EasyChair Preprint

№ 6315

Tool Alternatives to Facilitate Data Quality Monitoring

Monica Rosa Lopez Guayasamin and
Nel̀ stor Darì l o Duque-Méndez

August 17, 2021

Alternativas de herramientas para facilitar el monitoreo de la Calidad de Datos

Monica Rosa Lopez Guayasamin[1][0000-1111-2222-3333] and Néstor Darío Duque-Mendez[2][1111-2222-3333-4444]

[1] Universidad Nacional de Colombia, Sede Manizales, Colombia
[2] Universidad Nacional de Colombia, Sede Manizales Colombia

Abstract. Companies are managing large volumes of data, which must be optimally stored in adequate infrastructure and scalable. However, the big problem for using these data is the visualization for controlling the quality of the same; since quality is such a relevant issue, there are still issues to be faced following the importance it deserves.

The work presented in this article provides a tour of tools that allow a business to monitor and control the quality of their data; different tools that facilitate this process was used and evaluated. Due to the importance of and supported in the exploratory review, Completeness and Validity are defined as the dimensions to be incorporated in the evaluation exercise. First, a data quality scan is started with free tools; then the exercise is continued with a very powerful and easy-to-use analytical tool such as SPSS Modeler, and finally, work is done on a DQS Data Quality Server licensed tool that facilitates the exercise of defining rules and their applicability in the business.

Keywords: Data Quality, Dimensions, Visualization, Tools.

Resumen. Las empresas están manejando grandes volúmenes de datos, los cuales deben ser almacenados de manera óptima en una infraestructura adecuada y que sea escalable. Sin embargo, el gran problema para el uso de estos datos es la visualización para el control de la calidad de los mismos; pues siendo la calidad un tema tan relevante, aún hay asuntos por enfrentar acordes con la importancia que merece.
El trabajo presentado en este artículo ofrece un recorrido sobre herramientas que permiten a los negocios monitorear y controlar la calidad de sus datos; se hicieron usaron y evaluaron diferentes herramientas que facilitan este proceso. Por la importancia que revisten y soportados en la revisión exploratoria, se definen Completitud y Validez como las dimensiones a incorporar en ejercicio de evaluación. Se inicia una exploración de calidad de datos con herramientas libres; luego se continua el ejercicio con una herramienta de analítica, muy poderosa y de fácil uso como SPSS Modeler y por último se trabaja sobre una herramienta licenciada

DQS Data Quality Server que facilita el ejercicio de definición de reglas y su aplicabilidad en la empresa.

Palabras Clave: Calidad de datos, Dimensiones, Visualización, Herramientas.

1 Introducción

El problema de calidad de datos se ha trabajado en las empresas, pero en ocasiones no ha tenido la suficiente importancia hasta que la consolidación de los datos y el uso de estos, no han dado las señales esperadas por los negocios. Uno de los puntos importantes y que ocupa mucho tiempo en un estudio analítico es la revisión de la información, pues es ahí donde el analista de datos se enfrenta a dos retos importantes en este proceso: uno eliminar los datos con problemas, y el otro es limpiar la data, proceso en el cual se pueden perder señales importantes de estos datos.

A la fecha, los estudios de calidad de datos a nivel mundial se ha centrado en temas puntuales como Calidad en Datos Abiertos [1], [2], [3], Evaluación de Calidad [4], Riesgos en Calidad de Datos [5], Calidad en Big Data [6], Herramientas de Calidad [7], [8], Problemas Calidad de Datos [9], Calidad Datos Gobierno [10], [11], [12], Métricas de Calidad [13], [11], [14], [15] entre otros.

En la figura 1 se presenta un consolidado de las temáticas abordadas en los artículos que hicieron parte de este análisis asociado al tema de calidad de datos.

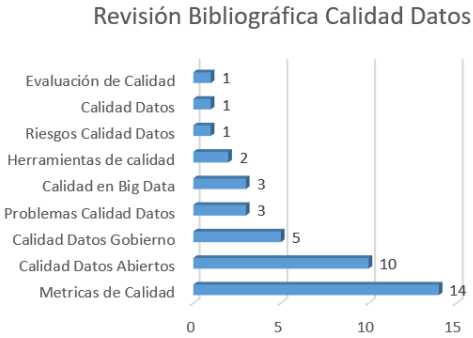

Fig. 1 Revisión Bibliográfica Calidad de Datos

En las empresas, los procesos deben garantizar calidad y confiabilidad en la información de la cual son responsables; sin embargo, uno de los puntos mas complicados es como responder ante esta tarea de manera oportuna y con las herramientas adecuadas. Por lo tanto, el reto es contar con una herramienta que permita de manera automática hacer el seguimiento y control de la calidad de los datos y que dicho control pueda ser monitoreado en el tiempo.

Este documento, comparte la experiencia que se ha tenido con diferentes herramientas tecnológicas para facilitar el monitoreo de la calidad de datos sobre un grupo de atributos los cuales son definidos de manera importante para el negocio. Se contribuye a la empresa definiendo una forma de como implementar control y seguimiento a la calidad mediante un ejercicio articulado y ordenado que inicia desde la parametrización de los datos, el análisis de la información en las dimensiones de calidad de Completitud y Validez, la consolidación y almacenamiento de los procesos de calidad y por último la visualización del comportamiento de esta data.

2 Metodología

Según [16] Cross Industry Standard Process for Data Mining - Crisp-DM tiene gran acogida en la industria, siendo utilizado por más de 160 empresas e instituciones de todo el mundo, que surge en respuesta a la falta de estandarización, siendo planteada como una metodología imparcial o neutra respecto a la herramienta que se utilice para el desarrollo de almacén de datos y Data Mining. El presente trabajo utilizó esta metodología y aunque esta es una metodología que se utiliza para identificar y hacer modelamientos analíticos, se utilizan sus bases para apoyar la consolidación de los ejercicios de calidad de datos que se realizan para este trabajo. Para este proceso fue necesario realizar las fases definidas en la metodología para tal fin:

- Comprensión del Negocio. Etapa en la que es necesario el acompañamiento de los usuarios funcionales pues se identifican las necesidades del proceso en cuanto a calidad, permitiendo definir los objetos de datos y los atributos prioritarios donde se deben aplicar las métricas de calidad.
- Comprensión de los datos. En esta etapa se procede con la identificación de las fuentes de datos asociados a los objetos identificados. Se realizó la construcción de las consultas necesarias sobre las diferentes plataformas a trabajar.
- Preparación de los datos. Etapa en la cual se realizan las configuraciones de las diferentes plataformas para el análisis de los datos que son particulares para cada ambiente de trabajo.
- Modelamiento. En este ejercicio se aplicaron diferentes herramientas:
    - Data Cleaner. Software de licenciamiento corto de forma libre en el que la configuración define las conexiones a las bases de datos fuente y la inclusión de las variables a revisar en cada modelamiento. En la Figura 2 se visualiza el uso de los objetos Patter Finder (Identificación de patrones), String Analizer (Identificación de cadenas), Number Analizer (Identificación de patrones en números), Character set distribution (Patrones en cadenas) que hacen parte de la dimensión Validez

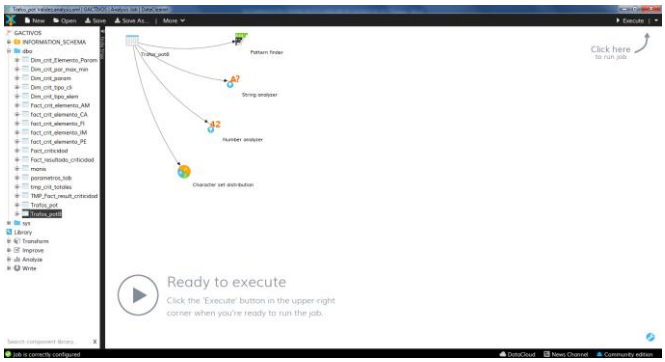

Fig. 2 Un ejemplo de la configuración de la herramienta para un objeto en Data Cleaner.

- Spss Modeler. Software licenciado de análisis de texto y minería de datos de IBM, la configuración y modelamiento de los ejercicios de validez se realizan en la plataforma utilizando los componentes básicos para la transformación de los datos. La Figura 3 permite visualizar un ejemplo del uso de componentes como Derivar (transforma valores), Agregar (implementa funciones), Filtro (excluye datos), Fundir (consolida datos) para poder modelar la data original y transformarla de acuerdo con las reglas definidas en validez.

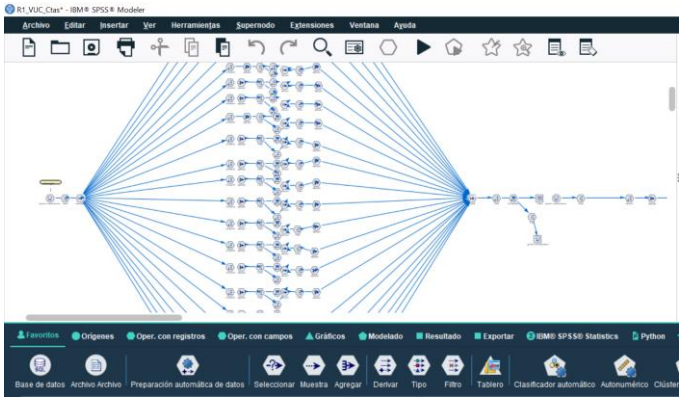

Fig. 3 Modelamiento de datos en la Plataforma Spss Modeler

- DQS Data Quality Services permite que un administrador de datos mantenga la calidad de los datos mediante la administración y configuración de variables. La configuración se hace por medio de la definición de dominios, como se aprecia en la Figura 4, en donde se parametrizan reglas por cada uno, lo cual facilita el modelamiento de los datos para que al ejecutar el procesamiento se obtengan las respuestas acordes a las expectativas del usuario.

Fig. 4 Modelamiento de datos en la Plataforma Data Quality Server

- Evaluación. En esta etapa se consolidan los resultados de las diferentes plataformas y se comparte al usuario funcional dicho resultado, teniendo claro que el criterio de validación de los resultados es una tarea asignada a este usuario.
- Despliegue. Esta etapa abre muchas opciones que dependen de los proyectos trabajados y del interés en la visualización. Para el ejercicio donde se hace uso de la plataforma libre, se consolidaron los resultados XML en una hoja de cálculo tipo Excel y se procedieron a trabajar graficas dinámicas para dicha visualización de resultados. Para los otros dos ejercicios se consolida la información en unas tablas resultado desde las cuales se procede a realizar un despliegue de datos mediante una plataforma de visualización de datos en un Dashboard.

3 Experimentación

En esta sección se exponen los 3 escenarios de pruebas con el fin de dar claridad sobre cada uno de los pilotos realizados.

3.1 Análisis Distribución (Data Cleaner).

Para este ejercicio se define que los objetos a analizar son activos de distribución prioritarios en la empresa por su nivel de criticidad y salud de activos. En este ejercicio se identifican los atributos relevantes por cada activo y un porcentaje de peso para cada atributo hasta llegar a un 100% del peso total. Una vez realizado el paso anterior, se procede al procesamiento de la calidad de los datos con la herramienta Data Cleaner; los resultados de los ejercicios se exportan a XML para luego ser consolidados en Excel y posteriormente se procede con la visualización para el usuario final en Excel.

A continuación, en la Figura 5 se puede visualizar un ejercicio de ponderación de objetos y variables principales asociadas en cada objeto.

DESCRIPCION CLASIFICACION (ACTIVO)		ATRIBUTO DEL ACTIVO			
DESCRIPCIÓN	PONDERACIÓN (%)	NOMBRE DEL ATRIBUTO	PONDERACIÓN (%)	RESULTADO CALIDAD DE DATOS DEL ATRIBUTO	RESULTADO CALIDAD DE DATOS DEL ACTIVO
CONDUCTOR PRIMARIO	8,33%	TENSION_CONN	10%	1	0,083
		NUMERO_FASES	10%		
		MATERIAL	10%		
		AISLAMIENTO	10%		
		TIPO_AISLAMIENTO	10%		
		CIRCUITO_CONN	10%		
		LOCALIZACION_CONN	10%		
		USO	10%		
		CALIBRE	10%		
		LONGITUD	10%		
RECONECTADOR, 13,2 y 33	8,33%	NUMERO_FASES	12,5%	1	0,083
		CORR_NOM	12,5%		
		TIPO_GES	12,5%		
		TIPO_AISLAMIENTO	12,5%		
		CORR_INT	12,5%		
		BIDIRECCIONAL	12,5%		
		I_NOMCOR	12,5%		
		FASES_CONN	12,5%		

Fig.5 Parametrización de porcentajes de variables en Excel

En la figura 6 se visualiza como desde la plataforma Data Cleaner se realiza el montaje del modelamiento del comportamiento de validez de los datos de la entidad customer.

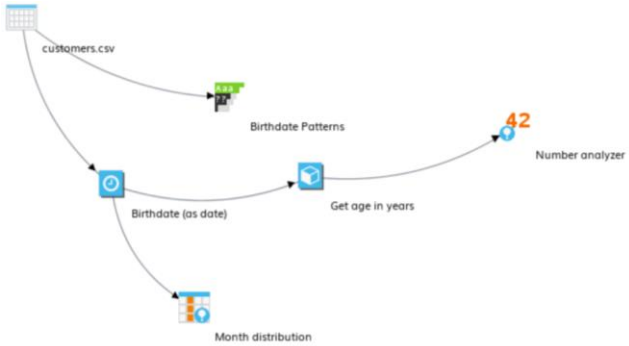

Fig. 6 Configuración de flujo de validez variable Customer

Una vez ejecutado el anterior proceso se obtiene el resultado de las variables programadas para analizar su comportamiento y luego permitir descargar su resultado a un archivo plano. Un ejemplo de este ejercicio se visualiza en la Figura 7.

Names completeness

(Completeness analyzer) (given_name,family_name,company)

Incomplete records (12)

Save dataset

id	given_name	family_name	company	address_line	post_code	city	country	email	job_title	birthdate	income_amount
5014	James	Langord		14, Cook Stre...	OL162LZ	Rochdale	GBR	James.La...	IT Support	1970-1-10	96349.0
5057		Wiltshire	Archer Da...	335, Gander ...	SM3 9QX	Sutton	GBR		Data Arc...	1981-11-9	30107.0
5062		Seekjær	Ingram M...	Sanatorievej ...	7140	Stouby		R..Seekjær...	Developer	1950-1-16	0.0
5087	Karen	Molloy		124, Park Av...	EN6 5EL	Potters Bar	GBR	Karen.Mol...	Sr	1963-7-2	36970.0
5097	Sven	Owen					DEU	Sven.Owe...	CEO	2008-1-26	105524.0
5104	Jody	Cook		ROOM 216,2...	MD 21201-...	BALTIMO...	USA	Jody.Cook...	dba	1981-2-21	189033.0
5080			General El...	Fairview, Holt...	B98 9AT	Redditch	GBR		CEO	1997-11-26	85237.0
5083	Alistair Whit...		Ingram M...	39, Merafield ...	PL7 1TL	Plymouth	GBR	Alistair.W...	DBA	1995-6-18	86548.0
5039	Boyd A	Sabouni		ROUTE 206 S...	NJ 7921				manager	1928.9.16	0.0
5025		Hayden	Prudential...	55 FARMING...	CT 6105	HARTFORD	USA	HEATHER...	System A...	1905.2.4	158159.0
5007	Gerhard	Jacobi		Friedhofstr. 8	64372	Ober-Ram...	DE	Gerhard.J...	student	1955-5-18	0.0
5100	Reinhard	Frings		Auf der Kapel...	91781	Weissenb...		Reinhard....	mr.	1980-8-20	0.0

Fig. 7 Ejemplo de comportamiento de los datos en Completitud

3.2 Análisis Comercial - Spss Modeler

Para el uso de la herramienta se optó para el ejercicio la exploración de calidad de datos en el área comercial, donde existe un elemento que consolida todas las variables del negocio de las cuales se requiere hacer. En la Figura 8 se muestra la distribución de porcentajes para valorar la calidad de acuerdo a lo referenciado por el usuario final.

DESCRIPCION DE LA CLASIFICACION (ACTIVO)		ATRIBUTO DEL ACTIVO			
DESCRIPCION	PONDERACION (%)	NOMBRE DEL ATRIBUTO	PONDERACION (%)	RESULTADO CALIDAD DE DATOS DEL ATRIBUTO	RESULTADO CALIDAD DE DATOS DEL ACTIVO
Caracterizacion de Cuentas	100%	AÑO	4%	25,5%	25,5%
		cant_pqrs	4%		
		cantidad_visitas_giip	4%		
		cantidad_visitas_giip_con_irr	4%		
		cart_30	4%		
		cart_60_90_120	4%		
		consumo_subsistencia_kwh	4%		
		consumos_kvarh	4%		
		consumos_kwh	4%		
		dias_pago	4%		
		dias_suspendido	4%		
		estaba_suspendido	4%		
		facturacion_kvarh	4%		
		facturacion_kwh	4%		
		facturacion_sitio	4%		
		MES	4%		
		nis	4%		
		realizo_pago	4%		
		tiene_consumo_kwh_no_cero	4%		
		valor_compensado	4%		
		valor_contribucion	4%		
		valor_pago	4%		
		valor_subsidio	4%		
		valor_total_facturado	4%		

Fig. 8 Distribución de porcentajes de valoración

Posteriormente, mediante un modelo de analítico sobre la plataforma Spss Modeler se realiza una exploración de las cuentas, denominado caracterización de cuentas. En este modelamiento, lo que se busca es transformar la data original para posteriormente adecuarla y analizar algunos patrones de comportamiento dado la definición de validez de algunas variables. En la Figura 9 se visualiza como el usuario functional define los criterios de evolución de algunas variables como cantidad de PQRS y cantidad de visitas en GIIP.

cant_pqrs			cantidad_visitas_giip		
BUENO	MALO	ATIPICO	BUENO	MALO	ATIPICO
Mayor a 0 y menor que 4	Menor a 0	Mayor a 4	Mayor = 0 y Menor = 2	Menor 0	Mayor a 3

Fig. 9 Configuración de criterios de validez

Con base en la anterior definición de variables, se procede a implementar en la Plataforma SPSS Modeler el modelo requerido para el análisis de variables como se presenta en la Figura 10 donde se define, mediante super nodos, la configuración de las diferentes variables a analizar en el ejercicio de calidad de datos.

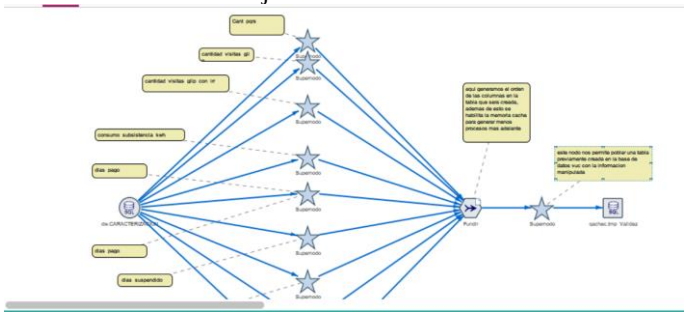

Fig. 10 Modelamiento de variables en SPSS Modeler

3.3 Análisis Comercial - DQS Data Quality Server

Para la implementación del proyecto de calidad con esta plataforma se respetaron las mismas condiciones del ejercicio anterior, lo que significa que se toman como referencia la misma estructura y las mismas variables. Al cambiar la plataforma de modelamiento se requiere cambiar la configuración de las variables, por lo tanto, para Validez se pretende distinguir los datos buenos, malos y atípicos de un conjunto ingresado a partir de bases de datos SQL y para Completitud se pretende distinguir cuáles son los datos que están completos y cuáles incompletos de un conjunto ingresados, a partir también de las bases de datos de SQL. Un ejemplo de este ejercicio se presenta en las Figura 11 y Figura 12 donde se configura una de las variables dentro de la herramienta.

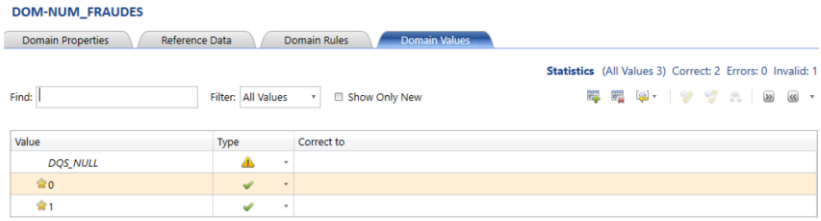

Fig. 11 Configuración de variables en DQS

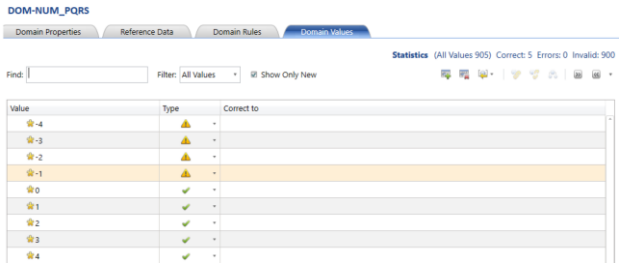

Fig. 12 Configuración de variable PQR en DQS

Una vez configuradas las variables, se procede a configurar un proceso de ETL haciendo un llamado a la ejecución de la base de conocimiento configurada en DQS mediante la cual se ejecutan los dominios anteriormente configurados y cuyo resultado se lleva a una tabla temporal donde finalmente reposa la información procesada. Lo anteriormente descrito se observa en la Figura 13.

Fig. 13 Proceso de ETL para llevar resultados a tablas

Posteriormente se realiza la consolidación en las estructuras finales donde se lleva la información procesada y actualizada de la validación de calidad de datos.

4 Resultados

En esta sección se presenta la forma como los diferentes ejercicios de calidad de datos se han manejado, para el caso de estudio.

4.1 Análisis distribución - Data Cleaner

Después de realizar la consolidación de datos, se procedió a integrar los resultados de la ejecución de la calidad de datos. Inicialmente este procesamiento se realizó como una fase exploratoria, se usó la herramienta Excel para consolidar datos mediante tablas y gráficos dinámicos. En las Figura 14 y Figura 15 se puede visualizar el resultado de las dimensiones Completitud y Validez aplicado a los datos de distribución.

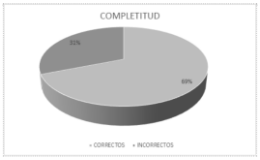

Fig. 14 Visualización Completitud datos

Fig. 15 Visualización Validez datos

4.2 Datos *comercial* – SPSS Modeler

Teniendo en cuenta que el resultado de la ejecución periódica del modelo implementado con esta plataforma se procesa y se inserta en un objeto de base de datos con una variable de temporalidad cada mes; la visualización se realiza con la plataforma Power BI donde la fuente de información es la consolidación de data en el tiempo realizado con la ejecución del modelo implementado. Un ejemplo de este resultado se registra en las Figuras 16 y 17 donde se muestra el reporte inicial consolidado de datos, y el reporte de la dimensión Validez de algunas variables.

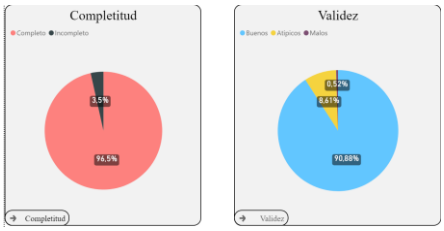

Fig. 16 Visualización datos comercial – origen SPSS Modeler

Fig. 17 Visualización datos comercial dimensión Validez – origen SPSS Modeler

4.3 Datos *comercial* – DQS

Para la implementación de este ejercicio, se mantuvo la filosofía utilizada en el trabajo de calidad realizado con la Plataforma SPSS Modeler. Las variables estaban ya definidas y los criterios o reglas a implementar también. El punto adicional en este ejercicio fue consolidar el historial de ejecuciones en el proceso de calidad de datos a los objetos de comercial. En la Figura 18 se aprecia la forma de filtrar y visualizar el resultado de la calidad de datos en el tiempo por atributo, fecha u objeto, y en la Figura 19 se visualiza un ejemplo de medición de variables de comercial

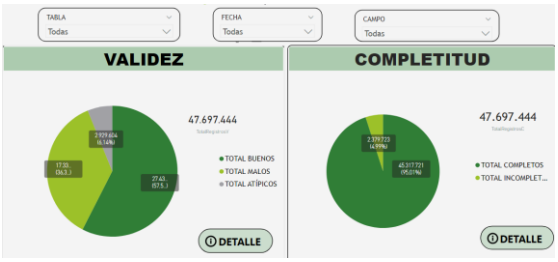

Fig. 18 Consolidado datos *comercial* - origen DQS

Fig. 19 Visualización dimensión Validez - origen DQS

5 Conclusiones

Este trabajo recopila un ejercicio de aplicación de diferentes herramientas para identificar la calidad de datos. Es importante resaltar que estos ejercicios se han realizado en diferentes momentos del tiempo desde el 2018 a la fecha y han requerido un esfuerzo en aprendizaje de las herramientas y plataformas tecnológicas aquí mencionadas. En

las diferentes etapas se ha ido mejorando en la visualización de la calidad de datos de un objeto en un periodo de tiempo, pero es importante destacar que dicha mejora se debe a que las herramientas actuales han facilitado este proceso.

Como aprendizaje se reconoce que las plataformas libres si bien facilitan las actividades, no siempre vienen con los componentes requeridos para trabajar de manera completa en un ejercicio y normalmente toca ajustar con nuevas tareas los resultados entregados por dichas plataformas. Este ejercicio exige algunas configuraciones y un poco de tiempo adicional para la entrega de resultados.

Por otro lado, herramientas como SPSS Modeler son una ventaja para cualquier empresa, ya que aparte de facilitar la generación de modelos analíticos permiten cerrar el ciclo completo de consolidación de calidad de datos. Pero este tipo de herramientas tienen un alto costo en licenciamiento, por lo cual no es viable su implementación en cualquier empresa.

Finalmente, con la herramienta DQS se tiene una ventaja importante frente a las demás: la configuración de variables es parametrizada de forma aislada en una base de conocimiento, la cual es procesada posteriormente con un ejercicio de ETL que se adecua perfectamente a la necesidad de calidad del usuario y permite ser aplicada en muchos dominios, si se requiere. Esta independencia en la parametrización garantiza que el ajuste de variables para su ejecución facilita a los procesos de Tecnología ser más oportunos en dicha tarea.

Con este trabajo se puede continuar explorando en otras dimensiones de calidad ya que la herramienta DQS facilita la implementación de nuevos dominios de acuerdo con unas definiciones y estándares identificados por el usuario final.

6 Reconocimientos

A la Universidad Nacional de Colombia Sede Manizales, Institución que ha apoyado este proceso de investigación y a la Central Hidroeléctrica de Caldas CHEC S,A. E,S,P empresa que ha facilitado los datos, el conocimiento y los recursos para la realización de este trabajo.

Referencias

[1] S. Sadiq and M. Indulska, “Open data: Quality over quantity,” *Int. J. Inf. Manage.*, vol. 37, no. 3, pp. 150–154, 2017, doi: 10.1016/j.ijinfomgt.2017.01.003.

[2] V. Estrada-Galinanes and K. Wac, “Visions and Challenges in Managing and Preserving Data to Measure Quality of Life,” *2018 IEEE 3rd Int. Work. Found. Appl. Self* Syst.*, pp. 92–99, 2019, doi: 10.1109/fas-w.2018.00031.

[3] I. Mergel, A. Kleibrink, and J. Sörvik, “Open data outcomes: U.S. cities between product and process innovation,” *Gov. Inf. Q.*, vol. 35, no. 4, pp. 622–

632, 2018, doi: 10.1016/j.giq.2018.09.004.

[4] H. H. Ahmed, “Data quality assessment in the integration process of linked open data (LOD),” in *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, 2018, doi: 10.1109/AICCSA.2017.178.

[5] A. Colborne and M. Smit, “Identifying and mitigating risks to the quality of open data in the post-truth era,” *Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017*, vol. 2018-Janua, pp. 2588–2594, 2018, doi: 10.1109/BigData.2017.8258218.

[6] P. Zhang, F. Xiong, J. Gao, and J. Wang, *Data Quality in Big Data Processing: Issues, Solutions and Open Problems*. .

[7] H. Wahl, “LEIWI – A Tool to Compute the Quality of Life Using Open Data,” no. Iscit, pp. 116–120, 2018.

[8] W. Xia, Z. Xu, and C. Mao, “User-driven filtering and ranking of topical datasets based on overall data quality,” *Proc. - 2017 14th Web Inf. Syst. Appl. Conf. WISA 2017*, vol. 2018-Janua, no. 1, pp. 257–262, 2018, doi: 10.1109/WISA.2017.24.

[9] A. Nikiforova, “Open Data Quality Evaluation: A Comparative Analysis of Open Data in Latvia,” *Balt. J. Mod. Comput.*, vol. 6, no. 4, pp. 363–386, 2018, doi: 10.22364/bjmc.2018.6.4.04.

[10] R. Machova and M. Lnenicka, “Evaluating the Quality of Open Data Portals on the National Level,” *J. Theor. Appl. Electron. Commer. Res.*, vol. 12, no. 1, pp. 21–41, 2017, doi: 10.4067/S0718-18762017000100003.

[11] A. Vetrò, L. Canova, M. Torchiano, C. O. Minotas, R. Iemma, and F. Morando, “Open data quality measurement framework: Definition and application to Open Government Data,” *Gov. Inf. Q.*, vol. 33, no. 2, pp. 325–337, 2016, [Online]. Available: http://dx.doi.org/10.1016/j.giq.2016.02.001.

[12] A. Whitmore, “Using open government data to predict war: A case study of data and systems challenges,” *Gov. Inf. Q.*, vol. 31, no. 4, pp. 622–630, 2014, doi: 10.1016/j.giq.2014.04.003.

[13] A. Abella and M. O. C. De-pablos-heredero, “INDICADORES DE CALIDAD DE DATOS ABIERTOS : EL CASO DEL PORTAL DE DATOS ABIERTOS DE BARCELONA Open data quality metrics : Barcelona open data portal case.”

[14] C. Batini and M. Scannapieca, *Data-Centric Systems and Applications: Data Quality Concepts, Methodologies and Techniques*. 2006.

[15] A. Abella, M. Ortiz-De-urbina-criado, and C. De-Pablos-heredero, “Meloda 5: A metric to assess open data reusability,” *Prof. la Inf.*, vol. 28, no. 6, pp. 8–10, 2019, doi: 10.3145/epi.2019.nov.20.

[16] P. Orv, F. Vrq, X. Dgrv, and S. Od, “Metodología crisp para la implementación Data Warehouse,” *Tecnura*, vol. 14, no. 26, pp. 35–48, 2010, doi: 10.14483/22487638.6685.