

CALIDAD DE DATOS EN SISTEMAS DE GESTIÓN ACADÉMICA UNIVERSITARIA BASADOS EN ISO/IEC 25012

MARLENY PERALTA ASCUE

marlenyperaltaascue@gmail.com / ORCID: 0000-0002-0708-8565
Escuela Universitaria de Posgrado Universidad Nacional Federico Villarreal,
Lima, Perú
Universidad Nacional Micaela Bastidas de Apurímac, Perú

Resumen

El presente artículo tiene como propósito proponer un modelo de evaluación de calidad de datos desarrollada sobre la base principal del estándar ISO/IEC 25012, aplicada a un sistema de gestión académica universitaria, con el objetivo de mejorar la calidad de los datos. El modelo propuesto se desarrolla desde la perspectiva del consumidor de datos y desde la visión de la calidad de datos inherentes. La muestra estuvo conformada por los datos almacenados en el sistema de gestión académica de la Universidad Nacional Micaela Bastidas de Apurímac, con un total de 22 tablas, 154 atributos y 319 685 registros. El modelo comienza con el análisis de los requisitos de calidad de datos, como insumo principal para su evaluación, y finaliza con un plan de mejora, el mismo que es implementado de manera automática utilizando herramientas de limpieza de datos y código SQL. Las características que inciden en problemas de calidad de los datos son: la precisión, consistencia, conformidad y actualidad. Finalmente se concluye que es posible mejorar la calidad de los datos aplicando el modelo propuesto, el mismo que puede ser usado para crear y generar valor a través de la exploración, explotación y análisis de datos en beneficio de la gestión académica universitaria.

PALABRAS CLAVES: modelo de evaluación / calidad de datos / sistema de gestión académica / ISO/IEC 25012

QUALITY OF DATA IN UNIVERSITY ACADEMIC MANAGEMENT SYSTEMS BASED ON ISO / IEC 25012

Abstract

This article proposes a data quality evaluation model developed on the primary basis of the ISO/IEC 25012 standard, applied to a University Academic Management System, to improve data quality. The proposed model is developed from the perspective of the data consumer and the vision of inherent data quality. The sample consisted of the data stored in the Academic Management System of the Universidad Nacional Micaela Bastidas, Apurímac, Perú, with 22 tables, 154 attributes, and 319 685 records. The model begins with data quality requirements as the primary input for its evaluation and ends with an improvement plan, automatically implemented using data cleaning tools and SQL code. The characteristics that affect data quality problems are accuracy, consistency, compliance, and timeliness. Finally, I conclude that it is possible to improve the quality of data by applying the proposed model, which can be used to create and generate value through the exploration, exploitation, and analysis of data for the benefit of university academic management.

KEYWORDS: evaluation model / quality of data / Academic Management System
/ ISO/IEC 25012

1. INTRODUCCIÓN

Es muy común que las organizaciones desconozcan el nivel de calidad de sus datos almacenados y el valor que estos pueden tener para generar grandes beneficios económicos, sociales y científicos. Los avances tecnológicos de los últimos años han hecho que la cantidad de los datos generados en las organizaciones aumenten exponencialmente en cuanto a su almacenamiento y volumen. Es cada vez más frecuente escuchar expresiones como “los datos son la nueva moneda”, “el nuevo petróleo del siglo XXI”, “los datos son la nueva mina”. En consecuencia, se puede decir que “nuestro mundo se está datificando” (Caballero *et al.*, 2017).

La calidad de datos se ha convertido en un elemento clave para la calidad de las decisiones y acciones en una organización. Sin embargo, descubrir la calidad de un conjunto de datos es un trabajo fundamental que requiere de un proceso de evaluación en función de ciertos requisitos y características de calidad (Zhang *et al.*, 2019).

Estudios realizados por DalleMule y Davenport (2017) señalan que del 100 % de los datos estructurados en una organización, menos del 50 % son utilizados de manera activa para tomar decisiones; así como que más del 70 % de los trabajadores tienen disponibilidad de datos a los que no deberían acceder y que el 80 % del tiempo de los analistas se dedica solo a capturar y preparar datos.

La compañía multinacional PowerData, de origen español, que proporciona servicios especializados de gestión de datos con presencia en países como Argentina, Chile, Colombia, Ecuador, México, Perú y Uruguay, aplicó una encuesta realizada por Price Waterhouse Coopers, en el año 2018, cuyos resultados fueron que el 75 % de 600 empresas a las que encuestaron reportaron problemas significativos como resultado de datos incorrectos, debido a la mala calidad de los datos.

Esto nos conduce a enfocar nuestro interés en los estándares definidos por ISO/IEC 25 000, conocido como el estándar SQuaRE (*System and Software Quality Requirements and Evaluation*), que es un conjunto de normas internacionales compuesto por cinco divisiones. Una de estas corresponde a la familia 2501n, que se focaliza en modelos de calidad de productos de *software* y, como parte de esta familia, se desprende el estándar ISO/IEC 25012 que se focaliza en la calidad de los datos. (ISO/IEC 25 000, 2021).

El estándar ISO/IEC 25012 propone un modelo de calidad de datos enfocado en sistemas de información en un entorno particular; identifica quince características de calidad que se pueden evaluar desde dos perspectivas: inherente y dependiente del sistema, como se ve en la figura 1 (ISO/IEC 25012,2018). Dichas características poseen métricas que permiten medir y obtener valores para establecer los niveles de calidad mediante fórmulas.

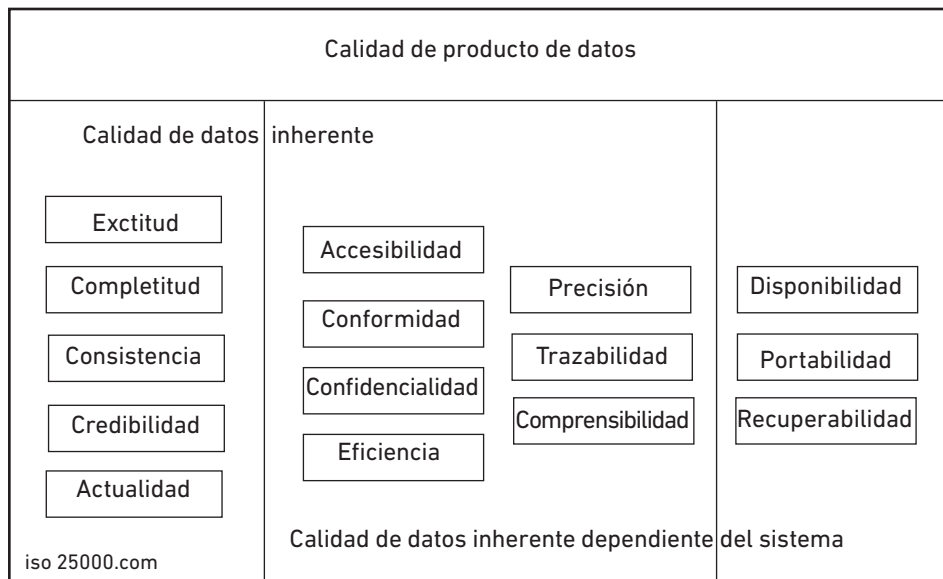


Figura 1. Características de calidad de datos

Fuente: ISO/IEC 25012 (2008)

La presente investigación tuvo como contexto de estudio del sistema de gestión académica (SISGA) de la Universidad Nacional Micaela Bastidas de Apurímac (UNAMBA), el mismo que se encarga de procesar las matrículas vía intranet e internet del semestre académico correspondiente dentro del calendario académico vigente (UNAMBA, 2018).

El panorama actual de la universidad se ve enfrentado a datos inexplorados que pueden tener cierta relevancia y ser un potencial para crear valor. Sin embargo, al explorar y usar los datos del sistema de gestión académica en la generación de información analítica se presentaron los siguientes problemas: información incompleta; valores fuera de un rango referencial, errores de sintaxis, que están directamente relacionados con el formato de los datos. Datos inconsistentes, datos duplicados o incluso representaciones no adecuadas a la realidad, valores desactualizados, restricciones de integridad no definidas, representación ambigua en el modelo conceptual de datos y débil representación de reglas de consistencia e integridad en el esquema del modelo conceptual de datos.

De acuerdo con lo mencionado en el párrafo anterior, se percibe que los problemas identificados pueden dividirse en problemas al nivel de la instancia a la que están directamente relacionados con el contenido o valor de los datos, y problemas en el modelo conceptual de los datos a nivel de esquemas.

Estudios realizados por Calabrese *et al.* (2019) proponen una guía de medición con la finalidad de dar a conocer la condición actual de los datos a través de distintos aspectos definidos en ISO/IEC 25 012. La evaluación se basa en el proceso adoptado por ISO/IEC 25040. La investigación se ejecutó sobre una base de datos con información relacionada con farmacias, la cual fue obtenida en Argentina a través de un sitio web de datos abiertos.

Otras investigaciones como Fritz *et al.* (2017), presentan un proceso de selección de atributos de calidad de datos en sistemas de gestión de aprendizaje bajo la familia de normas ISO/IEC 25 000, para ser evaluados. Tomando como referencias características de calidad determinadas en ISO/IEC 25 010, 25 012 y 25 040, que precisan un modelo general que permite evaluar la calidad de un sistema de información o *software*. La investigación está enmarcada en la integridad de los datos estructurados almacenados en el sistema de gestión de aprendizaje, desde la perspectiva de seguridad de datos. En el estudio realizado por Manrique (2017), desarrolló un catálogo de reglas de negocio referentes a datos, basado en ISO/IEC 25 012, cuyo objetivo fue mejorar la gestión de reglas de negocio, garantizando la implementación de dichas reglas en sistemas de gestión de base de datos relacionales, sosteniendo la reutilización de las reglas en distintos proyectos de datos y organizaciones de trabajo mediante el uso de catálogos.

Finalmente, considerando que los datos son los activos más reconocidos en el mundo digital y calificados como activos digitales valiosos en las organizaciones, y considerando la importancia de la calidad de datos, es necesario medirla siguiendo un proceso de evaluación en función a ciertos requisitos y características de calidad (Zhang *et al.*, 2019). Por esta razón, la presente investigación cobra importancia al proponer un modelo que servirá como una guía de evaluación paso a paso, que permitirá mejorar la calidad de datos para los responsables de sistemas de información y administradores de base de datos quienes tienen la responsabilidad de garantizar el cumplimiento de estándares relacionados con la gestión de datos. Frente a esta situación problemática de datos descrita en el contexto de estudio, se plantea determinar en qué medida el desarrollo de un modelo de evaluación de calidad de datos basado en ISO/IEC 25 012, mejora la calidad de los datos en el sistema de gestión académica universitaria, evaluando la calidad en el valor del dato y en el modelo conceptual de datos.

2. MATERIALES Y MÉTODOS

El tipo de investigación es aplicada y el diseño de la investigación corresponde a un diseño experimental en su categoría de cuasi experimental (Hernández *et al.*, 2014).

La muestra está constituida por los datos almacenados en el SISGA desde el año 2010 hasta el 2017, con un total de veintidós tablas, 154 atributos y 319 685 registros,

datos que fueron proporcionados por el responsable de la Dirección de Servicios Académicos de la UNAMBA para la ejecución de proyectos de investigación. La investigación se realiza respetando la Ley de protección de datos personales, publicada en el año 2011, por el Gobierno del Perú (Ley 29733, 2011).

2.1 Descripción de la base de datos

Para el caso de estudio vamos a utilizar una porción de la base de datos del sistema SISGA. Este sistema fue desarrollado por personal interno de la institución y almacena sus datos en un gestor de base de datos relacional, por lo que se trabaja con una base de datos relacional. El sistema se encarga de la automatización del proceso de matrículas, consta de seis módulos principales que ofrecen las funciones de registro de estudiantes, docentes, planes curriculares, carga académica, gestiona el registro de matrícula, evaluación y localización de instituciones educativas.

Para evaluar los datos se necesitó:

1. Repositorio elegido para la evaluación: acceso a una copia de los datos del SISGA desde el año 2010 hasta el 2017. (En la tabla 1 se muestran los módulos, tablas y campos de la base de datos a evaluar).
2. Requisitos de calidad de los datos que deben cumplir los datos (especificación de requisitos).
3. El modelo conceptual de datos (esquema de la base de datos).

Tabla 1
Datos a evaluar del Sistema SISGA de la UNAMBA

Módulos	Tablas		N.º	N.º	N.º
	Nombre	Descripción	Tablas	Campos	Registros
Docente	Totales		4	25	922
	Facultad	Unidad de formación académica.	1	2	9
	Escuela	Escuela profesional encargada de impartir estudios de formación.	1	3	19
	Docente	Docentes universitarios que tienen como función enseñar.	1	17	886
Departamento académico	Unidades de servicio académico que reúnen a docentes de disciplinas afines.	1	3	8	

(continúa)

(continuación)

Estudiante	Totales		5	38	10 054
	Sede	Centro de la Universidad que se ubican en distritos de provincias de la Región de Apurímac.	1	1	3
	Estudiante	Estudiante universitario que se encuentra matriculado.	1	17	5015
	DetEstudiante	Detalle de información académica del estudiante.	1	16	5015
	Tipo Estudiante	Representa el estado actual del estudiante.	1	2	6
	Modalidad	Modalidades de ingreso a la universidad.	1	2	15
Currículo	Totales		5	26	3446
	Plan Estudios	Es el esquema estructurado que garantiza el proceso de enseñanza - aprendizaje	1	5	33
	DetPlan	Detalle del plan de estudios	1	7	1821
	Estudios	Materia que se imparte en una escuela.	1	10	1552
	Asignatura	Nombre de la categoría en la que debe de corresponder una asignatura.	1	2	38
	Categoría	Representa el modo de estudio de la asignatura.	1	2	2
Carga Académica	Totales		2	14	11 545
	Carga Académica	Distribución de carga académica por escuela y semestre académico.	1	3	249
	DetCarga Académica	Detalle de la distribución de la carga académica por asignaturas, escuela y docente en un semestre.	1	11	11 296
Matrícula	Totales		4	40	291 602
	Semestre	Semestre académico	1	13	21
	Matrícula	Evento que acredita la condición de estudiante universitario.	1	12	54 415
	DetMatrícula	Detalle de matrícula del estudiante en asignaturas en un semestre.	1	13	237 161
	Condición	Condición de matrícula del estudiante.	1	2	5
Localización de IIEE	Totales		2	11	2116
	Ubigeo	Ubicación geográfica de la institución educativa.	1	7	1856
	Colegio	Datos de instituciones educativas	1	4	260
Totales			22	154	319 685

Elaboración propia

3. RESULTADOS

El modelo de evaluación propuesto, llamado en adelante SDQAM (en inglés, *SQuaRE – based Data Quality Assessment Model*) Modelo de Evaluación de Calidad Datos basado en SQuaRE, describe un marco de evaluación de la calidad tomado como base de un subconjunto de características definidas en el modelo de calidad de datos ISO/IEC 25 012 y métricas de medición de ISO/IEC 25 024 planteado por la familia SQuaRE. En el ámbito de modelos de evaluación se ha integrado el modelo de evaluación de productos de *software* ISO/IEC 25 040 y el modelo de medición y evaluación GOCAME (en inglés, *Goal-Oriented Context-Aware Measurement and Evaluation*), considerada como la estrategia de propósito general que abarca la medición y evaluación. Esto se basa en el sentido en que los consumidores de datos necesitan usar los datos, cuyos valores se espera que sean correctos, actualizados y completos; y que la evaluación de los datos vaya más allá de aspectos tecnológicos.

En la figura 2 se muestra el modelo de evaluación propuesto compuesto por cuatro fases y seis actividades, en las que también se definen los artefactos como los productos de entrada y salida, técnicas y herramientas utilizadas para cada una de las actividades.

En la selección de características del modelo ISO/IEC 25 012, se aplicó una encuesta considerando la valoración de los expertos ingenieros responsables de sistemas de gestión académica de dos universidades de Apurímac; se empleó el Método de Jim Brosseau para incluir el criterio del investigador (evaluador) en la selección de características apropiadas (Clarrus Consulting Group, 2010). Finalmente, se decidió establecer el conjunto de las características para medir la calidad de datos: exactitud, completitud, consistencia, credibilidad, actualidad, conformidad, precisión y extensibilidad. En la tabla 2 se muestra la definición de las características inherentes de calidad de datos.

Tabla 2
Características inherentes de calidad de datos según el estándar ISO/IEC 25012

Característica	Definición
Exactitud	Se refiere al grado en el cual el dato tiene atributos que representan correctamente el valor verdadero del atributo esperado de un concepto o evento en un contexto específico de empleo.
Completitud	Se refiere al grado en el cual el dato del sujeto asociado con una entidad tiene valores para todos los atributos esperados e instancias de entidad relacionadas en un contexto específico de uso.
Consistencia	Se refiere al grado en el cual el dato tiene los atributos que son libres de contradicción y son coherentes con otros datos en un contexto específico de uso.
Credibilidad	Se refiere al grado en el cual el dato tiene atributos que son considerados como verdaderos y auténticos en un contexto específico de uso.

(continúa)

(continuación)

Actualidad	Se refiere al grado en el cual el dato tiene los atributos que son correctos en un período de contexto específico de uso.
Conformidad	Se refiere al grado en el cual el dato tiene atributos que se adhieren a normas, convenciones o regulaciones vigentes y reglas similares relacionadas con la calidad de datos en un contexto específico de uso.
Precisión	Se refiere al grado en el que el dato tiene atributos que son exactos o que proporcionan la discriminación en un contexto específico de uso.
Entendibilidad	Se refiere al grado en el cual el dato tiene atributos que le permiten ser leído e interpretado por usuarios, y es expresado en lenguajes apropiados, símbolos y unidades en un contexto específico de uso.

Fuente: ISO/IEC 25012 (2018)

3.1 Descripción de las fases del modelo SDQAM

Fase 1: Requerimientos: enfocada en determinar los requerimientos de evaluación de los datos que se deben cumplir, para ello se requiere de la actividad (A1) establecer los requisitos de evaluación de datos.

- (A1). Establecer requisitos de evaluación:

El objetivo de esta actividad es definir y especificar los requisitos de calidad de datos de acuerdo con un contexto específico que caracteriza la necesidad de información. En las tabla 3 y 4 se visualizan los productos de entrada y salida, así como la ficha de especificación de requisitos.

Tabla 3

Productos de entrada y salida de la actividad A1 "Establecer requisitos de evaluación"

Productos de entrada	Productos de salida	Técnicas y Herramientas
Copia de los datos a evaluar. Requisitos de calidad de datos Modelo de datos.	Ficha de especificación de requisitos de calidad de datos por cada tabla. (véase la tabla 2)	Estudio de documentación. Entrevistas Observación.

Elaboración propia

Tabla 4
 Ficha de especificación de requisitos

Ficha de especificación de requisitos	Código:	
Tabla: Nombre de la tabla a evaluar		
Reglas sintácticas		
Atributos	Descripción del atributo	Dominio por intención
Atributo 1		
Atributo 2		
Reglas semánticas		
Atributo 1		
Atributo 2		
Especificación del esquema conceptual		

Elaboración propia

La Ficha de especificación de requisitos está compuesta de:

1. *Reglas sintácticas*, reglas que están definidas según el dominio por intención, que consiste en especificar el tipo de dato correcto (por ejemplo, definir carácter 25 para el nombre del docente) y si acepta o no valores nulos.
2. *Reglas semánticas*, reglas que se definen según el dominio de extensión, que consiste en especificar o enumerar los valores que van a pertenecer al dominio (por ejemplo, condición del docente: ordinario, extraordinario y contratado, según Ley 30220) y de acuerdo con un formato definido o asignado.
3. *Especificación del modelo conceptual*, consiste en mapear los elementos del diagrama de entidad y relación, de acuerdo con algunas especificaciones de requisitos.

Fase 2: Preparación: enfocada a definir las características, métricas, indicadores y criterios de decisión de acuerdo con el contexto, utilizando el estándar ISO/IEC 25 012 e ISO/IEC 25 024. Se identifica a los *stakeholders* que conocen las reglas de negocio que deben cumplir los datos y que pueden afectar a su calidad. Los *stakeholders* pueden ser los responsables finales de usar los datos, analistas de datos y administradores de datos (Caballero *et al.*, 2018).

Esta fase se compone de dos actividades (A2) y (A3), las que se describen a continuación.

- (A2) Especificación de la evaluación

Consiste en seleccionar las características definidas en el modelo de calidad ISO/IEC 25 012, de acuerdo a la importancia de los *stakeholders* o consumidores de datos. Se adoptaron las métricas definidas en el estándar ISO/IEC 25 024. Una métrica define la forma cómo llevar a cabo la medición. Las métricas están orientadas a medir la calidad del valor del dato (nivel de instancia) y la calidad del modelo conceptual de datos (nivel de esquema). En las tablas 5 y 6 se determinan los productos de entrada y salida, así como el artefacto de especificación de métricas.

Tabla 5
Productos de entrada y salida de la actividad (A2) "Especificación de la evaluación"

Productos de entrada	Productos de salida	Técnicas y herramientas
Especificación de requisitos de calidad de datos. Modelo de calidad de referencia ISO/IEC 25 012 Medidas de calidad de referencia ISO/IEC 25 024	Ficha de especificación de métricas. (véase la tabla 4)	Encuesta. Método de análisis comparativo Jim Brosseau.

Elaboración propia.

Tabla 6
Ficha de especificación de métricas

Dimensión de la calidad	Nombre de la dimensión
Tipo	Puede ser: inherente, inherente y dependiente, dependiente
Métrica	Nombre asignado a la métrica de calidad
Propósito	Motivo por el cual se selecciona la métrica
Entrada	Especificación de requisitos
Fórmula y cálculo de datos	Establece la fórmula de medición y especifica los significados de los datos utilizados.
Interpretación	Proporciona el rango y comprensión del significado.
Métodos de aplicación	Manera en que se va aplicar la métrica
Granularidad	Puede ser: celda, tupla, columna, tabla o conjunto de datos.

Elaboración propia

- (A3) Diseño de la evaluación

Esta actividad tiene como objetivo definir la escala de valores de los indicadores y criterios de decisión (niveles de aceptabilidad) como se puede ver en la figura 3. Para cada métrica se define un indicador parcial que indique el valor medido y por cada característica se define un indicador global que determine el nivel de satisfacción alcanzado de acuerdo con la especificación de requisitos; en la tabla 7 se muestran los productos de entrada y salida.



Figura 3. Escala de valores para proceso de evaluación

Fuente: Calahorrano (2007)

Tabla 7

Productos de entrada y salida de la actividad (A3) "Diseño de la evaluación"

Productos de entrada	Productos de salida	Técnicas y herramientas
Especificación de métricas	Indicador y escala de valores Criterios de decisión	Fórmulas de sumarización

Elaboración propia

Fase 3: Implementación: enfocada a aplicar métodos y herramientas de medición de calidad de datos, para obtener las medidas por cada una de las métricas seleccionadas. Esta fase está compuesta de la actividad (A4).

- (A4) Ejecución de la evaluación

El objetivo de esta actividad es realizar la medición y cálculo de indicadores mediante código SQL e informe de resultados. Se generó una plantilla de consultas SQL para medir las métricas. En la tabla 8 se muestran las entradas y salidas de la actividad.

Tabla 8

Productos de entrada y salida de la actividad (A4) "Ejecución de la evaluación"

Productos de entrada	Productos de salida	Técnicas y herramientas
Especificación de métricas e indicadores.	Plantilla de consulta SQL Resultados de la evaluación	MySQL Consultas SQL
Datos y modelo de la base de datos		Procedimientos almacenados

Elaboración propia

A continuación, determinamos la evaluación de dos métricas como ejemplo:

1. Evaluación de la métrica: Exactitud semántica de datos.

Consiste en evaluar los datos con un valor verdadero correspondiente a un dominio referencial, datos existentes y válidos.

Cálculo de la métrica:

$$X = 1 - A/B$$

A = Número de datos incorrectos en su valor.

B = Número de datos para los cuales se requiere la exactitud según su dominio.

Forma de cálculo:

Mediante la siguiente consulta SQL se verificó los valores inexistentes y los valores fuera de un rango referencial de acuerdo con la ficha de especificación de requisitos. Se realizó la suma de todos los valores obtenidos por cada atributo para llegar al resultado de la métrica.

```
SELECT COUNT(*) FROM TABLA WHERE LENGTH(TRIM(ATRIBUTO))=0 OR ATRIBUTO IS NULL OR ATRIBUTO NOT IN (Valor referencial);
```

2. Evaluación de la métrica: Completitud de atributos.

Consiste en evaluar la proporción de valores nulos que existe en cada atributo.

Cálculo de la métrica:

$$X = A/B \text{ (para cada atributo)}$$

A = Conteo de valores no nulos para un atributo

B = Total de valores evaluados por registro

Forma de cálculo:

Se verificó la completitud en los 154 atributos mediante la siguiente consulta SQL.

```
SELECT COUNT(*) FROM TABLA WHERE ATRIBUTO IS NOT NULL OR LENGTH(TRIM(ATRIBUTO))<>0;
```

Fase 4: *Mejora*: enfocada en diseñar un plan de mejora en el que se indican acciones de cambio y se implementan dichas acciones utilizando métodos y herramientas de un proceso denominado limpieza de datos, que se emplea para mejorar la calidad de los datos sin afectar su naturaleza. Luego se mide la mejora introducida de acuerdo a los rangos de aceptación.

Esta fase se compone de dos actividades (A5) y (A6), las cuales se describen a continuación:

- (A5). Plan de mejora

Esta actividad tiene como objetivo producir un plan de mejora en el cual se indiquen las acciones a corregir. Las tablas 9 y 10 muestran los productos de entrada y salida de la actividad, así como el artefacto de especificación del plan de mejora.

Tabla 9

Productos de entrada y salida de la actividad A6 "Implementar el plan de mejora"

Productos de entrada	Productos de salida	Técnicas y Herramientas
Resultados de la evaluación	Plan de mejora	
Ficha de especificación de requisitos		

Elaboración propia

Tabla 10

Ficha del plan de mejora

Código de la especificación de requisitos	Atributo	Acción de cambio	Método/Técnica

Elaboración propia

- (A6). Implementar el plan de mejora.

Esta actividad tiene como objetivo implementar, a partir del plan de mejora, los cambios correspondientes en las entidades o tablas evaluadas, originando como salida una nueva versión del valor del dato y del modelo conceptual de datos evaluados.

En la tabla 11 se muestran los productos de entrada y salida de la actividad.

Una vez efectuados los cambios es necesario medir y evaluar, para comprender la ganancia lograda en la mejora.

Tabla 11
Productos de entrada y salida de la actividad A6 "Implementar el plan de mejora"

Productos de entrada	Productos de salida	Técnicas y herramientas
Plan de mejora	Repositorio de la base de datos limpios.	Herramientas ETL
Plantilla de consulta SQL		Código SQL

Elaboración propia

3.2 Resultados de evaluación de calidad por características

Para obtener los resultados se realizó una prueba de calidad de datos antes y después de aplicar el modelo propuesto SDQAM. Los resultados del Pre-Test se obtuvieron realizando un análisis previo para conocer la calidad de los datos tomando como referencia dimensiones o características de calidad más comunes. Los resultados del Post-Test se obtuvieron implementando el plan de mejora siguiendo la metodología propuesta. Las métricas de cada característica fueron calculadas mediante consultas SQL (en inglés, *Structured Query Language*). Los valores fueron definidos mediante la unidad de porcentaje, que determina el nivel de cumplimiento según el modelo sistemático para estimar la calidad de los sistemas de *software*, mientras más cercano a 1 es aceptable la calidad de datos (Calahorrano, 2007). Los resultados de la medición de los datos en la base de datos del Sistema de Gestión Académica de la UNAMBA, se muestran en la tabla 12.

Tabla 12
Resultados de evaluación por características antes y después de aplicar el modelo SDQAM, en base al promedio de las métricas

Característica	Pre-Test	Post-Test
	Sin modelo	Con el modelo SDQAM
Exactitud	0,80	0,94
Compleitud	0,83	0,88
Consistencia	0,77	0,94
Credibilidad	0,79	0,93
Actualidad	0,30	0,57
Conformidad	0,64	0,81
Precisión	0,74	0,89
Entendibilidad	0,79	0,90
Calidad de datos	0,71	0,86

Elaboración propia

En la figura 4 se muestra el gráfico radar de las características evaluadas antes y después de aplicar el modelo propuesto y basadas en el promedio de las métricas.

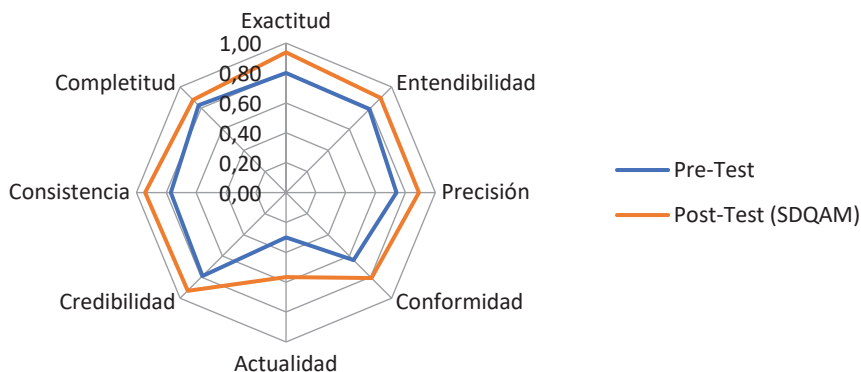


Figura 4. Gráfico radar de las características evaluadas antes y después de aplicar el modelo propuesto SDQAM, basadas en el promedio de las métricas
Elaboración propia

A partir de estos resultados, se puede decir que antes de aplicar el modelo propuesto de calidad de datos, se encontraban en niveles inaceptables y aplicando el modelo se logra mejorar significativamente la calidad de datos, a excepción de la característica de actualidad.

3.3 Resultados de evaluación de calidad del valor del dato por tablas

En la tabla 13 y en la figura 5 se presentan los resultados de las tablas de datos evaluadas sin modelo y con el modelo propuesto.

Tabla 13

Resultados de evaluación por tablas para medir la calidad del valor del dato antes y después de aplicar el modelo SDQAM

Tablas	Pre-Test Sin modelo	Post-Test Con el modelo SDQAM
Facultad	1,00	1,00
Escuela	1,00	1,00
Docente	0,52	0,68
Departamento académico	0,87	1,00
Sede	1,00	1,00
Estudiante	0,61	0,66
Det estudiante	0,68	0,70
Tipo estudiante	1,00	1,00
Modalidad	1,00	1,00
Plan estudio	1,00	1,00
Det plan estudio	0,75	0,86
Asignatura	0,91	0,99
Categoría	0,46	0,48
Modo estudio	0,95	1,00
Semestre	0,82	0,86
Carga académica	0,87	0,97
Detcarga académica	0,69	0,87
Matrícula	0,77	0,87
Det matricula	0,70	0,83
Condición	0,70	0,97
Ubigeo	1,00	1,00
Colegio	1,00	1,00

Elaboración propia

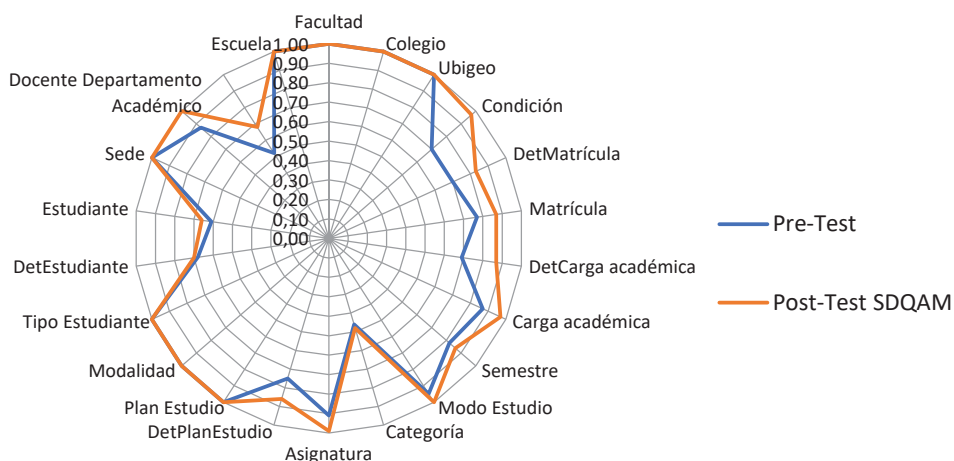


Figura 5. Gráfico radar de tablas evaluadas para medir la calidad del valor del dato antes y después del modelo propuesto SDQAM.

Elaboración propia

3.4 Resultados de evaluación de calidad del esquema conceptual de datos por tablas

En la tabla 14 y en la figura 6 se presentan los resultados de las tablas evaluadas.

Tabla 14

Resultados de evaluación por tablas para medir la calidad del esquema conceptual de datos antes y después de aplicar el modelo SDQAM.

Tablas	Pre-Test	Post-Test
	Sin el modelo propuesto	Con el modelo SDQAM
Facultad	1,00	1,00
Escuela	0,96	1,00
Docente	0,77	1,00
Departamento académico	0,96	0,96
Sede	0,00	1,00
Estudiante	0,66	1,00
Detestudiante	0,66	1,00
Tipo estudiante	0,00	1,00
Modalidad	0,00	1,00
Plan estudio	0,86	0,99
Detplan estudio	0,81	1,00

(continúa)

(continuación)

Asignatura	0,73	0,99
Categoría	0,00	1,00
Modo estudio	0,00	1,00
Semestre	0,76	0,83
Carga académica	0,74	0,99
Detcarga académica	0,57	0,91
Matrícula	0,75	1,00
Det matrícula	0,66	0,91
Condición	0,00	1,00
Ubigeo	1,00	1,00
Colegio	1,00	0,99

Elaboración propia

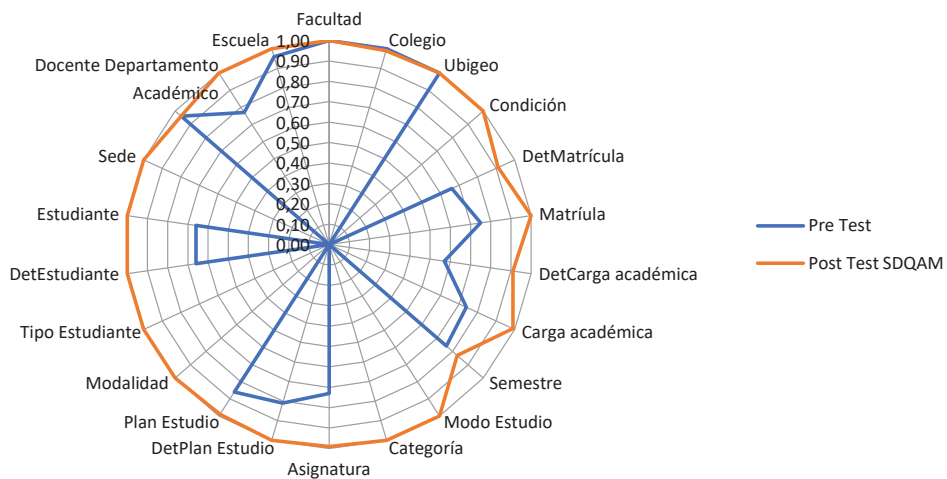


Figura 6. Gráfico radar de tablas evaluadas para medir la calidad del esquema conceptual de datos antes y después del modelo propuesto SDQAM.

Elaboración propia

Al obtener los resultados del proceso de evaluar la calidad del esquema conceptual de datos sin el modelo propuesto, se puede observar que varias tablas tienen un porcentaje de 0 %, esto se debe a que carecen de completitud en la definición de restricciones en los atributos y que varios elementos de la estructura del modelo no son entendibles, los mismos que se lograron mejorar.

4. DISCUSIÓN

La presente investigación ha conceptualizado un modelo de evaluación de calidad de datos tomando como base para la elaboración el estándar ISO/IEC 25 012, desde la perspectiva del consumidor de datos y desde la visión de la calidad de datos inherentes. Esto se basa en el sentido en que los consumidores de datos necesitan usar los datos, cuyos valores se espera que sean correctos, actualizados y completos; y que la evaluación de los datos vaya más allá de aspectos tecnológicos.

En la propuesta del modelo de evaluación se detallan los artefactos utilizados; se integró dos modelos: el modelo de evaluación de *software* ISO/IEC 25 040 (ISO/IEC 25 040, 2011) y el modelo GOCAME (Becker *et al.*, 2015), Estrategia integrada de Medición y Evaluación enfocada en un contexto y orientada a objetivos. Por lo que, el objetivo del modelo propuesto es evaluar y mejorar la calidad de los datos. Por su parte, Calabrese *et al.* (2019), determinó una guía para medir y evaluar la calidad de datos con énfasis en el proceso definido por ISO/IEC 25 040, concluyendo este proceso solo en un informe final de resultados.

Para la selección de características de calidad de datos, Fritz *et al.* (2017), plantean evaluar la calidad de productos de datos almacenados enmarcados netamente en aspectos de seguridad de datos, tomando como base las características de calidad de la norma ISO/IEC 25 010 e ISO/IEC 25 040 como proceso de evaluación. La selección de características o atributos propuestos por Fritz depende mucho de la calidad del *software*, a diferencia de nuestro enfoque que está orientado a seleccionar características inherentes, implícitas en el dato y en el esquema conceptual de datos, casi obligatorias de un sistema de gestión académica, características que fueron seleccionadas según la importancia del consumidor de datos.

Manrique (2017) desarrolla un catálogo de reglas de negocios basado en ISO/IEC 25 012, utilizando sentencias SQL, reutilizables y procesables por cualquier gestor de base de datos. En nuestro estudio, generamos plantillas de consultas SQL que pueden ser instanciadas de acuerdo con las especificaciones de requisitos y pueden ser aplicados a otros contextos.

Realizando un análisis de los resultados obtenidos al evaluar la calidad del valor del dato se puede obtener análisis importantes de los registros catalogados como datos de mala calidad o datos inaceptables en el sistema SISGA, como, por ejemplo, en los datos de los docentes y estudiantes con porcentajes inaceptables de calidad de datos, porque contienen datos inconsistentes, no estandarizados, desactualizados, generando inconvenientes en los consumidores de datos para poder usar los datos con nuevas aplicaciones.

5. CONCLUSIONES

Si bien la universidad cuenta con un sistema de gestión académica considerado como una fortaleza de la institución, esto se debilita al no contar con datos de calidad y esto se evidencia al evaluar la calidad de datos con características o dimensiones de calidad más comunes, donde se obtiene un 71 % de datos calificados como inaceptables. En consecuencia, la universidad, en cuanto a la calidad de datos, no le dio la debida importancia. Con la aplicación del modelo propuesto se obtiene un porcentaje del 86 %, calificado como aceptable, lo que mejora significativamente la calidad de datos. Por lo que tomar como referencia el estándar ISO/IEC 25 012 como modelo de calidad de datos fue muy importante para garantizar y asegurar la calidad de datos y así poder confiar en ellos al hacer uso de los mismos.

Es necesario trabajar de la mano con la especificación de requisitos, esto debido a que ciertas métricas de calidad lo requieren como insumo para su medición. El resultado obtenido de la mejora en el valor del dato se debe a las consideraciones establecidas en el reglamento de matrícula, reglamento general académico de la UNAMBA y regulaciones establecidas en la Ley 30220, definidas como reglas y restricciones de dominio.

Se tiene que considerar como un factor clave en la calidad de los datos, la calidad del modelo conceptual de los datos a nivel de esquema; esto se debe a que varias restricciones de datos no fueron implementadas en el esquema de la base de datos, razón por lo cual se incurre en el ingreso y registro de datos defectuosos. Por lo que, al aplicar el modelo de evaluación propuesto, mejora significativamente la calidad del modelo conceptual.

6. RECOMENDACIONES

Se recomienda a las universidades un cambio cultural que se centre en el valor del análisis de datos, teniendo en cuenta los beneficios que se pueden obtener a partir de datos de calidad.

REFERENCIAS

- Becker P., Papa F., y Olsina L. (2013) Enhancing the Conceptual Framework Capability for a Measurement and Evaluation Strategy. En Q.Z. Sheng, y J. Kjeldskov, *Lecture Notes in Computer Science*, 8295, 104-116. Springer. https://doi.org/10.1007/978-3-319-04244-2_11
- Clarrus Consulting Group Inc. (2010). *Software Quality Attributes: Following All the Steps*.
- Caballero, I., Rodriguez, M., y Fernandez, C. (Noviembre del 2017). Calidad de datos digitales certificada. *Aenor*, (331), 36-39. <https://portal.aenormas.aenor.com/revista/331/calidad-datos-certificada.html>
- Caballero, I., Gómez, A., Gualo, F., Merino, J., Rivas, B., y Piattini, M. (2018). *Calidad de Datos*. Editorial RAMA.
- Calabrese, J., Esponda, S., Pasini, A. C., Boracchia, M., y Pesado, P. M. (Octubre del 2019). Guía para evaluar calidad de datos basada en ISO/IEC 25012. XXV Congreso Argentino de Ciencias de la Computación, 694-706. <http://sedici.unlp.edu.ar/handle/10915/91086>
- Calahorrano, A. I. (2007). Herramienta para evaluar la calidad del código fuente generado en C ANSI [Tesis de Título Profesional, Escuela Politécnica Nacional]. Repositorio Digital Institucional de la Escuela Politécnica Nacional. <http://bibdigital.epn.edu.ec/handle/15000/559>
- Congreso de la República de Perú Ley N.º 29733, *Ley de Protección de Datos Personales*. (2011). http://www.pcm.gob.pe/transparencia/Resol_ministeriales/2011/ley-29733.pdf
- DalleMule, L., y Davenport T. (2017). *What's your Data Strategy?*. United States: Harvard Business Review. <https://hbr.org/2017/05/whats-your-data-strategy>.
- Fritz, E., Montejano, G., y García, P. (2017). Selección de atributos de calidad de datos en sistemas de gestión de aprendizaje bajo la familia de normas ISO/IEC 25000. *XIX Workshop de Investigadores en Ciencias de la Computación (WICC)*, 627-631. <http://sedici.unlp.edu.ar/handle/10915/62178>
- Hernández R., Fernandez, C., y Baptista, P. (2014). *Metodología de la investigación* (6.ª ed.). Mc Graw Hill.
- ISO25000. (30 de septiembre del 2021). *La familia de normas ISO/IEC 25000*. <http://iso25000.com/index.php/normas-iso-25000>
- ISO/IEC 25012. (2008). ISO/IEC 25012:2008 Software engineering –Software product Quality Requirements and Evaluation (SQuaRE)– Data Quality model. <https://www.iso.org/standard/35736.html>

- ISO/IEC 25040. (2011). ISO/IEC 25040: 2011 System and Software Engineering – Systems and Software Product Quality Requirements and Evaluation (SQuaRE) – Evaluation Process. International Organization for Standardization. <https://www.iso.org/standard/35765.html>
- Ley 30220, Ley Universitaria 30220. (2014). Diario *El Peruano*. <https://diariooficial.elperuano.pe/pdf/0021/ley-universitaria-30220.pdf>
- Manrique, A. (2017). *Desarrollo de un catálogo de reglas de negocio referentes a datos, basado en ISO/IEC 25012 y SBVR*. [Tesis de Magister, Universidad de Castilla-La Mancha]. Repositorio institucional de la UCLM. <http://hdl.handle.net/10578/16697>
- UNAMBA (2018). *Reglamento de matrícula*. Universidad Nacional Micaela Bastidas de Apurímac, Perú. <https://www.unamba.edu.pe/transparencia/transparencia-universitaria/reglamentosacademicoss/reglamento-de-matr%C3%ADcula-2.html>
- Zhang, R., Indulska, M., y Sadiq, S. (Julio, 2019). Discovering Data Quality Problems. *Business y Information Systems Engineering*, 61, 575–593. <https://doi.org/10.1007/s12599-019-00608-0>