

Modelo de Calidad de Datos Académicos Basado en ISO/IEC 25000

Academic Data Quality Model Based on ISO/IEC 25000

Presentación: 13 y 14 de septiembre de 2023

Claudio Carrizo

Universidad Tecnológica Nacional – Facultad Regional San Francisco
cjcarrizo77@gmail.com

Javier Saldarini

Universidad Tecnológica Nacional – Facultad Regional San Francisco
saldarinijavier@gmail.com

Marta Cabrera Villafañe

Universidad Nacional de La Rioja - La Rioja – Argentina
marta.cabreravilla@gmail.com

Carlos Salgado

Facultad de Ciencias Físico - Matemáticas y Naturales Universidad Nacional de San Luis - San Luis – Argentina
csalgado@unsl.edu.ar

Alberto Sánchez

Facultad de Ciencias Físico - Matemáticas y Naturales Universidad Nacional de San Luis - San Luis – Argentina.
alfanego@unsl.edu.ar

Mario Peralta

Facultad de Ciencias Físico - Matemáticas y Naturales Universidad Nacional de San Luis - San Luis – Argentina.
mperalta@unsl.edu.ar

Resumen

Los datos representan actualmente un recurso estratégico para cualquier organización, ya que, a partir de ellos, se puede descubrir conocimiento y apoyar el proceso de toma de decisiones. Los datos con baja calidad afectan de diferentes maneras la gestión institucional, es por este motivo que resulta imprescindible poder medir y evaluar la calidad de los mismos, a través de sus características. En este artículo se desarrolla un modelo de calidad basado en el estándar internacional ISO/IEC 25000, el cual está compuesto por un conjunto de características y medidas de calidad en relación a los datos, que servirán para una posterior evaluación. Los resultados del presente trabajo permitirán contribuir en reconocer la calidad que tienen los datos almacenados en el sistema de gestión académica de UTN San Francisco, con el fin de establecer acciones en pos de mejorar la calidad de datos.

Palabras clave: modelo de calidad, ISO/IEC 25000, datos académicos, utn

Abstract

Data currently represents a strategic resource for any organization, since it is possible to discover knowledge and support the decision-making process. Low-quality data affects institutional management in different ways, which is why it is essential to be able to measure and evaluate their quality through their characteristics. In this work, a quality model based on the international standard ISO/IEC 25000 is developed, which is composed of a set of characteristics and quality measures in relation to the data, which will be used for further evaluation. The

results of this work will help to recognize the quality of the data stored in the academic management system of UTN San Francisco, in order to establish actions to improve the quality of data.

Keywords: quality model, ISO/IEC 25000, academic data, utn

Introducción

Los Sistemas de Información (SI) están presentes en todo tipo de organizaciones y son un componente fundamental para el éxito de los negocios, ya que su implementación permite obtener una ventaja competitiva (Soto, 2014). Según la Norma ISO/IEC 25030:2007, en un modelo de sistema, los sistemas informáticos son parte de un SI, y están compuestos por cuatro elementos principales: hardware, sistema operativo, software de aplicación y datos. Los datos son hechos, valores o símbolos aislados que se encuentran de manera simple o cruda (Bernstein, 2009), que por sí solos, carecen de significado (Batini et. all, 2009).

En (Bernstein, 2009) se representa la pirámide del conocimiento, en donde se denota que el conocimiento es la adquisición del significado o análisis de la información necesaria para la toma de decisiones; los datos se encuentran en el nivel más bajo de la jerarquía de la pirámide, por lo tanto, sin la existencia de datos, no se llegaría nunca a obtener información, y por supuesto, mucho menos conocimiento. En (Beg & Hussain, 2008) se menciona que no basta la existencia de datos, ni la voluntad de basar las decisiones que se tomen sobre ellos, sino que también es crucial que éstos tengan la calidad adecuada.

En la Norma ISO/IEC 25012:2008 se define la calidad de datos como el "grado en que las características de los datos satisfacen necesidades implícitas y establecidas cuando son usados en condiciones específicas". Estas características o cualidades que deben poseer los datos para ser considerados como adecuados se denominan "dimensiones o características de calidad de los datos" (Pipino et. all, 2002). Esto quiere decir que la calidad de los datos está asociada a un conjunto de dimensiones que son los que la definen. Un objetivo fundamental de la definición de las dimensiones es poder establecer un lenguaje común, y también focalizar los problemas de calidad de datos, junto con las oportunidades de mejora (Javed & Hussain, 2003; Naveh & Halevy, 2000).

Por otra parte, existe actualmente el estándar ISO/IEC 25000, el cual provee dos normas que se enfocan específicamente hacia la calidad de los datos; la Norma ISO/IEC 25012 (ISO/IEC 25012, 2008) provee un conjunto de características de calidad de datos, las cuales tienen asociadas un conjunto de medidas de calidad de datos, provistas por la Norma ISO/IEC 25024 (ISO/IEC 25024, 2015).

En el ámbito de la UTN Facultad Regional San Francisco, la gestión académica de los alumnos de carreras de ingenierías se lleva adelante a través del uso de un sistema de información denominado SYSACAD. Según lo descrito anteriormente, y dada la importancia que tienen estos datos, es que en este trabajo se propone construir un modelo de calidad de datos académicos de alumnos de carreras de ingenierías de la UTN Facultad Regional San Francisco, por medio del uso del estándar ISO/IEC 25000. Este modelo contendrá un conjunto de características y medidas de calidad de datos asociados atributos académicos. La principal contribución de este trabajo consiste en reconocer la calidad que tienen los datos almacenados en el sistema de gestión de alumnos (SYSACAD), a fin de luego poder establecer acciones de limpieza y/o corrección, que conduzcan a la mejora de los mismos.

En lo sucesivo, el presente trabajo se divide en las siguientes secciones. En la Sección 2 se detalla la metodología utilizada. En la Sección 3 se realiza un estudio de la serie ISO/IEC 25000. En la Sección 4 se presenta el modelo de calidad de datos, instanciado a través de un caso de estudio. En la Sección 5 se elaboran las conclusiones.

2. Metodología

Para el presente trabajo se llevó adelante una investigación aplicada, donde el objeto de estudio fue más bien de tipo exploratorio.

Técnica de Relevamiento de Datos

Para el relevamiento de los datos, se envió al Área TICs de la UTN Facultad Regional San Francisco, un listado de atributos o campos asociados a datos de exámenes finales de alumnos de todas las carreras de ingenierías. Los registros de datos que fueron extraídos del sistema SYSACAD, fueron volcados a un documento en formato CSV.

Población y Muestra

Se trabajó con una población de 181053 registros de datos en total, que pertenecen a alumnos que han rendido exámenes finales entre los años 1970 y 2022. Del conjunto total se obtuvo una muestra de 37752 registros que corresponden a los siguientes años en los que los alumnos rindieron examen: 1996, 2005, 2010, 2020 y 2021. Cabe aclarar que estos 5 años fueron seleccionados al azar.

3. Normas ISO/IEC 25000

La Serie ISO/IEC 25000, también conocida como “SQuaRE” (del inglés, System and Software Quality Requirements and Evaluation), proporciona una guía que permite la definición de requisitos y evaluación de la calidad de sistemas y del software. En el marco de este artículo, se presenta en detalle las siguientes normas: ISO/IEC 25012 y ISO/IEC 25024.

3.1. Modelo de Calidad de Datos

Esta Norma Internacional se centra en la calidad de los datos como parte de un sistema informático, donde define un modelo general de la calidad para datos conservados en formato estructurado. Este modelo clasifica los atributos de la calidad en quince características consideradas por dos puntos de vista:

- **Calidad de Datos Inherentes:** se refiere al grado en que las características de la calidad de los datos tienen el potencial intrínseco de satisfacer las necesidades explícitas e implícitas cuando los datos se utilizan en condiciones especificadas.

- **Calidad de Datos Dependientes del Sistema:** se refiere al grado en que se alcanza y preserva la calidad de los datos dentro de un sistema informático cuando estos se utilizan en condiciones especificadas.

En la Figura 1, se pueden visualizar las 15 características que contiene la norma, clasificadas desde 2 puntos de vista:

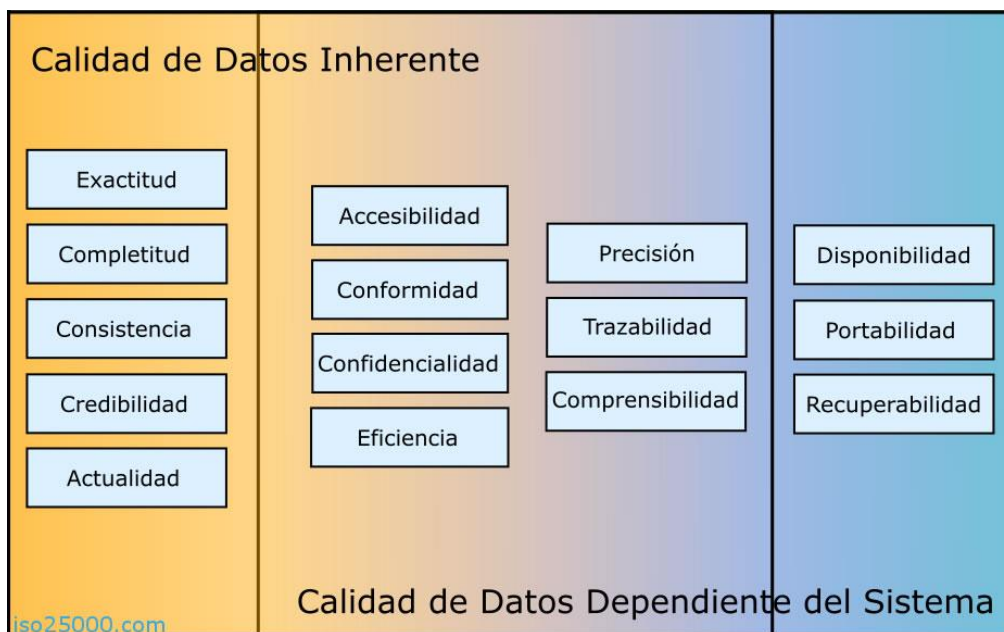


Figura 1. Características de Calidad de Datos (extraído de iso25000.com)

3.2. Medidas de Calidad de los Datos

La Norma ISO/IEC 25024 define un conjunto de medidas que permiten medir cuantitativamente la calidad de los datos, en términos de características definidas en la Norma ISO/IEC 25012. Contiene un conjunto de medidas de calidad de los datos para cada característica, las cuales se seleccionan de acuerdo al contexto en el cual se aplican. Cada medida está asociada a un valor objetivo, a elementos de medidas de la calidad, y a un método de medición.

4. Modelo de Calidad de Datos Académicos

Para obtener como resultado el modelo de calidad de datos asociados a atributos académicos se llevaron adelante las siguientes actividades que se describen en las subsecciones 4.1, 4.2, 4.3 y 4.4.

4.1. Actividad 1: Relevamiento y Descripción de Atributos Académicos

Se relevaron los atributos correspondientes a datos de exámenes finales rendidos por alumnos de ingenierías. Los atributos identificados fueron los siguientes: legajo, nombre y apellido, estado, año de ingreso, código especialidad, nombre materia, fecha examen, nota, aprobado. Como tarea adicional, también se realizó la descripción de cada atributo, en un diccionario de datos. Cabe destacar que dichos atributos luego deberán ser contrastados con la norma ISO/IEC 25012 e ISO/IEC 25024.

4.2. Actividad 2: Estudio e Identificación de Características y Medidas de Calidad

Se realizó un estudio de las normas asociadas a calidad de los datos (ISO/IEC 25012 e ISO/IEC 25024), en donde como resultado del mismo, se obtuvo una tabla compuesta por características y medidas de calidad de datos. En la Tabla 1 se puede observar un ejemplo.

Tabla 1. Relación entre característica y calidad de datos

Característica de Calidad (Punto de Vista)	Medida de Calidad (Punto de Vista)
Credibilidad (Inherente)	Credibilidad de Valores (Inherente) Credibilidad de Fuentes (Inherente) Credibilidad de Diccionario de Datos (Inherente) Credibilidad de Modelo de Datos (Inherente)
Confidencialidad (Inherente / Dependiente del Sistema)	Uso de Cifrado (Inherente) No vulnerabilidad (Dependiente del Sistema)
Recuperabilidad (Dependiente del Sistema)	Copia de Seguridad Periódica (Dependiente del Sistema)

4.3. Actividad 3. Mapeo de atributos con características y medidas de calidad de datos

Tomando como referencia la Tabla 1, se identificaron las características y medidas de calidad que aplicaban para cada atributo. También se dejó registro de aquellas características y/o medidas de calidad que no aplicaban. En la Tabla 2 se puede ver un ejemplo concreto.

Tabla 2. Aplicación de características y medidas de calidad para el atributo “Legajo”

Atributo: Legajo			
Característica de Calidad	Medida de Calidad	¿Aplica? (Sí/No)	Motivo (Aplica / No Aplica)
Compleitud	Compleitud de Registros	NO	
	Compleitud de Atributos	NO	
	Compleitud de archivos de datos	NO	
	Compleitud de valores de datos	SI	
Precisión	Precisión de Valores de Datos	SI	
	Precisión de Formato de Datos	SI	

4.4. Actividad 4. Obtención de un Modelo de Calidad de Datos

Finalmente se pudo obtener un modelo de calidad de datos académicos, en base a atributos que corresponden a exámenes finales de materias de ingenierías. El resultado de esta actividad se puede visualizar en la Tabla 3.

Tabla 3. Modelo de Características de Calidad de Datos junto con sus medidas asociadas

Atributos Académicos	Características de Calidad de Datos (Norma ISO/IEC 25012)	Medidas de Calidad de Datos (Norma ISO/IEC 25024)
Legajo	Compleitud	Compleitud de Valores de Datos
	Exactitud	Exactitud Sintáctica de Valores de Datos
	Exactitud	Exactitud de Valores de Datos en un Rango
	Precisión	Precisión de Valores de Datos
	Precisión	Precisión de Formato de Datos
Nombre y Apellido	Compleitud	Compleitud de Valores de Datos
	Exactitud	Exactitud Sintáctica de Valores de Datos
Año de Ingreso	Compleitud	Compleitud de Valores de Datos
	Exactitud	Exactitud Sintáctica de Valores de Datos
	Exactitud	Exactitud de Valores de Datos en un Rango
	Precisión	Precisión de Valores de Datos
	Precisión	Precisión de Formato de Datos
Código especialidad	Credibilidad	Credibilidad de Valores de Datos
	Compleitud	Compleitud de Valores de Datos
	Exactitud	Exactitud Sintáctica de Valores de Datos
	Exactitud	Exactitud de Valores de Datos en un Rango
	Precisión	Precisión de Valores de Datos
Nombre materia	Precisión	Precisión de Formato de Datos
	Credibilidad	Credibilidad de Valores de Datos
	Conformidad	Conformidad de Valores de Datos
	Compleitud	Compleitud de Valores de Datos
	Exactitud	Exactitud Sintáctica de Valores de Datos
Fecha examen	Conformidad	Conformidad de Valores de Datos
	Compleitud	Compleitud de Valores de Datos
	Exactitud	Exactitud Sintáctica de Valores de Datos
	Precisión	Precisión de Formato de Datos
Nota	Comprensibilidad	Comprensibilidad de Valores de Datos
	Compleitud	Compleitud de Valores de Datos
	Exactitud	Exactitud Sintáctica de Valores de Datos
	Exactitud	Exactitud de Valores de Datos en un Rango
	Precisión	Precisión de Valores de Datos
Aprobado	Precisión	Precisión de Formato de Datos
	Credibilidad	Credibilidad de Valores de Datos
	Compleitud	Compleitud de Valores de Datos
	Exactitud	Exactitud Sintáctica de Valores de Datos
	Exactitud	Exactitud Sintáctica de Valores de Datos
Estado	Precisión	Precisión de Valores de Datos
	Precisión	Precisión de Formato de Datos
	Compleitud	Compleitud de Valores de Datos
	Exactitud	Exactitud Sintáctica de Valores de Datos
	Precisión	Precisión de Valores de Datos
	Credibilidad	Credibilidad de Valores de Datos
	Credibilidad	Credibilidad de Valores de Datos

En la Figura 2 A continuación, se puede observar el modelo general de calidad de datos obtenido para los atributos académicos identificados.



Figura 2. Modelo de Calidad de Datos Resultante (elaboración propia)

Conclusiones

La calidad de los datos es crucial en cualquier contexto o ámbito; en este sentido, los datos almacenados en el sistema de gestión académica de UTN Facultad Regional San Francisco, deben mantener un cierto nivel de calidad, a fin de poder asegurar la correcta gestión de la información y la toma de decisiones.

En este trabajo se ha iniciado un proceso de evaluación de calidad de los datos almacenados en el sistema SysAcad, usando como marco de referencia el estándar ISO/IEC 25000, específicamente las Normas ISO/IEC 25040, ISO/IEC 25012 e ISO/IEC 25024.

Se ha presentado también la instanciación del proceso de evaluación de calidad a través de un caso concreto. Como puede observarse en la Tabla 5, se obtuvo un resultado de medición “aceptable” para la medida de calidad de datos “Completitud de Datos”, mientras que la medida “Exactitud de Datos en rango” el resultado fue “inaceptable”. En este último caso, y dada la relevancia de poder obtener valores correctos dentro un rango (de 2 a 10) para el dato “Nota Examen Final”, será necesario implementar un conjunto de acciones sobre el sistema SysAcad, a fin de que pueda garantizar la exactitud de los valores de datos, en término de los rangos correctos.

Referencias

- F. Soto, “Análisis de la problemática asociada con la baja calidad de datos en los sistemas de información”, Tesis de Maestría, Sistemas Informáticos, E.T.S.I. de Sistemas Informáticos (UPM), 2014.
- ISO/IEC 25030:2007. Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Quality requirements.
- J. H. Bernstein, "The Data-Information-Knowledge-Wisdom Hierarchy and its Antithesis", CUNY Kingsborough Community College, 2009.
- C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, “Methodologies for data quality assessment and improvement,” ACM Computing Surveys, vol. 41, no. 3, 2009.
- J. Beg and S. Hussain, “Data Quality-A problem and An Approach”, Wipro Technologies, 2008.
- ISO/IEC 25012:2008 Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model, ISO, 2012.
- R. Y. Leo, L., Pipino, L. Yang, W. L., & Wang, “Data quality assessment”, Commun. ACM, vol. 45, no. 4, pp. 211–218, 2002.
- Javed, B. & Hussain, S. (2003). Data quality – A problem and an approach. Wipro Technologies.
- Naveh, E. & Halevy, A. (2000). A hierarchical framework for a quality information system. Total Quality Management, Vol. 11, No. 1, p 87-111.
- “ISO/IEC 25000:2014, Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE), ISO-Guide to SQuaRE”, ISO, 2014.
- [24] “ISO/IEC 25024:2015 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality”, ISO, 2015.