria la utilización de métodos de razonamiento espacial. Entre las ventajas del modelo de servicios semánticos para IDE, que lo distingue esencialmente de otros existentes, se encuentra la relación semántica que se establece con ontologías de dominio hasta el nivel de atributos que, junto a la anotación de conocimiento en forma de restricciones de clases, habilita la recuperación de información vectorial basada en el conocimiento. (27) La figura 1 expone la interfaz de la consulta semántica espacial del modelo y un ejemplo de búsqueda (Ríos en Manicaragua), ejecutado a través del modelo de los servicios semánticos desarrollado para la IDERC.

En la IDERC se han experimentado algunas extensiones de los modelos semánticos, dando lugar a aplicaciones basadas en analíticas de datos con enfoque semántico. Un ejemplo de ello se propone un método de un sistema de recomendaciones sensible al contexto soportado por ontologías y basado en filtrado colaborativo, (28) el cual se divide en 3 fases: marco semántico y de datos, sistema de recomendaciones sensible al contexto basado en ontologías y visualización de las recomendaciones. (29) A partir de las limitaciones que suponen los sistemas de recomendaciones tradicionales ante conjuntos de datos voluminosos y variados, como los ofrecidos por el proyecto de la Web Semántica Linked Open Data o LOD, se diseñó una arquitectura para sistemas semánticos

de recomendaciones (figura 2) basada en *Big Data* y datos enlazados, con vistas a incorporar una base de conocimiento global de puntos de interés y facilitar la actualización de las preferencias de los usuarios, favoreciendo la mejora de las predicciones. (30)

Entre las diversas tecnologías y campos conceptuales que subyacen en Internet de las cosas (IoT), sobresalen la computación ubicua o pervasiva, y la sensibilidad al contexto y la semántica, que se complementan con redes de sensores, sistemas embebidos y comunicación inalámbrica, para conceder a los dispositivos o “cosas” la capacidad de medir información, interconectarse y transferirla.

La necesidad de gestionar eficientemente los datos recolectados en un entorno IoT, para afrontar adecuadamente los problemas que surgen de la heterogeneidad, el dinamismo y el tamaño de los datos, plantea retos a la comunidad científica. Algunos de esos retos están asociados con las tareas de representación, interpretación y razonamiento de los datos provenientes de sensores; así como, su integración semántica con otros datos abiertos en la web. En línea con estos desafíos se diseñó el modelo semántico de contexto extendido con Linked Open Data para Internet de las cosas, XSCM_4_IoT, el cual se soporta en una arquitectura de 4 capas y en la red ontológica de contexto CO_Net. El modelo se caracteriza por ser, además de semántico y extensible, abierto, interoperable y centrado en el usuario. El enriquecimiento semántico de datos contextuales usando el proyecto Linked Open Data, a partir de la publicación de los datos y recursos ontológicos de CO_NET, es un rasgo distintivo del modelo. La validación de XSCM_4_IoT en un escenario del Smart Campus en la Uni-

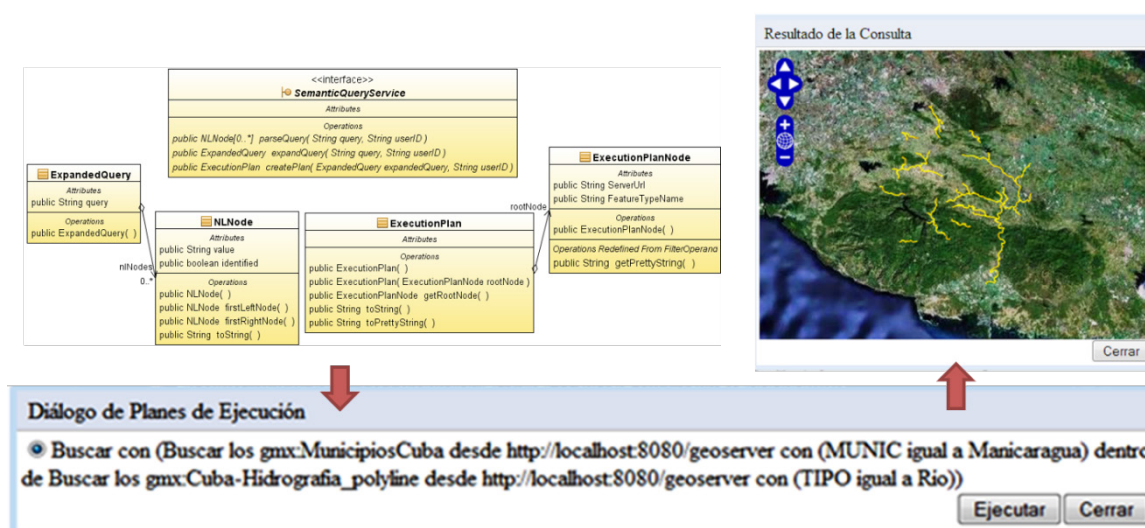


Fig.1. Consulta realizada en la IDERC extendida con el modelo de servicios semánticos. Patrón de interrelación de datos geoespaciales con la Web semántica, sensibilidad al contexto e Internet de las Cosas

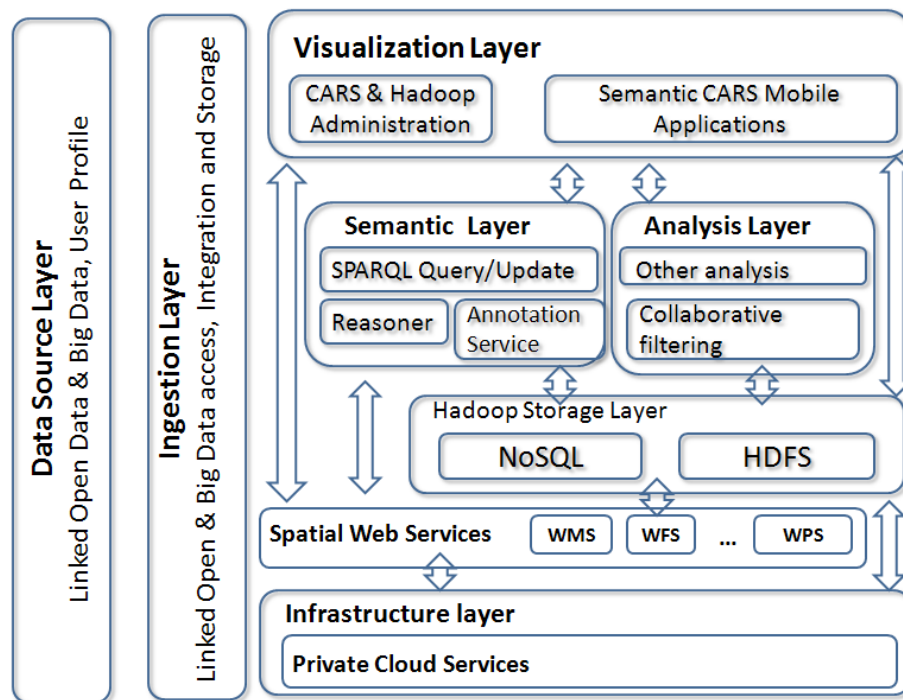


Fig. 2. Arquitectura para sistemas semánticos de recomendaciones basada en *Big Data* y datos enlazados ⁽³⁰⁾

versidad, permite comprobar la viabilidad del mismo para gestionar datos de contexto en un entorno IoT. ⁽³¹⁾ Después se desarrolló una extensión de dicho modelo para profundizar en la capa de razonamiento. ⁽³²⁾ La figura 3 refleja la arquitectura general del modelo XCM_4_IoT extendido con capacidades de razonamiento.

El componente de razonamiento semántico RTR_4_IoT diseñado para flujos de datos que se generan en tiempo real constituyó una solución a la principal limitación que presentan las tecnologías clásicas semánticas de razonamiento, las cuales no contemplan el escenario IoT, caracterizado por datos dinámicos y masivos. ⁽³²⁾

Patrón de interrelación de Big Data, información geográfica voluntaria e inteligencia de negocio geoespacial

La arquitectura de inteligencia geoespacial (GeoBI) y Big Data con VGI para análisis de crimen ofrecida por Herrera F *et al.* ⁽⁶⁾ refleja la interrelación de este patrón y fue utilizada en una aplicación móvil con GPS para denunciar incidencias criminales en la ciudad de Montevideo. La aplicación utiliza el servicio de objetos geográficos (WFS) para cargar sus datos, que se almacenan en una base de datos geográfica. El módulo de apoyo a la decisión se dispara para generar una alerta si la emergencia lo requiere. La información que está siendo almacenada es extraída por procesos ETL (extraer, transformar y cargar) que combinan la información reportada por

los usuarios y las instituciones existentes involucradas y las guardan en almacenes de datos de VGI. En este caso de estudio, los usuarios, como personal de bomberos, policía, etc., pueden consultar en todo momento el área de reclamos del visor web, y a su vez son alertados por la plataforma cuando un delito requiere atención inmediata. ⁽⁶⁾

Se han desarrollado otros escenarios donde se manifiesta este patrón, uno de los cuales fue aplicado en la maestría de Idorys Aguilar León para la integración de datos estadísticos de salud con la IDERC. ⁽³³⁾ La figura 4 muestra la arquitectura obtenida en esta investigación.

El enfoque de integración de datos en la figura 4 utiliza el estándar OGC TJS o Servicio de Unión de Tablas (OGC Table Joining Service, TJS) que es un servicio orientado a establecer vínculos entre una fuente de datos y un marco geográfico utilizando identificadores comunes.

El patrón de interrelación de Big Data, en su dimensión de gestión de datos voluminosos con Infraestructuras de Datos Espaciales, fue experimentado ⁽³⁴⁾ al configurar el servidor de datos geoespaciales GeoServer con trazas GPS de un clúster de Big Data, basado en GeoMesa, ⁽³⁵⁾ para publicar las posiciones de las trazas en forma de mapas. A través de un servidor de información geográfica como GeoServer, GeoMesa facilita la integración con una amplia gama de clientes de mapeo existentes al permitir el acceso a sus bases de datos y capacidades de transmisión sobre

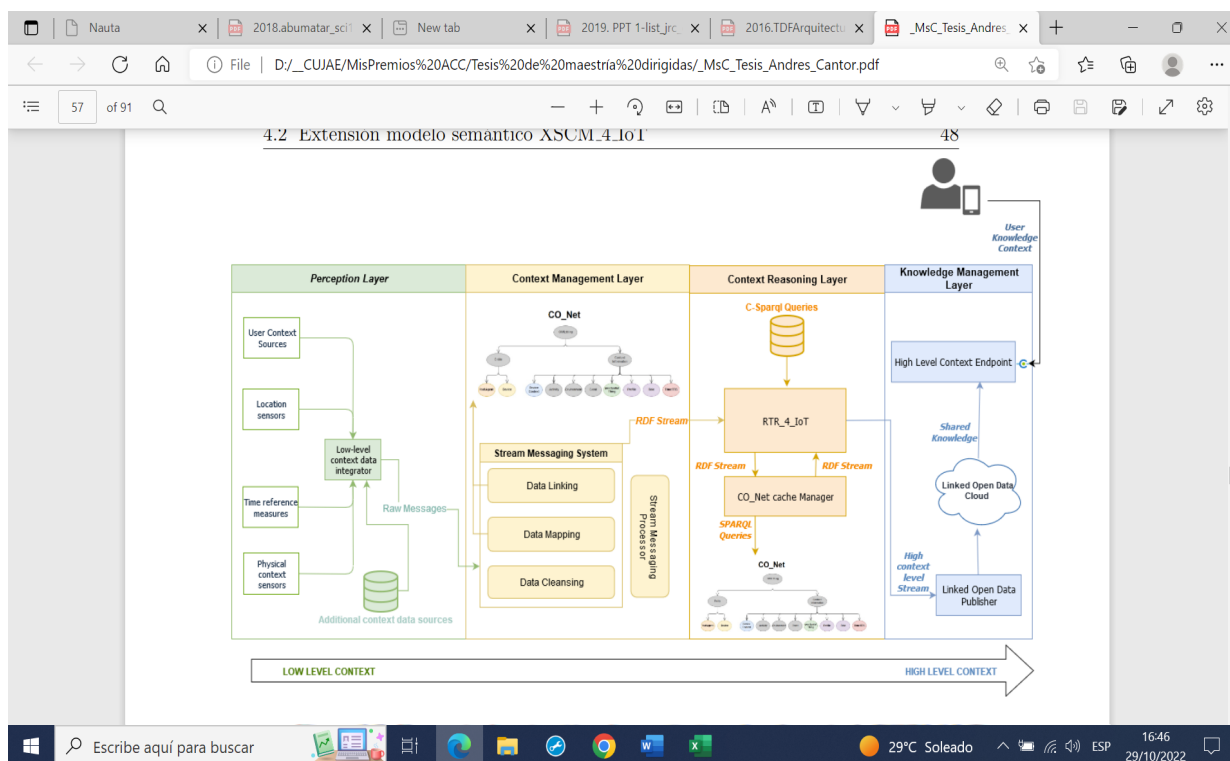


Fig. 3. Arquitectura general del modelo XCM_4-IoT extendido con capacidades de razonamiento ⁽³²⁾

API y protocolos estándar OGC. Estas interfaces también permiten que GeoMesa administre interfaces de usuario de mapas y sirva datos para análisis, como consultas, histogramas, mapas de calor y análisis de series de tiempo. El experimento desarrollado arrojó resultados satisfactorios, tanto para los tiempos de inserción de los datos, como para los tiempos de respuesta a solicitudes de mapas y consultas utilizando dicha interfaz para seleccionar los datos de una trayectoria en el sistema de control de flotas MovilWeb. Además, se habilita un grupo de análisis y la utilización de técnicas de minería de datos para buscar patrones y extraer información de los datos primarios, ⁽³⁴⁾ convirtiéndose en una potente herramienta de sistemas espaciales de apoyo a la decisión basados en datos.

Patrón de dinámicas de gobernanza de infraestructuras de datos espaciales

Después del auge principal en la popularidad de las IDE y más de 30 años desde que se articuló por primera vez su conceptualización, la comunidad geoespacial ha aprendido mucho sobre interoperabilidad y cooperación técnica y organizativa, aunque los resultados no siempre han sido los esperados. El panorama de las IDE a nivel mundial sigue siendo variado y complejo, y no es bien comprendido ni conocido por quienes están fuera de la comunidad geoespacial. ⁽³⁶⁾

Un aspecto que ha forzado a un cambio en los modelos de gobernanza de las IDE es la complejidad en cantidad y diversidad de actores que se involucran en la provisión y consumo de datos espaciales. Sobre la base en las transiciones de gobernanza clave encontradas en 2 casos de IDE analizados, ⁽¹²⁾ se concluyó que la gobernanza de las IDE es adaptativa, aunque, en el estudio la mayoría de los modelos de gobernanza no se mantuvieron por mucho tiempo, ya que no cumplían sus objetivos, no satisfacían a todos los actores o no estaban alineados con nuevas visiones e ideas.

En Sjoukema JW *et al.* ⁽¹³⁾ se ofrece un modelado basado en agentes y en las interacciones de gobernanza de IDE, desde un enfoque audaz e innovador. En la simulación del modelo a medida que se permite influir en las interacciones locales, se puede jugar con varias formas de interacción de gobernanza y examinar las compensaciones entre un enfoque de gobernanza jerárquico, de red o liberal. Se comprueba que es válido un enfoque de red para modelar las complejas interacciones entre las partes interesadas de una IDE. En cuanto a políticas presupuestarias, para el caso de estudio, la forma más sostenible resultó ser la que facilita continuidad de financiamiento, mientras el pago por uso tuvo casi el mismo éxito, y la política presupuestaria fragmentaria tuvo el peor desempeño.

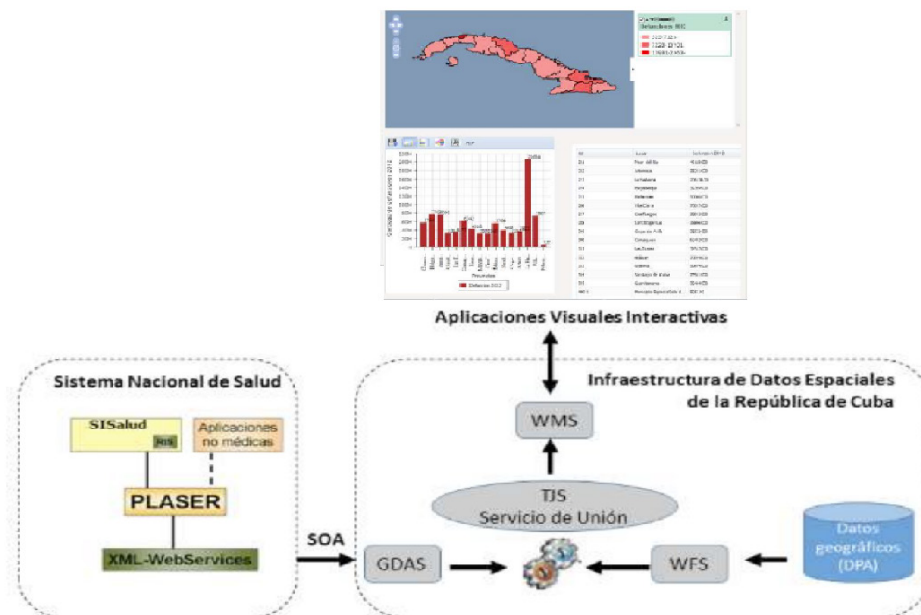


Fig. 4. Integración de estadísticas de salud con la IDERC ⁽³³⁾

Arquitectura de ecosistemas de datos basada en patrones evolutivos de infraestructuras de datos espaciales, tejidos de datos y grafos virtuales de conocimiento

A partir de los resultados obtenidos con los patrones evolutivos de IDE y tomando como referencia las tendencias en arquitecturas de datos descentralizadas *data mesh* ⁽¹⁶⁾ y *data fabric*, ⁽¹⁷⁾ así como los grafos de conocimiento, ⁽¹⁹⁾ se modeló una arquitectura de ecosistema de datos, ⁽³⁷⁾ instanciada al escenario geoespacial como prueba de concepto.

Las IDE por su naturaleza distribuida y su enfoque de datos como productos ofrecidos como interfaces de servicios basados en estándares, que garantizan interoperabilidad y accesibilidad desde la propia infraestructura de datos, tienen una similitud con algunos principios del concepto *data mesh* o tejido de datos. Entendiendo que *data mesh* es un diseño de arquitectura distribuida basada en dominios, donde se trata de “reconocer e identificar dominios de datos en una organización, así como construir una arquitectura basada en numerosos componentes que conforman estos dominios”, ⁽³⁸⁾ las IDE, en el marco de ecosistemas de datos, pueden aportar un tipo particular de dominio muy bien formalizado, con el objetivo de ofrecer productos de datos geoespaciales al ecosistema. El tejido de datos se obtiene entonces, a través de un grafo de conocimiento para integrar los productos de datos de dominios geoespaciales y de otros dominios no geoespa-

ciales, como los flujos de datos IoT provenientes de fuentes diversas, o datos estadísticos. En este enfoque se garantiza interoperabilidad entre los diferentes dominios, bajo una gobernanza computacional federada. Una arquitectura simplificada de ecosistema de datos basada en los patrones evolutivos de las IDE y soportada en la hibridación de arquitecturas de referencia de *data mesh* y *data fabric* con el uso de grafos de conocimiento, ⁽³⁷⁾ se muestra en la figura 5.

DISCUSIÓN

La experimentación realizada en el primer patrón con la migración de la IDERC a la nube (IDEaaS), evidenció una mejora significativa, expresada en una alta disponibilidad en los servicios y capacidades elásticas de escalado, lo cual garantizó el aumento del uso de los servicios y eliminó el problema de concurrencia de la IDE tradicional.

El modelo de servicios semánticos, analizado en el patrón relacionado con la *web* semántica, posibilitó la anotación de conocimiento en forma de restricciones de clases, lo cual ha podido ser utilizado para segmentar fuentes de datos respecto a sus atributos. Esta funcionalidad permite la recuperación de información vectorial teniendo en cuenta tales restricciones. Con la implementación del modelo en la IDERC se habilitaron futuros escenarios de aplicaciones semánticas, como el método de sistemas de recomendaciones semánticas basado en filtrado colaborativo, argumentado en el patrón que aporta sensibilidad al contexto, el cual tiene una versátil utilidad y fue demostrada

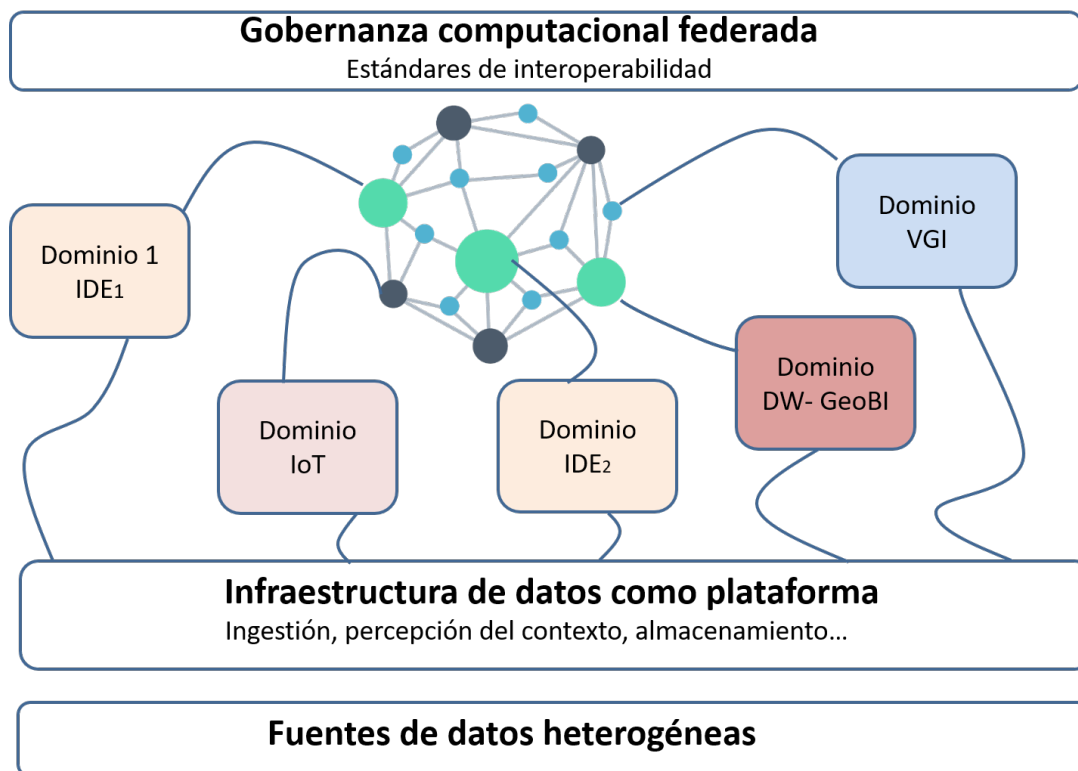


Fig. 5. Arquitectura de ecosistemas de datos basado en Data Mesh y grafos de conocimientos. Fuente: adaptada de Delgado-Fernández T⁽³⁷⁾

su precisión y eficacia para un escenario turístico, partiendo de trazas de GPS y preferencias del usuario simuladas.

Con la validación del modelo semántico XSCM_4_IOT, centrado en el usuario e independiente del dominio, que fuera validado en un escenario de *smart campus* en la Universidad Nacional de Colombia, se impacta en la resolución de los retos de sensibilidad al contexto, interoperabilidad y extensibilidad que plantean el dinamismo y diversidad de los datos y las condiciones de heterogeneidad de los ambientes IoT. Por otra parte, la publicación de la información contextual y de los recursos ontológicos en la nube de datos enlazados (LOD) enriquece la información del contexto local y, por estar compartida en este espacio con una descripción formal, tiene también el potencial de enriquecer otros recursos de la red. La extensión de este modelo para el tratamiento de flujos de datos de Internet de las cosas, con la inclusión del razonador semántico en tiempo real RTR_4_IoT, le otorgó capacidades de "inteligencia" al modelo original que cuenta con la característica de ser interoperable, jerárquico, escalable, configurable, extensible, adaptable y abierto, al ser diseñado usando exclusivamente tecnologías de código abierto pertenecientes al ecosistema para el tratamiento de *Big Data* de *Hadoop*.

Basado también en el escenario de datos complejos, el patrón de interrelación de datos geospaciales con *Big Data*, VGI e inteligencia de negocio refleja el impacto de las nuevas

arquitecturas en las soluciones a problemas de la sociedad, siendo un ejemplo de ello, su aplicación para denunciar incidencias criminales en la ciudad de Montevideo, y la integración de estadísticas de salud con la IDERC para obtener tableros digitales, con un gran potencial para la generación de mapas estadísticos de cualquier sector.

Para cerrar los patrones evolutivos se analizan las tendencias más modernas de gobernanza de IDE en el escenario de la iniciativa europea INSPIRE. El principal aporte en este sentido es el énfasis en las redes de actores y la autoorganización de tales iniciativas, así como la colaboración con una mayor inclusión de los usuarios.

El modelo de ecosistemas de datos expresado en la arquitectura basada en *data mesh* y *data fabric*-KG, absorbe y fusiona los patrones analizados y facilita la integración de datos complejos heterogéneos, abriendo nuevas oportunidades de potenciación del impacto de las IDE en el escenario dinámico de la transformación digital.

Conclusiones

Las IDE tradicionales ven limitadas sus capacidades de interconectarse de forma ágil y autosostenible debido a la heterogeneidad de datos y diversidad de actores presentes en la era actual. Esto junto a la emergencia de nuevas tecnologías

digitales está condicionando una transición hacia ecosistemas digitales ágiles, flexibles y autosostenibles.

Los patrones evolutivos de las IDE experimentados en la IDERC y en escenarios latinoamericanos, en particular Colombia y Uruguay, a partir de la incorporación de tecnologías habilitadoras de la transformación digital, sustentan el diseño de la nueva arquitectura de tejidos de datos geoespaciales, la cual responde a 4 principios fundamentales: propiedad del dominio, datos como producto, plataforma de autoservicio de datos y gobernanza computacional federada, y es la base para el potencial desarrollo de nuevos ecosistemas digitales en el contexto geoespacial.

La flexibilidad, autosostenibilidad y agilidad, requeridas en tales ecosistemas, ha forzado el estudio de las principales dinámicas de gobernanza de las IDE, manifestándose la importancia de asegurar su desarrollo en escenarios locales, como las ciudades inteligentes y sostenibles.

Dada la complejidad de los actores y el diverso panorama tecnológico y de datos, que caracteriza la era moderna, existe una tendencia de experimentación de la gobernanza en entornos dinámicos a partir de la simulación basada en agentes.

Una forma más flexible de gobernanza, especialmente con mayor énfasis en las redes y la autoorganización y una mayor inclusión de los usuarios, puede asegurar la adaptabilidad y sostenibilidad de los ecosistemas de datos.

Con esta contribución que ofrece una nueva arquitectura de ecosistemas de datos instanciada en tejidos de datos geoespaciales, fundamentada en patrones evolutivos de las IDE que responden al paradigma de la transformación digital, se abren nuevas oportunidades de investigación, desarrollo e innovación hacia la proliferación de ecosistemas digitales abiertos, flexibles, ágiles y autosostenibles.

REFERENCIAS BIBLIOGRÁFICAS

1. Masser I, Rajabifard A, Williamson I. Spatially enabling governments through SDI implementation. *International Journal of Geographical Information Science*. 2008 Jan 1;22(1):5-20. Disponible en <https://doi.org/10.1080/13658810601177751>
2. Delgado T. Infraestructuras de Datos Espaciales en países de bajo desarrollo tecnológico. Implementación en Cuba. 2005. Tesis doctoral, ITM, Comisión de Geodesia y Cartografía.
3. Delgado T, Cruz R. Construyendo Infraestructuras de Datos Espaciales a nivel local. CUJAE, La Habana, 2009. 130p.
4. Craglia M. Spatial Data Infrastructures. In: James Wright, editor. *International Encyclopedia of Social and Behavioral Sciences*. Elsevier; 2015. 130-5 p. JRC82372. Disponible en <http://dx.doi.org/10.1016/B978-0-08-097086-8.72058-7>
5. Minghini M, Sarretta A, Napolitano M. OpenStreetMap Contribution to Local Data Ecosystems in COVID-19 Times: Experiences and Reflections from the Italian Case. *Data*. 2022 Mar 31;7(4):39. Disponible en <https://doi.org/10.3390/data7040039>
6. Herrera F, Sosa R, Delgado T. GeoBI and big VGI for crime analysis and report. In 2015 3rd International Conference on Future Internet of Things and Cloud 2015 Aug 24:481-8. IEEE. Disponible en <https://ieeexplore.ieee.org/document/7300856>
7. Cromptvoets J, Rajabifard A, Van Loenen B, Fernández TD. A multi-view framework to assess SDIs. Wageningen University, RGI. 2008. Disponible en <https://eng.unimelb.edu.au/csdl/publications/multi-view-framework>
8. Grus L, Cromptvoets JW, Bregt AK, van Loenen B, Fernandez TD, Vandenbroucke D. Evaluating the application of the multi-view spatial data infrastructure assessment framework. *Journal of Spatial Science*. 2011 Jun 1;56(1):121-41. Disponible en <https://doi.org/10.1080/14498596.2011.567438>
9. Delgado T. La Infraestructura de Datos Espaciales de la República de Cuba, avances y perspectivas. Novena Conferencia Cartográfica Regional de las Naciones Unidas para América. Consejo Económico y Social de Naciones Unidas. Nueva York. E/CONF.99/IP.15, 2009. Disponible en https://mdgs.un.org/unsd/geoinfo/RCC/docs/rcca9/ip/9th_UNRCCA_econf.99_IP15.pdf
10. Delgado-Fernández T, Cromptvoets J. Evaluating Spatial Data Infrastructures in the Caribbean for sustainable development. In Proceedings of the 10th GSDI conference 2008 Feb 29:25-9. Disponible en <https://www.spatialist.be/download/pub/TS1.1paper.pdf>
11. Delgado, T., & Cromptvoets, J. Infraestructuras de datos espaciales en Iberoamérica y el Caribe. La Habana: IDICT. 2007. Disponible en http://redgeomática.rediris.es/idedes/IDEs_en_Iberoamerica.pdf
12. Sjoukema JW, Bregt A, Cromptvoets J. Evolving spatial data infrastructures and the role of adaptive governance. *ISPRS International Journal of Geo-Information*. 2017 Aug 16;6(8):254. Disponible en <https://doi.org/10.3390/ijgi6080254>
13. Sjoukema JW, Samia J, Bregt AK, Cromptvoets J. Governance interactions of spatial data infrastructures: an agent-based modelling approach. *International Journal of Digital Earth*. 2021 Jun 3;14(6):696-713. Disponible en <https://doi.org/10.1080/17538947.2020.1868585>
14. Delgado-Fernández T. Taxonomía de transformación digital. *Revista Cubana de transformación digital*. 2020;1(1):4-23. [Consultado 15 ago 2022]. Disponible en: <https://rctd.uic.cu/rctd/article/view/62>
15. Delgado T. Una arquitectura de Ecosistemas de Datos Espaciales. XVI Convención y Feria INFORMATICA 2016: Conectando sociedades 2016;1-6. ISBN 978-959-289-122-7
16. Dehghani Z. Data Mesh: Delivering Data-Driven Value at Scale (1.ed-preview version), O'Reilly Media, Inc. 2022. [Consultado 5 sep 2022]. Disponible en: <https://www.oreilly.com/library/view/data-mesh/9781492092384/>
17. IBM, "Data fabric," 2021. [Internet]. [Consultado 4 sep 2022]. Disponible en: <https://www.ibm.com/analytics/data-fabric>
18. Östberg PO, Vyhmeister E, Castañé GG, Meyers B, Van Noten J. Domain Models and Data Modeling as Drivers for Data Management: The ASSISTANT Data Fabric Approach. *IFAC-PapersOnLine*. 2022 Jan 1;55(10):19-24. [Consultado 4 sep 2022]. Disponible en: <https://doi.org/10.1016/j.ifacol.2022.09.362>
19. Hogan A, Blomqvist E, Cochez M, d'Amato C, de Melo G, Gutierrez C, Gayo JE, Kirrane S, Neumaier S, Polleres A, Navigli R. Knowledge graphs. arXiv preprint arXiv:2003.02320. 2020; Mar 4. [Consultado 2 ago 2022]. Disponible en: <https://arxiv.org/abs/2003.02320>

20. de Oliveira EF, Silveira MS. Open government data in Brazil a systematic review of its uses and issues. In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age 2018 may 30:1-9. Disponible en <https://doi.org/10.1145/3209281.3209339>
21. Kilov H, Linington PF, Romero JR, Tanaka A, Vallecillo A. The reference model of open distributed processing: Foundations, experience and applications. *Computer Standards & Interfaces*. 2013 mar 1;35(3):247-56. Disponible en <https://doi.org/10.1016/j.csi.2012.05.003>
22. Hjelmager J, Moellering H, Cooper A, Delgado T, Rajabifard A, Rapant P. An initial formal model for spatial data infrastructures. *International Journal of Geographical Information Science*. 2008 nov 1;22(11-12):1295-309. Disponible en <https://doi.org/10.1080/13658810801909623>
23. Cooper AK, Moellering H, Hjelmager J, Rapant P, Delgado T, Laurent D, ... A spatial data infrastructure model from the computational viewpoint. *International Journal of Geographical Information Science*. 2013 jun 1;27(6):1133-51. Disponible en <https://doi.org/10.1080/13658816.2012.741239>
24. Capote J. L. Modelo de Servicios Semánticos en la IDERC. Tesis doctoral. Instituto Técnico Militar José Martí. La Habana, 2011.
25. Delgado T. Semántica espacial y descubrimiento de conocimiento para desarrollo sostenible. La Habana: Cujae. 2009. ISBN: 978-959-261-282-2, 2009.
26. Cruz R. Modelo de Infraestructura de Datos Espaciales basada en computación en la nube. Tesis doctoral. La Habana: ITM José Martí. La Habana, 2011
27. Capote JL. Modelo de Servicios Semánticos en la IDERC. Tesis doctoral. Instituto Técnico Militar José Martí. La Habana, 2011.
28. González G. Método para sistemas de recomendaciones sensibles al contexto y basados en ontologías. Tesis doctoral. Instituto Técnico Militar José Martí. La Habana, 2014
29. González Suárez G, Delgado Fernández T, Capote Fernández JL, Cruz Iglesias R. Filtrado espacial, semántico y colaborativo para apoyar decisiones en entornos ubicuos. *Revista Cubana de Ciencias Informáticas*. 2015 jun;9(2):66-80. Disponible en http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2227-18992015000200005
30. Delgado T, Capote JL, Gonzalez G, Cruz R. Big Data & Geospatial Linked Open Data to generalize Context-Aware Recommender Systems. In Proc. of 27th International Cartographic Conference, Rio de Janeiro. Disponible en https://icaci.org/files/documents/ICC_proceedings/ICC2015/papers/7/533.html
31. Cangrejo-Aljure LD. Modelo semántico de contexto extendido con Linked Open Data para IoT. Tesis doctoral. Universidad Nacional de Colombia-Sede Bogotá. 2019. Disponible en <https://repositorio.unal.edu.co/handle/unal/75520>
32. Canto-Albarracín, AF. Extensión del modelo semántico de contexto XSCM_4_IoT para la gestión y el razonamiento de flujos de datos en tiempo real. Tesis de Maestría. Sistemas e Industrial. Universidad Nacional de Colombia-Sede Bogotá. 2020. Disonible en <https://repositorio.unal.edu.co/handle/unal/77533>
33. Aguilar I, Delgado T. Integración de estadísticas de salud con servicios de información geográfica: visualización geoespacial interactiva de estadísticas de mortalidad. VI Convención Internacional Informática 2016: Conectando sociedades, (pág. 8). La Habana. 2016. ISBN 978-959-289-122-7
34. Capote F. & Cruz R. Tecnologías de Big Data geoespacial en el Centro de Información Geoespacial de Geocuba. *Revista Cubana de Transformación Digital*, 1(2):75-84. Disponible en <https://rctd.uic.cu/rctd/article/view/48>
35. Hughes JN, Annex A, Eichelberger CN, Fox A, Hulbert A, Ronquest M. Geomesa: a distributed architecture for spatio-temporal fusion. In *Geospatial informatics, fusion, and motion video analytics V 2015 may 21;9473:128-40*. SPIE. Disponible en <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9473/1/GeoMesa-a-distributed-architecture-for-spatio-temporal-fusion/10.1117/12.2177233.short?SSO=1>
36. Coetzee S, McCormack B, Mohamed-Ghouse EZ, Scott AG. Towards a sustainable geospatial ecosystem beyond SDIs. 2021 EUROGI. [Consultado 4 sep 2022]. Disponible en https://ggim.un.org/meetings/GGIM-committee/11th-Session/documents/Towards_a_Sustainable_Geospatial_Ecosystem_Beyond_SDIs_Draft_3Aug2021.pdf
37. Delgado-Fernández T. Arquitectura de referencia de ecosistemas de datos basada en Data Mesh & Data Fabric. *Revista Cubana de Administración Pública y Empresarial*. 2022 nov 19;6(3):e249. Disponible en <https://dialnet.unirioja.es/servlet/articulo?codigo=8921779>
38. Hokkanen S. Utilization of Data Mesh Framework as a Part of Organization's Data Management (Master's thesis, Itä-Suomen yliopisto) 2021. Disponible en <https://erepo.uef.fi/handle/123456789/26142>

 Recibido: 20/02/2024

 Aprobado: 20/03/2024

Agradecimientos

Se agradece al Grupo Empresarial GEOCUBA por el apoyo que brindó a algunas de las investigaciones aquí incluidas. A la Comisión de Estándares de Datos Espaciales de la Asociación Cartográfica Internacional, en cuyas sesiones de trabajo se analizaron, discutieron y validaron algunos modelos conceptuales de infraestructuras de datos espaciales que se incluyen en este artículo.

Conflictos de intereses

Los autores declaran que no existen conflictos de intereses relacionados con el contenido del artículo con las instituciones a las que pertenecen ni con ninguna otra entidad o institución.

Contribuciones de los autores

Conceptualización: Tatiana Delgado Fernández, José Luis Capote Fernández, Joep Crompvoets, Rafael Cruz Iglesias, Libia Denisse Cangrejo Aljure, Guillermo González Suárez

Curación de datos: José Luis Capote Fernández, Rafael Cruz Iglesias, Libia Denisse Cangrejo Aljure, Guillermo González Suárez

Análisis formal: Tatiana Delgado Fernández, José Luis Capote Fernández, Joep Crompvoets, Rafael Cruz Iglesias, Libia Denisse Cangrejo Aljure, Guillermo González Suárez

Adquisición de fondos: Tatiana Delgado Fernández, Joep Crompvoets, Libia Denisse Cangrejo Aljure

Investigación: Tatiana Delgado Fernández, José Luis Capote Fernández, Joep Crompvoets, Rafael Cruz Iglesias, Libia Denisse Cangrejo Aljure, Guillermo González Suárez

Metodología: Tatiana Delgado Fernández, José Luis Capote

Fernández, Joep Crompvoets, Rafael Cruz Iglesias, Libia Denisse Cangrejo Aljure, Guillermo González Suárez

Administración del proyecto: Tatiana Delgado Fernández, Joep Crompvoets

Recursos: Tatiana Delgado Fernández, Joep Crompvoets, Libia Denisse Cangrejo Aljure

Software: José Luis Capote Fernández, Rafael Cruz Iglesias, Guillermo González Suárez

Supervisión: Tatiana Delgado Fernández, José Luis Capote Fernández, Joep Crompvoets

Validación: Tatiana Delgado Fernández, José Luis Capote Fernández, Joep Crompvoets, Rafael Cruz Iglesias, Libia Denisse Cangrejo Aljure, Guillermo González Suárez

Visualización: Tatiana Delgado Fernández, José Luis Capote Fernández, Joep Crompvoets, Rafael Cruz Iglesias, Libia Denisse Cangrejo Aljure, Guillermo González Suárez

Redacción-borrador original: Tatiana Delgado Fernández

Redacción-revisión y edición: Tatiana Delgado Fernández, José Luis Capote Fernández, Joep Crompvoets, Rafael Cruz Iglesias, Libia Denisse Cangrejo Aljure, Guillermo González Suárez

Financiamientos

El desarrollo de esta investigación fue favorecido por proyectos de investigación y desarrollo y para intercambios y estancias de investigación: Proyecto CYTED: (IDEDES) Evaluación y potenciación

de las Infraestructuras de Datos Espaciales de América Latina y el Caribe para desarrollo sostenible (Intercambios de investigadores 2006-2010); Departamento de Sistemas e Industrial, Facultad de Ingeniería de la Universidad Nacional de Colombia (recursos para estancias de investigación 2014-2019); Universidad de Melbourne (recursos para estancia de investigación 2013); Oficina Nacional de Hidrografía y Geodesia (recursos para el desarrollo de algunas de estas investigaciones).

Cómo citar este artículo

Delgado Fernández T, Capote Fernández JL, Crompvoets J, Cruz Iglesias R, Cangrejo Aljure LD, González Suárez G. Patrones evolutivos de infraestructuras de datos espaciales para el modelado de ecosistemas de datos. An Acad Cienc Cuba [internet] 2024 [citado en día, mes y año];14(1):e1539. Disponible en: <http://www.revistaccuba.cu/index.php/revacc/article/view/1539>

El artículo se difunde en acceso abierto según los términos de una licencia Creative Commons de Atribución/Reconocimiento-NoComercial 4.0 Internacional (CC BY-NC-SA 4.0), que le atribuye la libertad de copiar, compartir, distribuir, exhibir o implementar sin permiso, salvo con las siguientes condiciones: reconocer a sus autores (atribución), indicar los cambios que haya realizado y no usar el material con fines comerciales (no comercial).

© Los autores, 2024.

