

Requisitos de Calidad de Datos: Un aporte en el ámbito del desarrollo de sistemas de información

Data Quality Requirements: A contribution to the Information Systems Development

Claudio Carrizo

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

Javier Saldarini

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

Angélica Caro

Facultad de Ciencias Empresariales - Universidad del Bio Bio - Chillán – Chile
mcaro@ubiobio.cl

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

En la actualidad los Sistemas de Información (SI) cumplen un rol fundamental en las organizaciones, ya que permiten gestionar los procesos de negocio y brindar información para la toma de decisiones. Para que la información sea de utilidad es necesario que tenga un alto nivel de calidad, por lo tanto es crucial que también los datos del SI sean de calidad. En este sentido, resulta importante considerar la incorporación de aspectos de calidad de datos de manera temprana en el desarrollo de los SI, específicamente en fase de requisitos. El propósito de este trabajo consiste en obtener requisitos de calidad de datos, a través del estudio y análisis de las Normas ISO/IEC 205012 e ISO/IEC 25024, los cuales serán incorporados en historias de usuario. El principal aporte de este trabajo consiste en introducir requisitos de calidad relacionados a los datos en el ámbito del desarrollo de los SI.

Palabras clave: requisitos de calidad de datos, ISO 25000, desarrollo de sistemas de información

Abstract

At present, Information Systems (IS) play a fundamental role in organizations, since they allow managing business processes and providing information for decision-making. For the information to be useful, it must have a high level of quality, therefore it is crucial that the IS data also be of quality. In this sense, it is important to consider the incorporation of data quality aspects early in the development of the IS, specifically in the requirements phase. The purpose of this work is to obtain data quality requirements, through the study and analysis of the ISO / IEC 205012 and ISO / IEC 25024 Standards, to be incorporated into user stories. The main contribution of this work is to introduce quality requirements related to data in the field of IS development.

Keywords: data quality requirements, ISO 25000, information systems development

1. Introducción

Hoy en día, las organizaciones en todos sus niveles necesitan contar con un SI que permita a los usuarios poder llevar adelante los procesos de negocio y también que permita contar con información oportuna y correcta para la toma de decisiones (Zina Houhamdi et al., 2019). Si bien existen muchas definiciones de SI, (Andreu et al., 1991) lo definen como “aquella agrupación formal de los procesos los cuales se ejecutan sobre una colección de datos e información estructurada según los requerimientos de las compañías, donde se encarga de recopilar, elaborar y distribuir de forma selectiva la información que ha sido requerida para la ejecución de la organización en mención, y también para los dinamismos de dirección y control que corresponda, que contribuyen en cierta parte sobre la secuencia de toma de decisiones requeridos con el fin de lograr cumplir con las funciones de la organización manteniendo su misión”. En este sentido, resulta lógico pensar que la información que permiten brindar estos sistemas, debe tener un alto nivel de calidad, para que los usuarios puedan cumplir con las metas propuestas en las organizaciones. Para lograr esto último, es necesario que la colección de datos que utilizan los procesos, también tenga un alto nivel de calidad.

Si bien existen varios autores que abordan la calidad de los datos desde diferentes perspectivas (Wang & Strong et al., 1996) (L. Pipino et al., 2002), existe una serie de estándares de calidad cuya denominación es ISO/IEC 25000 (ISO/IEC 25000, 2014), o también conocida en la industria como SQUARE (Systems and software engineering-Systems and software Quality Requirements and Evaluation), la cual contiene 2 normas que aplican sobre la calidad de los datos. La Norma ISO/IEC 25012 (ISO/IEC 25012, 2008), permite definir un modelo general con características de calidad de datos, y la Norma ISO/IEC 25024 (ISO/IEC 25024, 2015) ofrece las medidas o métricas para cada una de las características del modelo de calidad de datos.

Una manera de poder garantizar la calidad de la información, es incluyendo Requisitos de Calidad de Datos (RCD) de manera temprana, en el desarrollo de los SI. Si bien se han encontrado trabajos que consideran la inclusión de aspectos de calidad de datos en etapas tempranas de los procesos de desarrollo de software (A. Rodríguez et al., 2012) (A. Rodríguez et al., 2013) (C. Guerra-García et al., 2012) (C. Guerra-García et al., 2016), ninguna de ellas plantea la obtención de requisitos de calidad de datos, a través de la serie SQUARE.

El propósito de este trabajo consiste en proponer un conjunto de requisitos de calidad que serán aplicados a datos especificados en historias de usuario. Estos RCD serán obtenidos a través de un proceso de definición y análisis

de las Normas ISO/IEC 25012 e ISO/IEC 25024. El principal aporte de este trabajo consiste es contribuir hacia la incorporación de requisitos de calidad relacionados con los datos en formato estructurado, en el ámbito del desarrollo de los SI.

En lo sucesivo, el presente trabajo se divide en las siguientes secciones. En la Sección 2 se realiza un estudio de la serie ISO/IEC 25000. En la Sección 3 se desarrolla el caso de estudio. En la Sección 4 se presentan los resultados obtenidos en el caso de estudio. Finalmente, en la Sección 5 se elaboran las conclusiones.

2. Estudio de la Serie ISO/IEC 25000 (SQUARE)

Como se mencionó anteriormente, para la obtención de los requisitos de calidad de datos, se necesitó previamente realizar un estudio en general de la serie ISO/IEC 25000 (SQUARE), y en particular de las Normas ISO/IEC 25030, ISO/IEC 25012 e ISO/IEC 25024.

2.1. Norma ISO/IEC 25030: Requisitos de Calidad del Software/Datos

En esta Norma se ofrece una serie de recomendaciones para obtener requisitos de calidad para el producto de software/datos, donde se destaca un proceso de definición y análisis que se realiza sobre las características y medidas de calidad. En el caso de requisitos de calidad en relación a los datos, se lleva adelante este proceso, utilizando como referencia las Normas ISO/IEC 25012:2008 e ISO/IEC 25024:2015, las cuales se describen a continuación.

2.2. Norma ISO/IEC 25012:2008. Modelo de Calidad de Datos

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

- **Calidad de Datos Inherentes** (ISO/IEC 25012): 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** (ISO/IEC 25012): 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 Tabla 1 se pueden visualizar las 15 características del modelo de calidad de datos según sus puntos de vista inherente y dependiente del sistema. La existencia del símbolo “X” indica la pertinencia de las características para la calidad de los datos requeridos o evaluados desde puntos de vista inherentes y/o dependientes del sistema.

Características	Calidad de los datos	
	Inherente	Dependiente del Sistema
Exactitud	X	
Completitud	X	
Coherencia	X	
Credibilidad	X	

Actualidad	X	
Accesibilidad	X	X
Conformidad	X	X
Confidencialidad	X	X
Eficiencia	X	X
Precisión	X	X
Trazabilidad	X	X
Comprensibilidad	X	X
Disponibilidad		X
Portabilidad		X
Recuperabilidad		X

Tabla 1. Características del Modelo de Calidad de Datos

Cabe destacar que por cada característica que contiene el modelo de calidad de datos, existe un conjunto de medidas de calidad de datos, provistas por la Norma ISO/IEC 25024, la cual se describe a continuación.

2.3. Norma ISO/IEC 25024:2015. Medidas de Calidad de Datos

Esta Norma Internacional define medidas de calidad de datos para medir cuantitativamente la calidad de los datos en términos de características definidas en la Norma ISO/IEC 25012:2008. Contiene un conjunto básico de medidas de calidad de los datos para cada característica. En la Tabla 2 se muestra un ejemplo de una medida de calidad que brinda esta Norma.

Característica de Calidad de Datos (CCD)	Compleitud
Punto de Vista CCD	Inherente
Medida de Calidad de Datos (MCD)	Compleitud de Valores de Datos
Punto de Vista MCD	Inherente
Descripción de la MCD	Compleitud de valores de datos para un elemento de datos en un archivo de datos
Función de Medición	$X = A / B$ A = número de valores de datos para un elemento de datos en un archivo de datos conectado a los valores esperados B = número de valores de datos esperado para un elemento de datos en un archivo de datos

Tabla 2. Ejemplo de la medida de calidad de Datos “Compleitud de Valores de Datos”.

3. Caso de Estudio: Sistema Web E-Commerce

A través de este caso de estudio se pretende demostrar la inclusión de RCD en una especificación de requisitos funcionales en formato de historias de usuario. Cabe aclarar, que dichos RCD se obtienen a partir del estudio y análisis de las características y medidas de calidad de datos, proporcionadas por la Normas ISO/IEC 25012 e ISO/IEC 25024.

El presente caso de estudio se llevó a cabo en el marco del Proyecto “Casa Bardo E-Commerce”, el cual fue desarrollado por una empresa de desarrollo de software local, donde se utilizó un ciclo de vida iterativo e incremental, a través del uso de la Metodología Ágil Scrum. El objetivo principal de este proyecto consistió en desarrollar un sistema web de venta online de artículos para el hogar, enfocado para clientes mayoristas. Por cuestiones de espacio en este artículo, sólo aplicaremos el caso de estudio sobre una funcionalidad del sistema, la cual será el “Registro de Mayorista”. En la Tabla 3 se puede observar un extracto de la especificación del requisito para esta funcionalidad, en formato de historia de usuario.

Id HU	01
Nombre	Registro de Mayorista
Descripción	COMO mayorista, QUIERO registrarme como mayorista PARA poder comprar productos
Detalle	- Los datos de registro de usuario son: Razón Social, Nombre de Fantasía, Cuil, calle, numero, localidad, código postal, provincia, teléfono, email, fecha de registro, contraseña (todos requeridos) - El formulario de registro debe ser de tipo responsivo.
Criterios de Aceptación	- Validar el ingreso obligatorio de todos los datos - Solicitar que ingreso de contraseña y confirmar contraseña - Validar que la contraseña esté encriptada - Validar formato correcto para el email

Tabla 3. Historia de Usuario “Registro de Mayorista”

Una vez definida la historia de usuario, el siguiente paso consistió en obtener los requisitos de calidad para cada dato especificado. Para ello se llevó a cabo en primera instancia una revisión de las Normas ISO/IEC 25012 e ISO/IEC 25024, lo que permitió luego realizar el análisis correspondiente, para definir las características y medidas de calidad que más aplicaban para cada dato. Uno de los RCD obtenidos fue el de “Complejidad de Valores de Datos”, debido a que en la historia de usuario se especificó que todos los datos debían ser requeridos u obligatorios. Para obtener como resultado este RCD, se realizó el análisis que se detalla en la Tabla 4.

Característica de Calidad de Datos (ISO/IEC 25012)	Medida de Calidad de Datos (ISO/IEC 25024)	Detalle de Análisis Realizado	¿Aplica como RCD?
Complejidad	Complejidad de Registros	La medida aplica para registros de datos y no para un dato en particular	No
	Complejidad de Archivos de Datos	La medida aplica para archivos de datos y no para un dato en particular	No
	Complejidad de Valores de Datos	La medida aplica para valores que tomen los datos	Sí
	Complejidad de Modelo Conceptual de Datos	La medida aplica a entidades del modelo conceptual y no para un dato en particular	No
	Complejidad para Metadatos	La medida aplica para metadatos y no para un dato en particular	No

Tabla 4. Análisis de Características y Medidas de Calidad de Datos

4. Resultados del Caso de Estudio

Como resultado del caso de estudio desarrollado, se analizaron 15 características y 63 medidas de calidad de datos, para obtener un conjunto de requisitos de calidad, los cuales se aplican a cada dato. Por cuestiones de espacio, en la Tabla 5 se muestran algunos de los RCD definidos para los datos “Razón Social”, “Cuil” y “Fecha de Registro”.

Dato	Requisitos de Calidad de Datos	Característica de Calidad de Datos (ISO/IEC 25012)	Medida de Calidad de Datos (ISO/IEC 25024)
Razón Social	Exactitud Sintáctica de valores de datos	Exactitud	Exactitud Sintáctica de datos
	Exactitud semántica de valores de datos	Exactitud	Exactitud Semántica de datos
	Completitud de valores de datos	Completitud	Completitud de Valores de Datos
Cuil	Credibilidad de valores de datos	Credibilidad	Credibilidad de Valores
	Completitud de valores de datos	Completitud	Completitud de Valores de Datos
	Eficiencia de formato de datos	Eficiencia	Eficiencia de formato de datos
Fecha Registro	Consistencia semántica de valores de datos	Consistencia	Consistencia semántica
	Completitud de valores de datos	Completitud	Completitud de Valores de Datos
	Eficiencia de formato de datos	Eficiencia	Eficiencia de formato de datos

Tabla 5. Resultado de Análisis de Características y Medidas de Calidad de Datos

Por último, se debe vincular a la historia de usuario “Registro de Mayorista” el resultado obtenido en la Tabla 5, para de esta manera poder incluir los RCD, junto con los requisitos funcionales y no funcionales especificados.

5. Conclusiones

La calidad de la información es considerada un aspecto fundamental para que los usuarios de las organizaciones puedan llevar adelante los procesos de negocios y la correcta toma de decisiones. En este trabajo se plantea la necesidad de incorporar aspectos de calidad de datos en fases de requisitos del desarrollo de los sistemas de información. Para lograr esto último, se hizo uso de la serie de estándares de calidad ISO/IEC 25000, en donde tomando como referencia las Normas ISO/IEC 25012 e ISO/IEC 25024, se logró llevar a cabo un proceso de análisis y definición de un conjunto de Requisitos de Calidad de Datos, los cuales pueden ser considerados para ser incluidos en el formato de historias de usuarios, junto con los requisitos funcionales y no funcionales.

Referencias

- A. Rodríguez, A. Caro, C. Cappiello, I. Caballero, “A BPMN Extension for Including Data Quality Requirements in Business Process Modeling”. International Workshop on Business Process Modeling Notation. Lecture Notes in Business Information Processing, vol.125, pp. 116-125. 2012.
- A. Rodríguez y A. Caro, “Especificación de Requisitos de Calidad de Datos en Modelos de Procesos de Negocio: BPiCD una propuesta”, Jornadas Chilenas de Computación, II Workshop on Business Process Management (WBPM). Temuco, Chile. 2013.

- Andreu, R., Ricart J. E. Y Valor, J. (1991): *Estrategia y Sistemas de Información*. Mc Graw-Hill, Madrid.
- C. Guerra-García, I. Caballero, R. Testillano, "Elicitation of Specific Requirements of Data Quality during the Web Portal Development", In *Proceedings of the 10th International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems and 1st International Workshop on Web Intelligence (WEBI-2012)*, pages 81-93, 2012.
- C. Guerra-García, I. Caballero, M. Cardenas-Juarez, R. Samano-Robles, "Adding quality in the user requirements specification: A first approach", 2016 Eleventh International Conference on Digital Information Management (ICDIM). IEEE.
- ISO/IEC 25000 Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models.
- ISO/IEC 25030:2007. Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Quality requirements.
- ISO/IEC 25012:2008. Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE). Data quality model.
- ISO/IEC 25012. Normas ISO 25000 Calidad de Software y Datos. <https://iso25000.com/index.php/normas-iso-25000/iso-25012>
- ISO/IEC 25024:2015. Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE). Measurement of data quality.
- Leo L. Pipino, Y.W.L. and Richard Y. Wang, "Data Quality Assessment, in *COMMUNICATIONS OF THE ACM*", 2002, p. 211 – 218.
- Wang & Strong, D.M. (1996), "Beyond accuracy: What data quality means to data consumers". *Journal of management information systems*, 5-33.
- Zina Houhamdi, Belkacem Athamena. Impacts of information quality on decision-making. *Global Business and Economics Review*, Vol. 21, No. 1, 2019.