

# Evaluación de Calidad de Datos en el ámbito de un Sistema de Gestión Académico mediante ISO/IEC 25000

## Data Quality Assessment in the Scope of an Academic Management System through ISO/IEC 25000

Presentación: 26 y 27 de octubre de 2022

### **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 son un recurso estratégico para cualquier organización, es por esto que se requiere que los mismos tengan la calidad adecuada. La mala calidad de datos afecta de diferentes maneras la gestión institucional, por este motivo resulta fundamental poder medir y evaluar la calidad que deben tener los datos, a través de ciertas características o dimensiones. El propósito del presente trabajo consiste en iniciar un proceso de evaluación de calidad de los datos que se encuentran almacenados en el Sistema de Gestión Académica (SysAcad), haciendo uso del estándar ISO/IEC 25000. Los resultados de este proyecto permitirán identificar el nivel de aceptabilidad en cuanto a la calidad que poseen los datos, como así también, a establecer acciones de limpieza y/o corrección, que conduzcan hacia la mejora de los mismos.

**Palabras clave:** evaluación de calidad de datos, ISO/IEC 25000, sistema de gestión académica

## **Abstract**

Data is a strategic resource for any organization, which is why it is required that they have adequate quality. The poor quality of data affects institutional management in different ways, for this reason it is essential to be able to measure and evaluate the quality that the data should have, through certain characteristics or dimensions. The purpose of this work is to start a process of evaluating the quality of the data stored in the Academic Management System (SysAcad), using the

ISO/IEC 25000 standard. The results of this project will allow us to identify the level of acceptability in terms of the quality of the data, as well as to establish cleaning and/or correction actions that lead to their improvement.

**Keywords:** Data Quality Assessment, ISO/IEC 25000, Academic Management System

## 1. Introducción

Los datos son “un término general para denotar alguno o todos los hechos, letras, símbolos y números referidos a, o que describen un objeto, idea, situación, condición u otro factor” (Maynard, 1982), y constituyen un elemento fundamental para la toma de decisiones objetivas en todos los niveles de una organización (Javed y Hussain, 2003; Levy, 2004; Naveh y Halevy, 2000). Es más, para una organización moderna, los datos constituyen uno de sus recursos estratégicos (Olson, 2002; Tayi y Ballou, 1998). Es importante destacar que el uso de datos como base para la toma de decisiones ha sido una práctica ampliamente recomendada, en contraposición al hecho de desarrollar la toma de decisiones con base en la intuición; el enfoque basado en hechos para la toma de decisión, es uno de los principios de gestión de la calidad, el cual plantea que las decisiones eficaces se basan en el análisis de los datos y la información (Norma Cubana ISO 9000, 2000).

En (Gil-Aluja, 2000; Javed y Hussain, 2003) se menciona que no es suficiente la existencia de datos ni la voluntad de basar las decisiones que se toman, además se requiere que éstos datos tengan la calidad adecuada. Es decir que, las decisiones que se tomen estén basadas en los datos, y se tenga la certeza de que estos datos estén libres de errores y posean atributos relevantes (Redman, 2001). La mala calidad de los datos afecta de diversas maneras la gestión institucional, provocando principalmente efectos negativos sobre la toma de decisiones.

El término calidad en relación con los datos toma sentido por el hecho de que los datos, al igual que los productos y servicios, deben adecuarse al uso que se les pretende dar. El término preciso para el uso en este caso implica que, dentro de cualquier contexto operacional, el dato que va a ser utilizado satisfaga las expectativas de los usuarios de los datos. Dichas expectativas se satisfacen en gran medida si los datos son útiles para lo que estos los necesitan, son fáciles de entender e interpretar, y además son correctos (Loshin, 2001; Redman, 2001). De estas definiciones se puede deducir que la calidad de los datos es un concepto relativo. Según la Norma ISO/IEC 25012 (ISO/IEC 25012, 2008), la calidad de datos es 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” (Abate, 1998; Pipino y otros, 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 los datos y las oportunidades de mejora (Javed y Hussain, 2003; Naveh y Halevy, 2000).

En este sentido, 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). Por otra parte, la Norma ISO/IEC 25040 (ISO/IEC 25040, 2011) proporciona una guía de referencia para llevar a cabo el proceso de evaluación de la calidad de los datos.

Según lo descrito anteriormente, y dada la importancia que tienen los datos en el contexto de un sistema de gestión académico, en este trabajo se propone iniciar un proceso de evaluación de calidad de datos de alumnos de carreras de ingenierías de la UTN Facultad Regional San Francisco, utilizando el estándar ISO/IEC 25000 (ISO/IEC 25000, 2014).

La principal contribución de este trabajo consiste en reconocer la calidad que tienen los datos almacenados en el sistema de gestión de alumnos de la UTN Facultad Regional San Francisco (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 proceso de evaluación 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 a cabo una investigación de tipo aplicada, donde el nivel de profundización en el objeto de estudio será de tipo exploratorio. Se tomó como población el registro de datos de todos los atributos de alumnos de carreras de ingenierías de la UTN Facultad Regional San Francisco desde el año 1970 hasta el año 2022. La recolección de los datos de alumnos del sistema SYSACAD se realizó a través un archivo de tipo csv. En primera instancia, se realizó una exploración de los datos, donde se utilizó la herramienta “RapidMiner”, a fin de poder visualizar el estado preliminar de los mismos. Luego se procedió a ejecutar el proceso de evaluación de calidad de estos datos, registrando los resultados obtenidos en un documento.

## 3. Estudio del Estándar ISO/IEC 25000

La Familia de Normas 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 trabajo, se presenta en detalle las siguientes normas: ISO/IEC 25012, ISO/IEC 25024 e ISO/IEC 25040.

### 3.1. Norma ISO/IEC 25012: 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 15 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.

### 3.2. Norma ISO/IEC 25024: Medidas de Calidad de los Datos

Esta Norma Internacional 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.

### 3.3. Norma ISO/IEC 25040: Proceso de Evaluación

Proporciona una descripción del proceso de evaluación de la calidad del producto de software/datos, y establece los requisitos para la aplicación de este proceso. Dicho proceso de evaluación consta de un total de cinco actividades, las cuales se describen a continuación:

- **Actividad 1: Establecer los requisitos de la evaluación**
  - Tarea 1.1. Establecer el propósito de la evaluación
  - Tarea 1.2. Obtener requisitos de calidad del producto
  - Tarea 1.3. Identificar las partes del producto que se deben evaluar
  - Tarea 1.4. Definir el rigor de la evaluación
- **Actividad 2: Especificar la Evaluación**
  - Tarea 2.1. Seleccionar los módulos de la evaluación
  - Tarea 2.2. Definir criterios de decisión para las métricas
  - Tarea 2.3. Definir criterios de decisión para la evaluación
- **Actividad 3: Diseñar la Evaluación**
  - Tarea 3.1. Planificar las actividades de la evaluación
- **Actividad 4: Ejecutar la Evaluación**

- Tarea 4.1. Realizar las mediciones
- Tarea 4.2. Aplicar los criterios de decisión para las métricas
- Tarea 4.3. Aplicar los criterios de decisión para la evaluación
- **Actividad 5: Concluir la Evaluación**
  - Tarea 5.1. Revisar los resultados de la Evaluación
  - Tarea 5.2. Crear el informe de evaluación
  - Tarea 5.3. Revisar la calidad de la evaluación
  - Tarea 5.4. Tratar los datos de la evaluación

#### 4. Instanciación del proceso de evaluación de calidad sobre datos del Sistema SysAcad

En esta sección se pretende mostrar a través de un caso de estudio acotado, como se llevó adelante del proceso de evaluación de calidad de los datos del sistema SysAcad. Para ejemplificar la instanciación, se tomará como referencia los datos de exámenes finales de alumnos de carreras de ingenierías, dictadas en la UTN Facultad Regional San Francisco. Por cuestiones de espacio, se presentarán a continuación las actividades más relevantes del proceso de evaluación.

##### 4.1. Actividad 1: Establecer los requisitos de la evaluación

###### *Tarea 4.1.1. Establecer el propósito de la evaluación*

La evaluación de calidad de datos del sistema SysAcad se realiza con el propósito de reconocer la calidad de los datos del sistema de gestión académica de UTN Facultad Regional San Francisco, a fin de iniciar un plan de acciones de mejora sobre los mismos.

###### *Tarea 4.1.2. Obtener requisitos de calidad*

Tomando como referencia las Normas ISO/IEC 25030 (ISO/IEC 25030,2007), ISO/IEC 25012 e ISO/IEC 2524, y de acuerdo a los resultados preliminares obtenidos en la exploración de los datos, se definió un conjunto de Requisitos de Calidad de Datos (RCD). Por cuestiones de espacio, en este artículo se trabajará con los siguientes requisitos de calidad en relación a los datos: “Complejidad de Valores de Datos” y “Exactitud de valores de datos en un rango”.

###### *Tarea 4.1.3. Identificar las partes del producto que se deben evaluar*

Por cuestión de espacio, en este caso de estudio se realizará la evaluación de calidad para el dato “Nota Examen Final”.

##### 4.2. Actividad 2: Especificar la Evaluación

###### *Tarea 4.2.1. Seleccionar los módulos de la evaluación*

De acuerdo a los requisitos de calidad mencionados en la Tarea 4.1.2, se definieron las medidas de calidad de datos correspondientes, las cuales se exhiben en la Tabla 1.

Tabla 1. Requisitos, Características y Medidas de Calidad de Datos

Requisitos de Calidad de Datos	Características de Calidad de Datos	Medidas de Calidad de Datos
Complejidad de Valores de Datos	Complejidad	Complejidad de Datos
Exactitud de valores de datos en un rango	Exactitud	Exactitud de Datos en rango

###### *Tarea 4.2.2. Definir criterios de decisión para las métricas*

Se definieron los criterios de decisión para las medidas de calidad expuestas en la Tarea 4.2.1, los cuales pueden visualizarse en la Tabla 2.

Tabla 2. Criterios de decisión definidos para medidas de calidad del dato “Nota Examen Final”

Dato	Medida de Calidad de Datos	Criterio de Aceptación
	Complejidad de Datos	$0 \leq X \leq 1$ Lo más cerca de 1 es lo mejor <ul style="list-style-type: none"> <li>● Aceptable (si <math>X \geq 0,99</math>) / Inaceptable (si <math>X &lt; 0,99</math>)</li> </ul>

Nota Examen Final	Exactitud de Datos en Rango	$0 \leq X \leq 1$ Lo más cerca de 1 es lo mejor <ul style="list-style-type: none"> <li>• Aceptable (si <math>X \geq 0,99</math>) / Inaceptable (si <math>X &lt; 0,99</math>)</li> </ul>
-------------------	-----------------------------	--

### 4.3. Actividad 4: Ejecutar la Evaluación

#### Tarea 4.3.1. Realizar las mediciones

En esta tarea se llevaron a cabo las mediciones de calidad para el dato “Nota Examen Final”, para obtener como resultado la valoración para cada medida. A continuación, en la Tabla 3 se provee un marco que permite guiar la ejecución de las mediciones, mientras que en la Tabla 4 se exhiben los valores obtenidos para cada medida.

Tabla 3. Marco de Referencia para la ejecución de las métricas de calidad para el dato “Nota Examen Final”

Medidas de Calidad	Vista	Función de Medición	Elementos de la Medición	Propiedad
Compleitud de Datos	Inherente	$X=A/B$	A= número de elementos de datos asociados con valores completos B= número de elementos de datos en total.	Valores de datos
Exactitud de Datos en rango	Inherente	$X=A/B$	A= número de elementos de datos que tiene valores incluidos correctamente en un rango específico. B= número de elementos de datos en total.	Valores de datos

Tabla 4. Resultado de valoración de las métricas de calidad para el dato “Nota Examen Final”

Requisitos de Calidad de Datos	Características de Calidad de Datos	Medidas de Calidad de Datos	Valor de X
Compleitud de Valores de Datos	Compleitud	Compleitud de Datos	$X = 181031/181052$ $X = 0,999$
Exactitud de valores de datos en un rango	Exactitud	Exactitud de Datos en rango	$X = 162290 / 181052$ $X = 0,89$

#### Tarea 4.3.2. Aplicar los criterios de decisión para las métricas

En esta tarea se aplicaron los criterios de decisión definidos en 4.2.2, cuyo resultado puede visualizarse en la Tabla 5.

Tabla 5. Resultados de aplicación de criterios de decisión para métricas de calidad para el dato “Nota Examen Final”

Medidas de Calidad de Datos	Resultados de Aplicación Criterio de Decisión
Compleitud de Datos	Aceptable
Exactitud de Datos en rango	Inaceptable

## 5. 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 “Compleitud 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

Abate, Marcey L & Diegert, Kathleen V. (1998). A Hierarchical Approach to Improving Data Quality. [http:// www.dataquality.com](http://www.dataquality.com) Abate, Marcey L & Diegert, Kathleen V. (1998). A Hierarchical Approach to Improving Data Quality. <http://www.dataquality.com>

ISO/IEC 25000:2014 - Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Guide to SQuaRE

ISO/IEC 25012:2008 - Software Engineering - Software Quality Requirements and Evaluation (SQuaRE) – Data Quality Model.

ISO/IEC 25024:2015 - Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Measurement of data quality.

ISO/IEC 25030:2007- Software engineering— Software product Quality Requirements and Evaluation (SQuaRE)— Quality requirements.

ISO/IEC 25040:2011- Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Evaluation process.

Gil – Aluja, J. (2000). Las decisiones y la incertidumbre. Barcelona

Javed, B. & Hussain, S. (2003). Data quality – A problem and an approach. Wipro Technologies.

Levy, S. (2004). Model Documents and forms for Organizing and Maintaining a Data Quality Program. [www.dataqualitymodeldocument.com](http://www.dataqualitymodeldocument.com).

Loshin, David. (2001). Integration and the Data Quality Imperative: The Data Quality Monitor. <http://www.datajunction.com>

Maynard, J. (1982). Dictionary of Data Processing. Londres, Inglaterra.

Naveh, E. & Halevy, A. (2000). A hierarchical framework for a quality information system. Total Quality Management, Vol. 11, No. 1, p 87-111.

Norma Cubana ISO 9000:2000: Sistemas de gestión de la calidad. Fundamentos y vocabulario.

Olson, J. (2002). Data Profiling: The Data Quality Analyst’s Best Tool. DM Direct, December. DMReview.com.

Pipino, Leo; Lee, Yang & Wang, Richard (2002). Data Quality Assessment. <http://web.mit.edu/tdqm/www/tdqmpub/PipinoLeeWangCACMApr02.pdf>

Redman, Thomas C. (2001). Sistemas de calidad de datos de segunda generación. Manual de Calidad de Juran. McGraw Hill

Tayi, G. & Ballou, D. (1998). Examining Data Quality. Communications of the ACM. Vol. 41, No. 2.