

Aplicación de un Spatial Data Warehouse en la gestión de proyectos de vinculación: Caso de estudio

Spatial Data Warehouse applied to social projects management: Case study

Guillermo Omar Pizarro Vásquez^{1,*}, Vanessa Alexandra Jurado Vite^{1,†}, y Shirley Guisella Coque Villegas^{1,‡}.

¹Universidad Politécnica Salesiana, Ecuador.

gpizarro@ups.edu.ec, vjurado@ups.edu.ec, scoque@ups.edu.ec

Fecha de recepción: 09 de noviembre de 2017 — **Fecha de aceptación:** 12 de diciembre de 2017

Cómo citar: Pizarro Vásquez, G., Jurado Vite, V., & Coque Villegas, S. (2018). Aplicación de un Spatial Data Warehouse en la gestión de proyectos de vinculación: Caso de estudio. *Journal of Science and Research: Revista Ciencia e Investigación*, 3(9), 19-24. <https://doi.org/10.26910/issn.2528-8083vol3iss9.2018pp19-24p>

Resumen—La Dirección Técnica de Vinculación con la Sociedad de la Universidad Politécnica Salesiana, Sede Guayaquil, es la encargada de gestionar proyectos de vinculación que proporcionan el acercamiento entre Universidad y Sociedad, en conjunto con las Carreras. El desarrollo de estos proyectos no cuenta con la ayuda de un sistema de información geográfico para establecer los lugares donde se ha intervenido. Para brindar apoyo a la toma de decisiones en el proceso de la gestión de proyectos en sus etapas iniciales, se aplicó un Spatial Data Warehouse y se diseñó e implementó algunos cubos de datos espaciales, siguiendo la metodología HEFESTO, que permitieron la visualización escalar y geográfica de proyectos ejecutados; además, de la identificación de sectores vulnerables para el desarrollo de nuevos proyectos.

Palabras Clave—Análisis de datos, Visualización de datos, Sistemas de Información geográfica, Toma de decisiones.

Abstract—The Technical Direction of Linkage with the Society of Salesian Polytechnic University (UPS) in Guayaquil, is in charge of managing social projects that provide an approach between University and Society, with the Careers. Geographic Information System (GIS) is not used when they want to establish the places where they have been intervened. To support decision-making in the process of project management in its early stages, a Spatial Data Warehouse was applied and some spatial data cubes were designed and implemented, following HEFESTO methodology, which allowed scalar and geographic visualization of executed projects; In addition, the identification of vulnerable sectors for the development of new projects.

Keywords—Data analysis, Data visualization, Geography information systems.

INTRODUCCIÓN

Las Universidades a nivel mundial desean unir academia, investigación y vinculación, lo cual, es complicado; sin embargo, para algunas esto ya es un hecho, como ocurre en Estados Unidos o en la mayoría de las universidades europeas (De Gortari, 1994). Una de las formas para acercar la Universidad con la Sociedad es mediante la vinculación, principalmente a través de los proyectos de vinculación o también denominados como proyectos sociales.

La Universidad Politécnica Salesiana (UPS) del Ecuador desde sus inicios se ha interesado en los proyectos sociales; por ello, cuenta con una Secretaría Técnica de Vinculación con la Sociedad (STVCS), quien gestiona las actividades con entidades externas a la Universidad de las tres sedes (Guayaquil, Quito y Cuenca), en cada Sede existe una Dirección Técnica de Vinculación con la Sociedad (DTVCS) quien administra

los procesos de vinculación en conjunto con las Carreras que se ofrecen.

En los proyectos de vinculación que gestiona la DTVCS, existe una necesidad para diseñar, elaborar, ejecutar y evaluar proyectos de vinculación con la sociedad de una manera eficiente, eficaz y pertinente. Para cubrir esta necesidad se diseñó e implementó un Spatial Data Warehouse (SDW) que ayudará en la toma de decisiones, específicamente en las etapas iniciales dentro de la gestión de los proyectos, siendo este el principal aporte de esta investigación.

Una de las motivaciones para llevar a cabo este estudio está basado en la localización geográfica de zonas vulnerables que permitan visualizar dónde ejecutar nuevos proyectos de vinculación; además, de conocer otras vulnerabilidades en los lugares donde ya se han ejecutados proyectos.

La inclusión de las tecnologías de la información en la gestión de proyectos de vinculación no ha sido un tema considerado por el momento, como se puede verificar en Álvarez Rosero (2014), Montoya Ríos et al. (2012), Padilla Gómez (2013), Vizcaíno Figueroa (2013), donde el problema tan solo es abordado desde las ciencias sociales o administrativas, sin considerar a las tecnologías como parte de la solución.

*Ingeniero en Sistemas Computacionales, Magíster en Gestión de Bases de Datos

†Ingeniera en Sistemas Computacionales, Magíster en Gerencia de Tecnologías de la Información

‡Ingeniera en Sistemas Computacionales, Magíster en Sistemas de Información Gerencial

Este trabajo de investigación propone el uso de un SDW para la gestión de proyectos de vinculación. La aplicación o uso del SDW se encuentra en la literatura científica, como se puede ver en Astriani and Trisminingsih (2016), Hasanah and Trisminingsih (2016), , donde se verifica el uso de una herramienta ETL espacial y de un SDW. En esta revisión de trabajos relacionados, se tuvo en cuenta sólo los artículos que hayan utilizado software de tipo código abierto o software libre.

El procedimiento ETL espacial se ha utilizado en los siguientes campos de aplicación: la agricultura Oubennaceur (211) con Talend Open Studio + Spatial Data Integrator (TOS+SDI) o en estudios ambientales Hasanah and Trisminingsih (2016) con GeoKettle; en estos casos se diseñan y ejecutan procedimientos ETL automáticos. También se hace mención de un procedimiento ETL manual, como en Pandolfi González and Bogantes González (2014).

La visualización de datos espaciales mediante un Spatial Online Analytical Processing (SOLAP), también es utilizado, como se evidencia en Gonzalez and Gonzalez (2013) con el uso de GeoMondrian para el análisis de datos demográficos de Costa Rica, o en Hasanah and Trisminingsih (2016) con el uso de SpagoBI para presentar los puntos calientes de biodiversidad.

En este artículo de investigación se presenta un caso de estudio sobre cómo se llevó a cabo la aplicación del SDW en el problema antes descrito, mediante las siguientes actividades:

- a) especificación sobre la relación entre los datos del Censo de Población y Vivienda (1990, 2001 y 2010) y los proyectos de vinculación existentes;
- b) la aplicación de la metodología de HEFESTO en el diseño del cubo de datos espacial; y
- c) visualización del análisis espacial multidimensional.

Este artículo se organiza de la siguiente manera: en la sección 2, se detalla la metodología aplicada, los elementos utilizados; en la sección 3, se detalla el caso de estudio presentado en la introducción; en la sección 4, la correspondiente discusión y resultados del caso de estudio; y en la sección 5, conclusiones y trabajos futuros.

METODOLOGÍA

El enfoque de esta investigación es cualitativa (Hernández et al., 2010), debido a que se elabora un análisis sobre el caso de estudio de la aplicación del SDW en la gestión de proyectos de vinculación; además, el alcance es descriptivo debido a que se especifican las propiedades o características del proceso en la gestión de los proyectos mediante el diseño y aplicación del SDW. Las técnicas utilizadas son: observación y entrevista; que se utilizaron con los responsables tanto de la STVCS como de la DTVCS.

Materiales

A continuación, se elabora una breve descripción de los elementos utilizados en este trabajo:

1. Talend Open Studio (TOS) (Talend Open Studio, 2017): es un Software Open Source para los procedimientos automáticos ETL (Extraction – Transformation – Load),

el procedimiento ETL se fundamenta en la extracción de datos externos para cargarlos en un almacén de datos, entre la extracción y la carga se deben definir reglas de transformación para que el paso de los datos pueda ser factible (Vaisman and Zimányi, 2014). TOS tiene una extensión que permite la integración de datos espaciales con otros tipos de datos, denominado como Spatial Data Integrator (SDI) (Prunayre et al., 2007).

2. Spatial Data Warehouse (SDW): es el almacén de datos espacial que se encuentra implementado en PostgreSQL (Maymala, 2015), considerando la extensión PostGIS (Marquez, 2015) para gestionar datos espaciales. La arquitectura del SDW de este trabajo de investigación se basa en (Pizarro et al., 2017).
3. On Line Transactional Processing (OLTP) (Vaisman and Zimányi, 2014): es la fuente externa transaccional (PostgreSQL + PostGIS) donde se almacenan los proyectos de vinculación registrados en la DTVCS, estos proyectos se encuentran georeferenciados; es decir, existe una coordenada geográfica relacionada a cada lugar donde se ejecutó el proyecto.
4. Spatial On Line Analytical Processing (SOLAP) (Vaisman and Zimányi, 2014): la herramienta SOLAP permite visualizar geográficamente la información almacenada en un SDW. Esta herramienta hace uso del Software SpagoBI (S.P.A. and I., 2017) y GeoServer (Foundation, 2014). Si se desea visualizar solo datos escalares, se puede contar con un OLAP (On Line Analytical Processing).
5. SpagoBI Server (Cazzin, 2012): es un software de código abierto para Inteligencia de Negocios que permite visualizar datos espaciales desde un SDW, en su componente de Location Intelligence. En esta investigación se utilizó SpagoBI Server + Location Intelligence como herramienta SOLAP, debido a que existe una comunidad de programadores que respaldan su mantenimiento, la facilidad de instalación y es usable para el usuario final.
6. Metodología HEFESTO (Bernabeu, 2010): proporciona los pasos a seguir en el análisis multidimensional, diseño e implementación de un Data Warehouse, como se muestra en la Figura 1, consta de cuatro fases: análisis de requerimientos, análisis de los OLTP (On Line requerimientos, análisis de los OLTP (On Line Transaction processing), modelo lógico del almacén de datos y la integración de datos. Estos pasos se ejecutan de manera secuencial.
7. Cubo de Datos Espacial: elemento que permite la extracción y visualización de datos desde un SDW hacia un SOLAP, su diseño se fundamenta en la metodología HEFESTO y la notación para su diagramación se basa en (Boulil et al., 2014).
8. Datos: Se tiene acceso a datos proporcionados por el INEC; tales como: a) archivos en formato "shapefile" de la provincia del Guayas de los años 2001 y 2010 (cantón y parroquias); además, de las zonas y sectores de 1990, 2001 y 2010 (INEC, 2017a); b) los datos de los censos de población y vivienda de la provincia del Guayas de los años 1990, 2001 y 2010 (INEC, 2017b). Como se



Figura 1. Metodología de HEFESTO.
Fuente: (Bernabeu, 2010).

puede verificar en la Figura 2. Los datos de los proyectos de vinculación ejecutados, fueron proporcionados por DTVCs de la UPS, Sede Guayaquil, de los años 2014 y 2015.

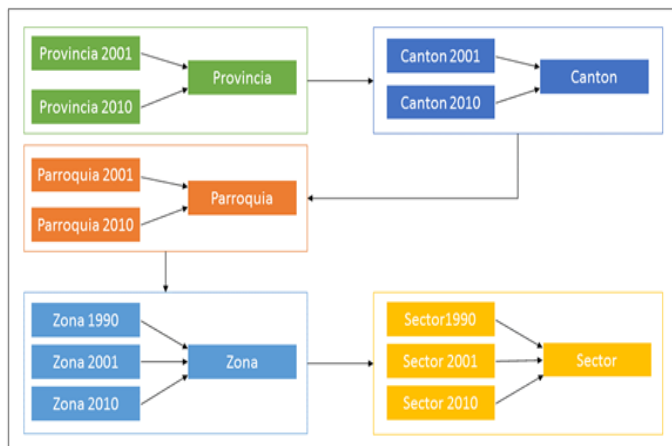


Figura 2. Cartografía del Censo de Población y Vivienda.
Fuente: (Pizarro Vásquez, 2017).

CASO DE ESTUDIO

Gestión de los Proyectos de Vinculación.-

Para la adecuada aplicación del SDW, fue necesario analizar la gestión de los proyectos de vinculación y su especificación se diagramó mediante el estándar SPEM 2.0 (Software & Systems Process Engineering Metamodel) (Domínguez and Bolaños, 2015); de tal manera, que se puede conocer en qué etapas se podría utilizar el SDW.

En la Figura 3, se puede determinar que el uso del SDW se puede aplicar en los pasos iniciales de la gestión de proyectos de vinculación, específicamente en: a) la búsqueda de idea de proyecto y en b) la elaboración de propuesta de proyecto, como se puede verificar en la Figura antes mencionada.



Figura 3. Elaboración de Propuestas de Proyectos de Vinculación.
Fuente: (Pizarro Vásquez, 2017).

Uso de la arquitectura de un Spatial Data Warehouse.-

La arquitectura está basada en Pizarro et al. (2017) y se tiene en cuenta los siguientes elementos:

- Fuente de Datos Heterogéneas:** se considera la base de datos donde se encuentra almacenados los proyectos de vinculación debidamente georeferenciados; además, de la cartografía censal (provincias, cantones, parroquias, zonas y sectores) en formato shapefile y los archivos del Censo de Población y Vivienda, en formato SPSS.
- Procedimiento ETL:** se utilizó la herramienta TOS + SDI para el paso de los datos al SDW; sin embargo, debido a que TOS tiene aún problemas para el procesamiento de datos en formato SPSS, se hizo la carga de archivos de los censos a una Base de Datos en PostgreSQL mediante un script en R, luego desde TOS se accedió a la base de datos reciente y se realizó el correspondiente procedimiento ETL automático para el paso de los datos al SDW.
De manera general, el proceso ETL ejecutado, contó con los siguientes pasos secuenciales: 1) mapeo preliminar; 2) conversión de datos; 3) mapeo de datos; 4) unión de datos; 5) datos procesados.
Luego de la selección de los datos de origen se detalla lo siguiente: en el mapeo preliminar, se realizaron las validaciones y separaciones de campos que se encontraban en una sola cadena de caracteres; en la conversión de los datos, se cambió el tipo de dato de los campos, según el diseño preliminar de los cubos de datos espaciales; en el mapeo de datos, se organizaron los datos para la unificación; en la unión de datos, se agrupan los datos de diferentes años; finalmente, en una sola tabla se registraron los datos procesados.
- Spatial Data Warehouse:** después del procedimiento ETL se conforma el SDW, para el correspondiente análisis multidimensional que se concreta en la creación de un cubo de datos espacial.
- SOLAP:** herramienta que permite visualizar los datos espaciales mediante un mapa que tiene información escalar asociada. En este trabajo también se permitió el acceso a la funcionalidad OLAP del SpagoBI Server para la respectiva revisión de datos escalares.

Diseño del Cubo de Datos Espacial mediante HEFESTO.-

Uno de los retos fue encontrar la manera de vincular los proyectos sociales con los datos censales. Existen investigaciones previas que consideraron los datos censales del Ecuador con una visión social utilizando las tecnologías de la información, como fue la de Molina and Osorio (2014) mostrando geográficamente un índice de condiciones de vida.

Nuestro análisis consistió en la selección de algunos de los programas de vinculación definidos por STVCS de la UPS (Universidad Politécnica Salesiana, 2016), que serían los siguientes: - profesionalización de jóvenes indígenas y afrodescendientes; - atención a población en estado de vulnerabilidad.

Se seleccionaron los dos programas antes mencionados debido a que son los que se encuentran relacionados con los datos del Censo de Población y Vivienda del INEC. Luego de elegir dos de cuatro programas dentro del proceso de servicios universitarios de la línea de intervención fortalecimiento organizacional; a continuación, la selección de algunos atributos del censo poblacional y vivienda; de esta manera, se procede al diseño de los cubos de datos espaciales del SDW que ayudaron a la toma de decisiones.

Es importante mencionar que se utilizó una operación geográfica de intersección para conocer en qué zonas censales (polígonos), se encontraban los proyectos sociales (puntos) finalizados. La intersección se lo hizo mediante TOS+SDI.

En la Figura 4 se muestra el diseño del cubo de datos espacial de los Proyectos de Vinculación, utilizando la notación propuesta en Boulil et al. (2014). Adicional a este cubo, se generó también los cubos de Población y Hogar.

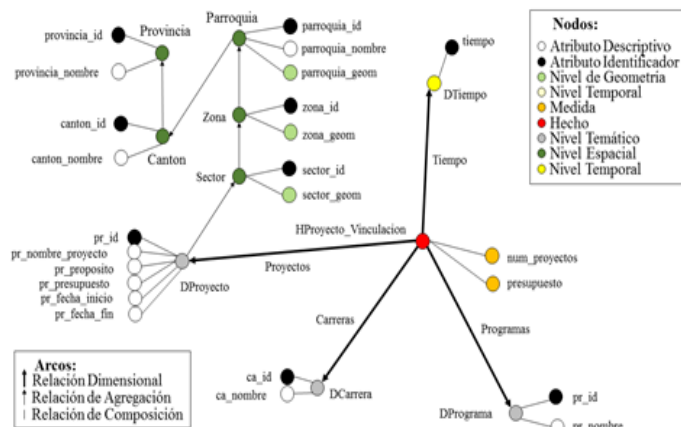


Figura 4. Diseño del Cubo de Datos Espacial de los Proyectos de Vinculación.

Fuente: Elaboración Propia.

Cubo de Datos Espacial y la implementación de los Reportes.-

Una vez que se tiene el SDW cargado con los datos antes descritos, se procede con el diseño de los cubos de datos espaciales y los reportes, desde la suite de SpagoBI Server, la debida configuración se encuentra documentada en Cazzin (2012).

Para los cubos de datos fue necesario el uso de SpagoBI Meta que se conecta al SpagoBI Server mediante el SDK de SpagoBI; en cambio, para los reportes, fue necesario el uso de SpagoBI Studio mediante el uso de iReport (TIBCO Software Inc, 2017).

Configuración del SOLAP.-

Para que se pueda visualizar el cubo de datos espacial fue necesario configurar las herramientas disponibles en SpagoBI + Location Intelligence como se encuentra especificado en Cazzin (2012)

RESULTADOS Y DISCUSIÓN

Los reportes obtenidos pueden ser de diferentes diseños, desde la vista del cubo de datos espacial se puede mostrar solo los datos escalares; además, de su respectivo informe gráfico.

Sin embargo, también se puede visualizar el cubo de datos espacial desde un mapa, como se presenta en la Figura 5, donde se muestran las zonas censales (norte de Guayaquil); en las cuales, se han ejecutado varios proyectos de vinculación por parte de la UPS, Sede Guayaquil. El color se gradúa de acuerdo al número de proyectos, es decir, donde existe un número máximo de proyectos se encuentra de color verde; donde hay menos, en amarillo.

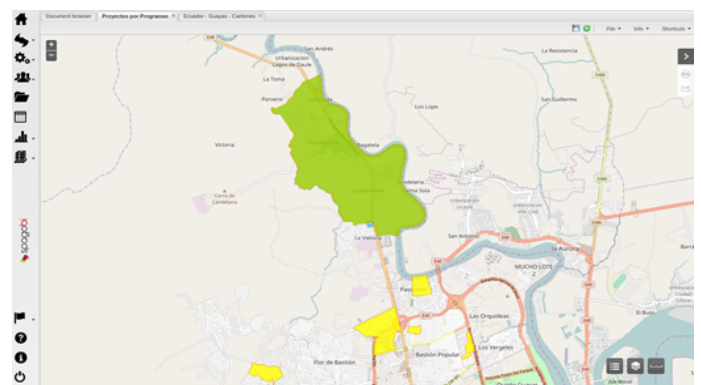


Figura 5. Mapa de Proyectos de Vinculación en el norte de Guayaquil.

Fuente: Elaboración Propia.

Al dar click en una de las zonas censales, se muestra un reporte con el detalle de información del Censo Poblacional con los atributos pertinentes, adicional a esto, se muestra una lista de los proyectos de vinculación ejecutados en ese lugar como se muestra en la Figura 6.

Este trabajo fue presentado a la representante de la DTVCS y de la STVCS quienes realizaron los siguientes comentarios puntuales: - “es una herramienta que ayudará a redactar la línea base de los proyectos que se vayan a elaborar a futuro”, este comentario es debido a que se pueden ver los datos del INEC en las zonas con información de interés para la DTVCS como se muestra en la Fig. 8; - “efectivamente puedo visualizar zonas donde existe población en estado de vulnerabilidad para tenerlos en cuenta en futuros proyectos”, esto debido a los datos que el INEC proporciona, ayuda

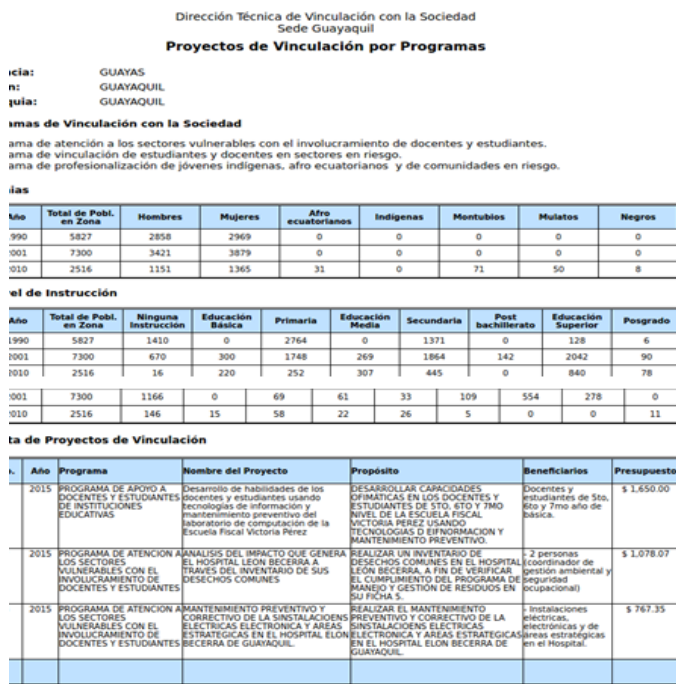


Figura 6. Reporte de relación entre datos del INEC con los proyectos de vinculación.

Fuente: Elaboración Propia.

a localizar algunos sectores vulnerables de los definidos en Constituyente (2008); tales como: personas con discapacidad, adultos mayores, niños, niñas y adolescentes.

CONCLUSIONES

Se diseñó e implementó un Spatial Data Warehouse para la visualización geográfica de proyectos de vinculación con la sociedad para la Dirección Técnica.

La Dirección Técnica ha logrado identificar sectores vulnerables en donde se pueden ejecutar nuevos proyectos de acuerdo a las características del sector o zona censal.

La información que proporciona el Spatial Data Warehouse, puede ser utilizada en la creación de nuevos proyectos, dando a los directores de proyectos, la oportunidad de tener un marco de trabajo referencial, o línea base como lo denominaron, sobre el entorno en donde se va a intervenir y su pertinencia con respecto a la Carrera que ejecuta el proyecto.

Debido a la información que brinda el Spatial Data Warehouse es posible determinar qué tipo de proyectos serían convenientes para el sector, además del número de personas a las que podría beneficiar el futuro proyecto, en concordancia con las Carreras ofertadas.

El procedimiento ETL fue automático y permitió que pueda ser ejecutado las veces que fueran necesarias sin necesidad de aplicar alguna técnica manual.

Como futuros trabajos se podrían aplicar técnicas de “machine learning”, específicamente los algoritmos de aprendizaje no supervisado, mediante redes neuronales para la debida clasificación de datos y así poder visualizar los “clusters” de manera geográfica y encontrar alguna relación existente.

AGRADECIMIENTOS

Este trabajo de investigación se realizó gracias a la colaboración de la Dirección Técnica de Vinculación con la Sociedad de la Universidad Politécnica Salesiana, sede Guayaquil; y al Vicerrector de la Sede.

REFERENCIAS BIBLIOGRÁFICAS

Álvarez Rosero, S. C. (2014). Diseño de un modelo de gestión por procesos para el área de Vinculación con la Colectividad de la Universidad Politécnica Estatal del Carchi.

Astriani, W. and Trisminingsih, R. (2016). Extraction, transformation, and loading (etl) module for hotspot spatial data warehouse using geokettle. *Procedia Environmental Sciences*, 33:626–634.

Bernabeu, R. D. (2010). Hefesto—data warehousing: Investigación y sistematización de conceptos-hefesto: Metodología para la construcción de un data warehouse. *Córdoba, argentina. Córdoba, Argentina.*

Boullil, K., Bimonte, S., and Pinet, F. (2014). Spatial olap integrity constraints: From uml-based specification to automatic implementation: Application to energetic data in agriculture. *Journal of Decision Systems*, 23(4):460–480.

Cazzin, G. (2012). Business intelligence with spagobi. *Italy, Padua.*

Constituyente, E. A. (2008). Constitución de la república del Ecuador.

De Gortari, R. (1994). La vinculación, parte de las políticas universitarias. *MÁ Campos y L. Corona (Coords.), Universidad y vinculación: Nuevos retos y viejos problemas*, pages 31–44.

Domínguez, V. H. M. and Bolaños, M. E. C. (2015). Spem: Software process engineering metamodel. *Revista Latinoamericana de Ingeniería de Software*, 3(2):92–100.

Foundation, O. S. G. (2014). GeoServer.

Gonzalez, D. B. and Gonzalez, L. P. (2013). Spatial data warehouses and solap using open-source tools. In *Computing Conference (CLEI), 2013 XXXIX Latin American*, pages 1–12. IEEE.

Hasanah, G. U. and Trisminingsih, R. (2016). Multidimensional analysis and location intelligence application for spatial data warehouse hotspot in indonesia using spagobi. In *IOP Conference Series: Earth and Environmental Science*, volume 31, page 012011. IOP Publishing.

Hernández, R., Fernández, C., and Baptista, P. (2010). Metodología de la investigación (5º ed.). México df: Ed.

INEC (2017a). Cartografía Estadística.

INEC (2017b). Censo de Población y Vivienda.

Marquez, A. (2015). *PostGIS Essentials*. Packt Publishing Ltd.

Maymala, J. (2015). *PostgreSQL for data architects*. Packt Publishing Ltd.

Molina, D. O. and Osorio, P. (2014). Segregación socioespacial urbana en cuenca, Ecuador. *Análitika: revista de análisis estadístico*, (8):27–38.

Montoya Ríos, F. J. et al. (2012). *Elaboración de una propuesta metodológica para la generación de ideas de proyectos a ser desarrollados a través de las prácticas universitarias*

- de la Facultad de Minas de la Universidad Nacional de Colombia Sede Medellín. PhD thesis, Universidad Nacional de Colombia, Sede Medellín.
- Oubennaceur, K. (2011). Structuration d'une base de données sur l'agriculture en Rhône-Alpes. *Université Jean Monnet de Saint-Etienne*.
- Padilla Gómez, R. R. (2013). Evaluación del grado de vinculación con la comunidad mediante los proyectos de tesis de los egresados de la universidad estatal de milagro. B.S. thesis, Universidad de Guayaquil. Unidad de Post-Grado Investigación y Desarrollo. Maestría en Docencia y Gerencia en Educación Superior.
- Pandolfi González, L. and Bogantes González, D. (2014). Geocr: una aplicación web de análisis dinámico para el soporte de decisiones basadas en datos convencionales y espaciales.
- Pizarro, G., Jurado, V., and Coque, S. (2017). Arquitectura de un Almacén de Datos Espacial. Guayaquil.
- Pizarro Vásquez, G. O. (2017). Diseño de un "spatial data warehouse" su incidencia en la generación de nuevos proyectos en la dirección técnica de vinculación con la sociedad de la universidad politécnica salesiana, sede guayaquil. Master's thesis, Universidad Técnica de Ambato. Facultad de Ingeniería en Sistemas, Electrónica e Industrial. Dirección de Posgrado. Maestría en Gestión de Bases de Datos III Versión.
- Prunayre, F. X., Chartier, B., Coudert, M., and Jacolin, Y. (2007). Spatial extension for Talend.
- S.P.A. and I., E. I. (2017). SpagoBI.
- Talend Open Studio (2017). Data Integration: Talend Enterprise Data Integration Services.
- TIBCO Software Inc (2017). iReport Designer — Jaspersoft Community.
- Universidad Politécnica Salesiana (2016). Plan Integrado de Vinculación con la Sociedad.
- Vaisman, A. and Zimányi, E. (2014). *Data warehouse systems*. Springer.
- Vizcaíno Figueroa, J. J. (2013). Diseño de un modelo de gestión para la vinculación de las instituciones de educación superior con la comunidad.