Revista Facultad de Ingeniería

Journal Homepage: https://revistas.uptc.edu.co/index.php/ingenieria
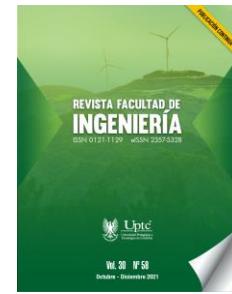
Business Intelligence for the Programs of the Secretaries of Health, Education and Planning in a Territorial Entity

María-Alejandra Varona-Taborda[1]

Jorge-Cesar Mosquera-Ramírez[2]

César-Augusto Medina-Moreno[3]

Diego-Fernando Lemus-Muñoz[4]

Carlos-Julián Muñoz-Hernández[5]

Christian-Gustavo Arias-Iragorri[6]

[1] M. Sc. Institución Universitaria Colegio Mayor del Cauca (Popayán-Cauca, Colombia). mvarona@unimayor.edu.co. ORCID: https://orcid.org/0000-0001-9545-4766

[2] Institución Universitaria Colegio Mayor del Cauca (Popayán-Cauca, Colombia). jcmosquera@unimayor.edu.co. ORCID: https://orcid.org/0000-0003-1655-2152

[3] Institución Universitaria Colegio Mayor del Cauca (Popayán-Cauca, Colombia). cmedina@unimayor.edu.co. ORCID: https://orcid.org/0000-0002-2335-8975

[4] Institución Universitaria Colegio Mayor del Cauca (Popayán-Cauca, Colombia). dlemus@unimayor.edu.co. ORCID: https://orcid.org/0000-0002-5066-0635

[5] Institución Universitaria Colegio Mayor del Cauca (Popayán-Cauca, Colombia). cjmunoz@unimayor.edu.co. ORCID: https://orcid.org/0000-0003-1982-802X

[6] Ph. D. Institución Universitaria Colegio Mayor del Cauca (Popayán-Cauca, Colombia). charias98@unimayor.edu.co. ORCID: https://orcid.org/0000-0002-4557-1611

Abstract

Territorial entities in Colombia are bound by the law to register and report to different instances the control information of the government programs they administer. However, so much information is distributed on various state and owned platforms that the result is processed and generated in different formats. This situation makes the comprehensive management of territorial data difficult, since, although the information exists, it is isolated, and its analysis is carried out independently by each party responsible for the process. The objective of this research is the Implementation of a business intelligence model that allows integration and analysis of data from the programs of the Health, Education, and Planning ministries for a Territorial Entity. Ralph Kimball's methodology was used, implementing a star topology model in the Datamart using MySQL as a database manager, an ETL system was built with the Pentaho tool which allows the extraction, transformation, and loading of the data in the Datamart. The cubes, reports and Dashboard are obtained with the management of tools such as Pentaho and Power BI, thus it is possible to make a correct interpretation of the resulting information. After applying Business Intelligence, it is possible to generate an adequate analysis of the information, allowing decision-making and application of new strategies to solve specific problems using control panels, visualization of indicators and generation of reports.

Keywords: business intelligence; datamart; ETL; Kimball methodology; territorial entity.

Inteligencia de negocios para los programas de las secretarías de salud, educación y planeación en un ente territorial

Resumen

Los entes territoriales en Colombia por ley deben registrar y reportar a diferentes instancias la información de control de los programas gubernamentales que administran. Sin embargo, es tanta la información distribuida en diversas plataformas estatales y propias, que el resultado es procesado y generado en

diferentes formatos. Esta situación dificulta el manejo integral de los datos territoriales, pues, aunque la información existe, se encuentra aislada y su análisis se realiza de manera independiente por cada parte responsable del proceso. El objetivo de esta investigación es la Implementación de un modelo de inteligencia de negocios que permite integración y análisis de los datos de los programas de las secretarías de Salud, Educación y Planeación para un Ente Territorial. Se empleó la metodología de Ralph Kimball implementando un modelo de topología estrella en el Datamart; utilizando como gestor de base de datos MySQL se construye un sistema ETL con la herramienta Pentaho que permite la extracción, transformación y carga de los datos en el Datamart. Se obtienen los cubos, reportes y Dashboard con el manejo de herramientas como Pentaho y Power BI y de este modo es posible realizar una correcta interpretación de la información resultante. Después de aplicar Inteligencia de Negocios se logra generar un adecuado análisis de la información, permitiendo la toma de decisiones y aplicación de nuevas estrategias para dar solución a problemas específicos mediante la utilización de tableros de control, visualización de indicadores y generación de reportes.

Palabras clave: datamart; ente territorial; ETL; inteligencia de negocios; metodología Kimball.

Inteligência de negócios para os programas das secretarias de saúde, educação e planejamento de um ente territorial

Resumo

As entidades territoriais na Colômbia por lei devem registrar e relatar a diferentes instâncias as informações de controle dos programas de governo que administram. Porém, tanta informação é distribuída em vários estados e plataformas próprias que o resultado é processado e gerado em diferentes formatos. Esta situação dificulta a gestão integral dos dados territoriais, pois, embora a informação exista, é isolada e a sua análise é efectuada de forma independente por cada um dos responsáveis pelo processo. O objetivo desta pesquisa é a implementação de um modelo de business intelligence que permita

a integração e análise de dados dos programas dos ministérios da Saúde, Educação e Planejamento de uma Entidade Territorial. Foi utilizada a metodologia de Ralph Kimball, implementando um modelo de topologia em estrela no Datamart; Utilizando o MySQL como gerenciador de banco de dados, é construído um sistema ETL com a ferramenta Pentaho que permite a extração, transformação e carregamento de dados no Datamart. Os cubos, relatórios e Dashboard são obtidos com a gestão de ferramentas como Pentaho e Power BI e desta forma é possível fazer uma interpretação correta da informação resultante. Após a aplicação de Business Intelligence, é possível gerar uma análise adequada das informações, permitindo a tomada de decisões e aplicação de novas estratégias para resolução de problemas específicos por meio da utilização de painéis de controle, visualização de indicadores e geração de relatórios.

Palavras-chave: datamart; entidade territorial; ETL; inteligência de negócios; metodologia Kimball.

María-Alejandra Varona-Taborda; Jorge-Cesar Mosquera-Ramírez; César-Augusto Medina-Moreno; Diego-Fernando Lemus-Muñoz; Carlos-Julián Muñoz-Hernández; Christian-Gustavo Arias-Iragorri

I. INTRODUCTION

In Colombia, the municipalities have in their administration different undersecretariats or portfolios for the management and fulfillment of the objectives to be developed by their elected representative, these are executed through the development or government plan. Each portfolio must have relevant information for decision-making and generation of impact projects for the population of the municipality, being the contexts of education, health, and social programs primary sources of information to generate management indicators that define part of the situational status of a specific population or population group. These are also determining factors in the social transformation and analysis of indicators of progress in overcoming poverty, improving the conditions of the vulnerable population, schooling, improving the population's health conditions, among others.

Currently, each program handles and manages the information from different platforms, systems, and structures in its different portfolios, independently of the management processes provided by the departmental and national government. In addition, all this information has as its main core the data provided by citizens, therefore, the territorial entity does not have the possibility of unifying, relating, and permanently managing this information. The provision of public services, programs or social projects requires the management of large volumes of data and information for the benefit of the community, which leads to the need of using tools that facilitate data extraction, processing and analysis, to facilitate the decision making in government agencies; where the main axis is the information provided by the citizen, which includes personal data, location, socioeconomic characterization, provision of health services, education and benefits of government programs, showing that the information is administered inefficiently, using manual methods, office automation tools between different information systems, both internal and external, with different data structures; this is why the quality and homogeneity of the data is not a constant, which makes the query, crossing and reporting of information between the different dependencies complex. Hence, it is important that the territorial entities begin to use Business Intelligence tools, which eases the consolidation, visualization, and analysis of information by

administrative personnel, generating reports in real time that facilitate the analysis of results, the identification of variations in information, facilitate decision-making and the implementation of strategies aimed at the welfare of the population.

The importance that the public sector needs to advance in preparing to exploit the data that entities constantly generate and collect and recognizing that these are a strategic asset of the nation [1] is very clear. For this reason, a Business Intelligence solution can contribute significantly to the government's development plan.

II. METHODOLOGY

This article presents a Business Intelligence solution aligned to descriptive research under the methodology developed by Ralph Kimball. Figure 1 shows the proposed life cycle.
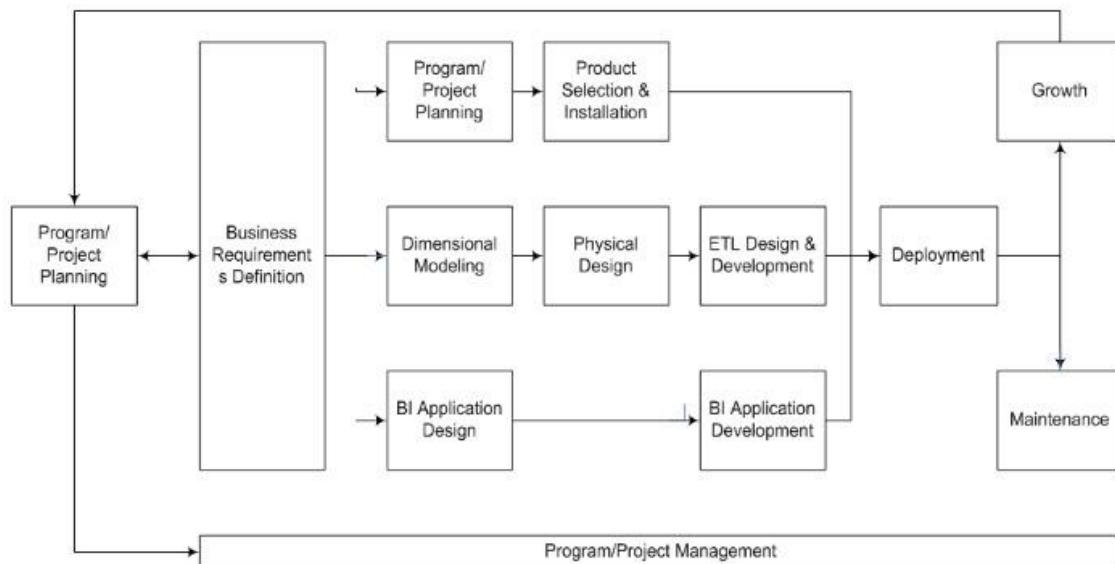
Fig. 1. Datawarehouse Life Cycle Flow [2].

For the development of a Data Warehouse / Business Intelligence solution, the following tasks were performed [3]:

1. Planning
2. Analysis of requirements

María-Alejandra Varona-Taborda; Jorge-Cesar Mosquera-Ramírez; César-Augusto Medina-Moreno; Diego-Fernando Lemus-Muñoz; Carlos-Julián Muñoz-Hernández; Christian-Gustavo Arias-Iragorri

3. Dimensional modeling
4. Physical Design
5. ETL system design
6. BI application specification and development

This article presents important aspects that emerged after performing the tasks defined in the chosen methodology. Table 1 shows the matrix obtained with the facts (metrics) and the corresponding dimensions (normalization) [4] according to the established requirements.

Table 1. Matrix of facts and dimensions.

Name	Description	Time	Location	Educational Institution	Lifecycle	Health_Dx (Purpose)
PSOCIALES_EDUCACION_1	Number of student population by location, educational institution and age range, beneficiaries of the Families in Action social program.	X	X	X	X	
SALUD_EDUCACION_1	Amount of student population by educational institution and age range that received promotion and prevention activities.	X	X	X	X	
SALUD_EDUCACION_2	Total number of adolescent pregnancies in the student population.	X		X	X	X
PSOCIALES_SALUD_1	Number of underage population by location, population group and age range affected by malnutrition belonging to the social program families in action.	X	X		X	X

Figure 2 represents the dimensional model which describes 4 facts and 5 relevant dimensions in the study [5].
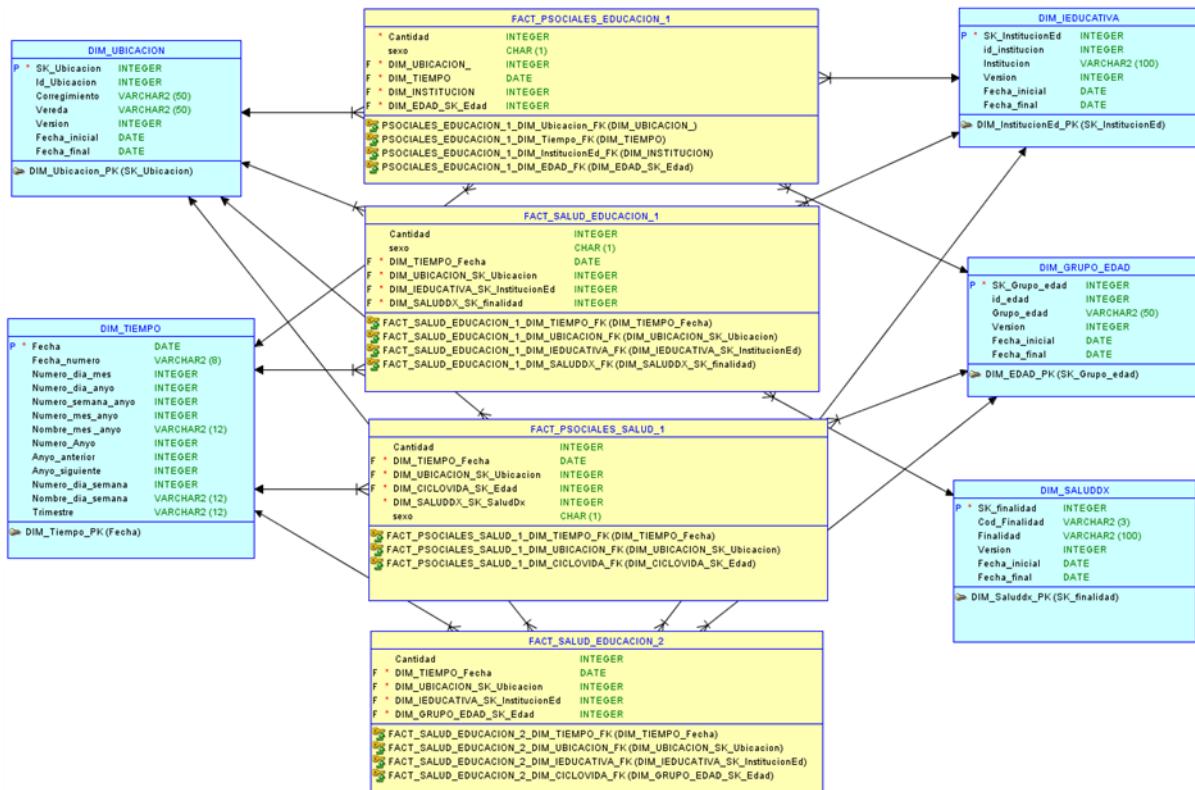

Fig. 2. Dimensional model.

III. RESULTS

This chapter represents the resulting Cubes, Reports and Dashboards using the Pentaho tool [6].

Table 2 shows the dimensions, hierarchies, and levels of Cube 1, which allows the generation and visualization of the historical information of the beneficiaries of Families in Action enrolled in the educational institutions of the municipality under analysis.

Table 2. Dimensions, hierarchies, and levels. Cube 1.

Cube name	Cube 1	
Description	Number of student population by location, Educational Institution, and age range beneficiaries of the Families in Action Program.	
Measures	Quantity	
Dimensions	Hierarchy	Levels

María-Alejandra Varona-Taborda; Jorge-Cesar Mosquera-Ramírez; César-Augusto Medina-Moreno; Diego-Fernando Lemus-Muñoz; Carlos-Julián Muñoz-Hernández; Christian-Gustavo Arias-Iragorri

Dim_Time	Time	Year, Trimester; Month
Dim_Location	Location	Corregimiento, Vereda
Dim_Lifecycle	Lifecycle	Age Group

Figure 3 shows the configuration of dimensions, hierarchies, and levels of Cube 1.
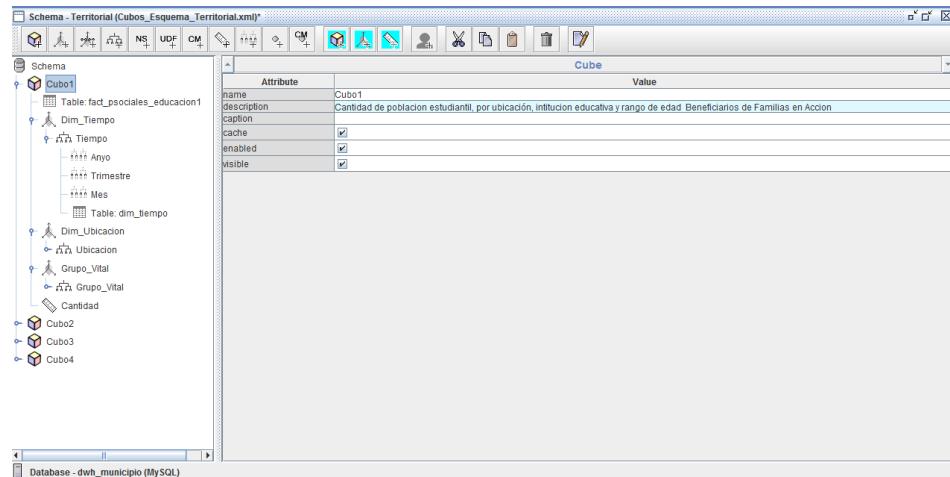
Fig. 3. Configuration of dimensions, hierarchies, and levels of Cube 1.

Figure 4 shows the elaboration of the Cube 1 report where the PIVOT-4J Plugin for Pentaho Server was used, which allows to elaborate different types of reports and use of the levels of the hierarchies of the built cube.

Fig. 4. PIVOT-4J view report - Cube 1.

Figure 5 shows the control panel with the total population records corresponding

to census lists by reservation and / or community initially uploaded and processed through the Extraction, Transformation and Loading (ETL) process [7], where the respective filter was applied, extracting the inconsistent records to which the respective inconsistency label was assigned through a transformation process with the Pentaho Data Integration tool. For a total of 32407 initial records, 1095 inconsistencies were obtained, out of which 768 had duplicate documents.

Fig. 5. Information debugging Dashboard.

Figure 6 shows the Dashboard with the results related to the population of the Families in Action social program that are currently studying, determined by year, gender, age, and educational institution. A total of 8465 enrolled in Educational Institutions was determined.
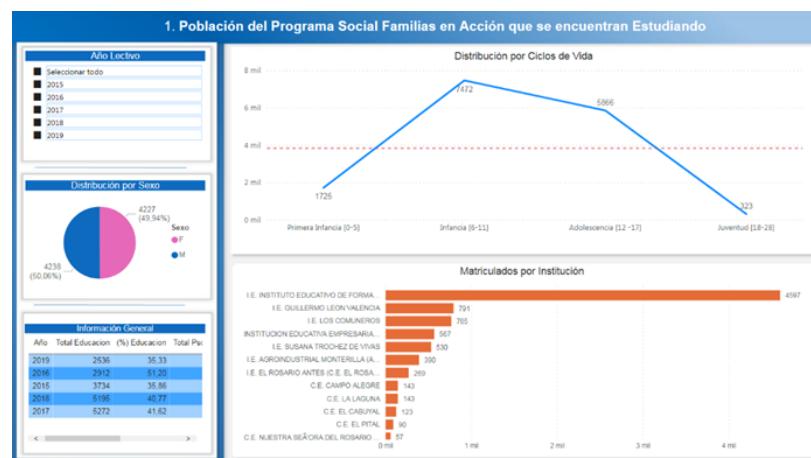
Fig. 6. Social Programs and Education Dashboard.

María-Alejandra Varona-Taborda; Jorge-Cesar Mosquera-Ramírez; César-Augusto Medina-Moreno; Diego-Fernando Lemus-Muñoz; Carlos-Julián Muñoz-Hernández; Christian-Gustavo Arias-Iragorri

Figure 7 shows the Dashboard, which combines the information on the student population with the information from the Individual Service Provision Registry – RIPS, to know the Promotion and Prevention activities carried out on this population in a territorial entity, the user can identify the activities applied or executed by life cycle, generating expectations of fulfillment and/or increase of activities not typical of the student population between the ages of 0 to 17 years in the framework of the Mandatory Health Plan - POS.
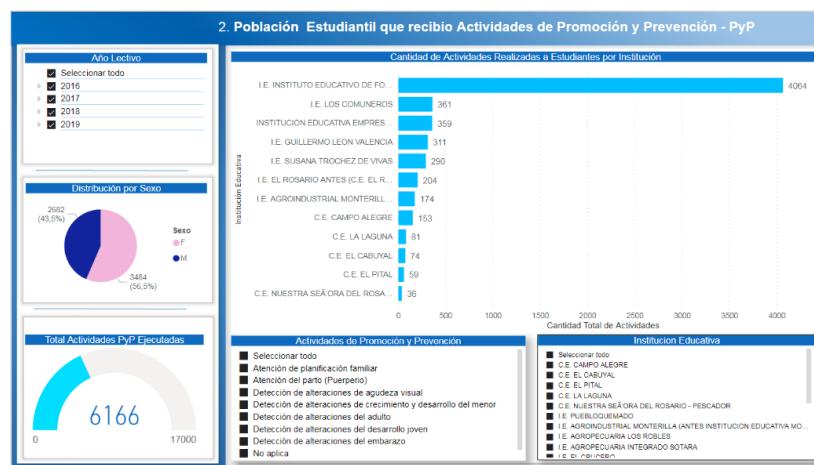

Fig. 7. Dashboard of the student population that received promotion and prevention activities.

Figure 8 shows the Dashboard, which combines the information on the student population with the information from the Individual Service Provision Registry - RIPS, identifying the adolescent pregnancies presented in the student population. For this case, a total of 553 adolescent pregnancies between 2016 and 2019 can be evidenced.

Fig. 8. Dashboard of pregnancies in student population.

Figure 9 shows the Dashboard that combines the information on the population in social programs with the information from the Individual Registry for the Provision of Services -RIPS, to know the children and adolescents who present a state of malnutrition, in this case a total of 44 minors affected by malnutrition can be seen between 2016 and 2019, and the highest percentage is determined in the early childhood stage.
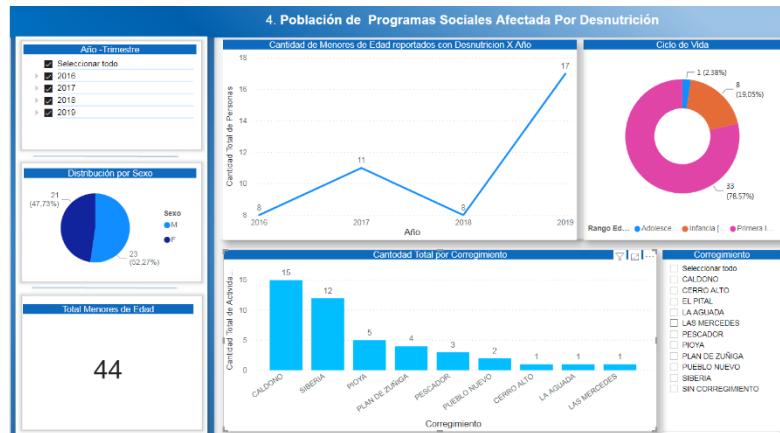
Fig. 9. Dashboard of population in social programs affected by malnutrition.

IV. DISCUSSION AND CONCLUSIONS

It is evident that public institutions are rich in data and information, but currently they do not have enough tools to exploit it and perform a better analysis. Thus, it is

María-Alejandra Varona-Taborda; Jorge-Cesar Mosquera-Ramírez; César-Augusto Medina-Moreno; Diego-Fernando Lemus-Muñoz; Carlos-Julián Muñoz-Hernández; Christian-Gustavo Arias-Iragorri

of great importance that government entities begin to venture into the implementation of new technologies such as Business Intelligence and Data Warehouse solutions that allow easy access and analysis of information, facilitating decision-making and the application of new strategies to solve specific problems through the use of control panels, visualization of indicators, and generation of reports.

By having an information system integrated into a territorial entity, the probability of inconveniences in the standardization and quality of the data decreases, which presents an opportunity for the adoption of more specific ETL processes that allow generating quality information.

It is essential to raise awareness among territorial entities about the importance of implementing a project of this type in order to have access to more sources of data and historical information and demonstrate the usefulness and importance of this type of technology in the official sector, to scale the solution to larger territorial entities, Departments (Governments).Also explore other technologies and tools to design an ecosystem in the cloud that allows the territorial entity to improve access and consultation of data.

With this solution, the opportunity to develop an own methodology for the documentation and implementation of projects of this type, based on the adoption of good industry practices arises. In the same way enriching the current scope of Data Warehouse with data mining techniques to obtain more information for interpretation and analysis.

AUTHORS' CONTRIBUTION

Maria-Alejandra Varona-Taborda: Supervision, Formal Analysis, Investigation, Writing –Review and editing.

Jorge-Cesar Mosquera-Ramírez: Conceptualization, Methodology, Writing – Review and editing.

César-Augusto Medina-Moreno: Conceptualization, Methodology, Writing – Review and editing.

Diego-Fernando Lemus-Muñoz: Conceptualization, Methodology, Writing – Review and editing.

Carlos-Julián Muñoz-Hernández: Conceptualization, Methodology, Writing – Review and editing.

Christian-Gustavo Arias-Iragorri: Supervision, Formal Analysis, Investigation, Writing –Review and editing.

REFERENCES

[1] DNP Departamento Nacional de Planeación, *Implementación de Big Data en Entidades Públicas*, 2017. : https://www.dnp.gov.co/Paginas/El-96,7-de-las-entidades-p%C3%BAblicas-deben-adelantar-acciones-para-prepararse-en-la-implementaci%C3%B3n-de-Big-Data-.aspx

[2] W. Widianty, "Data Warehouse Design with Kimball Method: Case Study of Fahrenheit Manufacturing Systems," *ComTech: Computer, Mathematics and Engineering Applications*, vol. VI, no. 4, pp. 604-612, 2015. https://doi.org/10.21512/comtech.v6i4.2200

[3] C. H. Bolaños Martinez, I. F. A. Damián, H. F. Muñoz Muñoz, A. F. Negrete Gómez, M. A. Tunubalá Morales, C. G. Arias Iragorri, M. A. Varona Taborda, "Inteligencia de Negocios para el Análisis de la Accidentalidad Vial en la Ciudad de Popayán," *Revista Ibérica de Sistemas e Tecnologias de Informação*, no. E38, pp. 130-141, 2020.

[4] S. L. Morales Cardoso, *Metodología para procesos de Inteligencia de Negocios con mejoras en la extracción y transformación de fuentes de datos, orientado a la toma de decisiones*, Alicante, 2019

[5] J. M. Pacheco Casadiego, *Metodología para elaborar el modelo conceptual de datos*, Ediciones Universidad Cooperativa de Colombia, Bogotá, 2017

[6] Y. Quispe Valero, "Desarrollo de una aplicación para la toma de decisiones en el proceso de adquisición utilizando Business Intelligence con la metodología Ralph Kimball en la Municipalidad Provincial de Lampa," Grade Thesis, Universidad Peruana Unión, Lima, Perú, 2018

[7] A. B. Farroñan Carranza, "Implementación de inteligencia de negocios con uso de la herramienta extracción, transformación y carga en las organizaciones para la toma de decisiones: una revisión sistemática," Universidad Católica Santo Toribio de Mogrovejo, Chiclayo, Perú, 2020