

Hacia un Modelo de Evaluación de Calidad para Datos Basado en la Norma ISO/IEC 25012

Carrizo Claudio*, Saldarini Javier*, Salgado Carlos+, Sanchez Alberto+, Peralta Mario+

*Grupo de I+D Calidad de Software - Facultad Regional San Francisco
Universidad Tecnológica Nacional

Av. de la Universidad 501 - San Francisco - Córdoba - Tel. 03564-421147
{ cjcarrizo77, saldarinijavier }@gmail.com

+ Departamento de Informática Facultad de Ciencias Físico-Matemáticas y
Naturales Universidad Nacional de San Luis

Ejército de los Andes 950 – C.P. 5700 – San Luis – Argentina
e-mail: {csalgado, mperalta, alfanego}@unsl.edu.ar

RESUMEN

Los datos representan el activo corporativo más importante, ya que a través del procesamiento de los mismos, permiten generar información de gran valor para las empresas en lo que respecta a las actividades de gestión, operación y toma de decisiones.

Uno de los problemas principales al que se enfrentan las empresas en general está relacionado con la falta de calidad de los datos que gestionan. Esto tiene un impacto directo en el desempeño de la empresa debido a que, al no contar con datos fiables, se cometen errores al tomar decisiones, se pierde competitividad y se daña la imagen corporativa. Por este motivo, se considera un factor clave poder garantizar la calidad de los datos.

El propósito de este trabajo consiste en

por características y sub-características de calidad de datos, las cuales deberán ser evaluadas a través de la aplicación de métricas e indicadores para un caso en concreto en el dominio de las empresas.

Con el modelo de datos propuesto se pretende brindar un aporte a la investigación científica y a las empresas

que deseen evaluar la calidad de sus datos en post de mejorar la gestión de los mismos.

Palabras clave: Modelo de Calidad – ISO 25012 – Métricas – Indicadores – Calidad de Datos.

CONTEXTO

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto N° P-031516. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se ha logrado un importante vínculo con el nivel nacional e se encuentra reconocido por el programa de Incentivos.

1. INTRODUCCIÓN

Los datos se han convertido actualmente en uno de los activos más

valiosos para las empresas debido a que estos permiten generar información de gran valor para la gestión, operación y toma de decisiones. La calidad de la información es un factor clave para cualquier organización debido a que el acierto de las decisiones depende de la calidad de los datos en los que se basan dichas decisiones [1].

La falta de calidad de datos es uno de los principales problemas a los que se enfrentan los responsables de sistemas de información y las empresas en general, pues representa claramente uno de los problemas "ocultos" más graves y persistentes en cualquier organización. [2]. Una pobre calidad de datos tiene un gran impacto en el desempeño de la empresa, genera importantes sobrecostos, daña la imagen corporativa y trae consecuencias como las que se detallan a continuación:

- Errores en la toma de decisiones.
- Pérdida de competitividad.
- Impacto en la retención de clientes.
- Ineficiencia operacional
- Costos no planificados

De acuerdo a lo expuesto anteriormente, y entendiendo que la gestión de datos constituye un recurso estratégico para las empresas, se hace imprescindible garantizar la calidad de los datos. Disponer de datos con alto nivel de calidad no sólo es importante para las empresas en lo que se refiere a una adecuada gestión y toma de decisiones eficaces, sino también para lograr fortalecer la relación con el cliente de manera de poder fidelizarlo y lograr la satisfacción del mismo.

La calidad de los datos es un concepto multidimensional [4] y frecuentemente es

definida como "datos apropiados para el uso" [3, 5]. Esto quiere decir que el usuario es quién determina si un conjunto de datos, usados en una determinada tarea y en un contexto específico, pueden ser usados para el objetivo previsto. Según la norma ISO/IEC 25012 [6], 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". Lo expresado anteriormente, determina la relevancia que tiene la participación del usuario a la hora de definir si un conjunto de datos es de calidad. En consecuencia, se puede concluir que el término "calidad de datos" es un concepto relativo ya que, al variar las expectativas de los usuarios respecto a los mismos, varían también las características que éstos deben tener para ser considerados adecuados. Esto significa que, la calidad de los datos está asociada a un conjunto de características que permiten, por un lado, establecer un lenguaje común y, por otro lado, focalizar los problemas de calidad de los datos y las oportunidades de mejora.

Recientemente la Organización de Estandarización Internacional (ISO) en conjunto con la Comisión Electrotécnica Internacional (IEC), han presentado la familia de normas ISO/IEC 25000 (SQuaRE - System and Software Quality Requirements and Evaluation) [7], la cual proporciona una guía para establecer los requisitos y evaluación de calidad de productos de software. Esta familia está compuesta por 5 divisiones, una de ellas es la ISO/IEC 2501n que especifica los requisitos para un modelo de calidad. A su vez, dentro de esta división se encuentra la norma ISO/IEC 25012 que proporciona un modelo de calidad de datos en donde se establecen las características de calidad de datos que se deben tener en cuenta a la hora de evaluar

las propiedades de un producto de datos determinado. La Calidad del Producto de Datos se puede entender como el grado en que los datos satisfacen los requisitos definidos por la organización a la que pertenece el producto. Son precisamente estos requisitos los que se encuentran reflejados en el modelo de calidad de datos mediante 15 características, las cuales están clasificadas en 2 grandes categorías: Calidad de Datos Inherente: y Calidad de Datos Dependientes del Sistema [8].

El objetivo de este trabajo consiste en elaborar un modelo basado en la Norma ISO/IEC 25012 que estará compuesto por características y sub-características de calidad de datos, las cuales deberán ser evaluadas a través de la aplicación de métricas e indicadores en el dominio de las empresas a través de un caso en concreto. Para lograr esto será necesario:

- Identificar y seleccionar las características y sub-características de calidad de datos basadas en el Estándar ISO/IEC 25012 para la elaboración del modelo.
- Definir y/o adaptar métricas e indicadores para cada una de las características y sub-características del modelo de calidad de datos elaborado basándose en el estándar ISO/IEC 25024 [9].
- Evaluar el modelo a través de la aplicación de las métricas e indicadores en un caso en concreto en el dominio de las empresas.

El modelo de calidad de datos resultante será un instrumento de suma utilidad para las empresas ya que permitirá evaluar la calidad de los datos que manejan y

determinar las mejoras necesarias para asegurar que los datos almacenados tengan la calidad deseada.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Una de las áreas prioritarias establecidas por la especialidad Sistemas de Información en la política de desarrollo de Ciencia y Tecnología de la Facultad Regional San Francisco es la de "Calidad de Software". Dentro de este contexto, la línea de investigación en la cual se está trabajando actualmente está relacionada con la temática "Calidad de datos".

Los principales ejes de trabajo de esta línea de investigación se detallan a continuación:

- Elaboración de un modelo compuesto por características y sub-características de calidad de datos para el dominio de las empresas.
- Elaboración de métricas e indicadores para cada una de las características y sub-características del modelo.
- Evaluación del modelo en un caso en concreto en el dominio de las empresas.

3. RESULTADOS OBTENIDOS/ESPERADOS

Dentro de los resultados esperados para esta línea de investigación, se han logrado alcanzar algunos al momento de esta publicación. La propuesta consiste en poder obtener un modelo que permita evaluar a través de características y sub-características la calidad de los datos que son gestionados por las empresas. Para poder lograrla se han fijado los siguientes objetivos:

- Se han estudiado, identificado y seleccionado las características y sub-características de calidad de datos basadas en el Estándar ISO/IEC 25012.
 - Se ha elaborado un modelo de calidad de datos general con las mejores características y sub-características seleccionadas. Y complementada con características obtenidas de la experiencia y en la de expertos del medio.
 - Se están analizando distintas métricas de la bibliografía para poder utilizarlas, redefinirlas y/o adaptarlas al dominio que se estudiará. Para ello también se hace uso del estándar ISO/IEC 25024.
 - Evaluar el modelo a través de la aplicación de las métricas e indicadores definidos en un caso de estudio en concreto.
- Difundir y divulgar los resultados obtenidos

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de investigación se trabaja en lo referente a distintos métodos de evaluación de calidad con la participación del Grupo de Investigación "Calidad de Software" perteneciente a la UTN Facultad Regional San Francisco, Córdoba.

Se está trabajando en el Proyecto de Tesis de la Maestría en Calidad de Software (Plan Ord. 017/09-CD) del Ing. Claudio Carrizo, de la Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Como así también se están llevando a cabo algunas tesinas

de grado para la Licenciatura en Ciencias de la Computación y trabajos finales en el marco de la Ingeniería en Informática.

5. BIBLIOGRAFÍA

- [1] Calidad de Datos - ISO/IEC 25012. Recuperado el 07/09/16 <http://www.aqclab.es/index.php/calidad-datos-iso-25012>
- [2] Introducción a la Calidad de Datos. Recuperado el 15/09/16. <http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/368784/introduccion-a-la-calidad-de-datos-definicion-control-y-beneficios>
- [3] D. Ballou and R. Wang. "Modeling information manufacturing systems to determine information product quality". *Management Science*. Vol. 44, pp. 462-484. 1998.
- [4] B.D. Klein. "User perceptions of data quality: Internet and traditional text sources". *Journal of Computer Information Systems*. Vol. 41, pp. 9-18. 2001.
- [5] D. Strong, Y. Lee and R. Wang. "Data Quality in Context". *Communications of the ACM*. Vol. 40, Issue 5, pp. 103-110. May, 1997.
- [6] ISO/IEC-25012. "ISO/IEC 25012: Software Engineering - Software Quality Requirements and Evaluation (SQuaRE) - Data Quality Model". 2008

- [7] La familia de normas ISO/IEC 25000. Recuperado el 05/10/2016. <http://iso25000.com/index.php/normas-iso-25000>
- [8] ISO/IEC 25012. Recuperado el 15/10/2016.
- [9] <http://iso25000.com/index.php/normas-iso-25000/iso-25012>
- [10] Norma ISO/IEC 25024:2015. Recuperado el 09/11/2016. <https://www.iso.org/standard/35749.html>