

HACIA LA DEFINICIÓN TEMPRANA DE REQUISITOS DE CALIDAD DE DATOS PARA PRODUCTOS DE INFORMACIÓN EN CONTEXTO ÁGIL

Carrizo Claudio*, Javier Saldarini*, Silvana Armando*, Julio Trasmontana*, Angélica Caro#, Carlos Salgado+, Albero Sanchez+, Mario Peralta+

*Facultad Regional San Francisco – Universidad Tecnológica Nacional
Av. de la Universidad 501 - San Francisco - Córdoba - Tel. 03564-421147
{jcarrizo77, saldarinijavier, silvana.armando, juliotrasmontana}@gmail.com

#Departamento de Ciencias de la Computación y Tecnologías de la Información - Facultad de Ciencias Empresariales, Universidad del Bío-Bío
Casilla 447, 3780000, Chillán, Chile
{mcaro}@ubiobio.cl

+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
{csalgado, alfanego, mperalta}@unsl.edu.ar

RESUMEN

Los Sistemas de Información (SI) permiten procesar datos en bruto para obtener como resultado Productos de Información (PIs), por lo cual es fundamental que dichos datos tengan un alto nivel de calidad, para finalmente poder obtener como resultado productos de información de calidad, que permitan apoyar los procesos de negocio y la toma de decisiones. En relación a los procesos de desarrollo de SI, resulta importante considerar la incorporación de aspectos asociados con la calidad de los datos en fases tempranas, específicamente en fase de requisitos. En procesos de desarrollo donde se utiliza la metodología ágil SCRUM, comúnmente no se definen requisitos que estén relacionados con la calidad de los datos, y enfocados hacia los productos de información. El propósito de esta línea de investigación es proponer un método para la definición y especificación temprana de Requisitos de Calidad de Datos, basado en el concepto de PIs y en la Familia de estándares ISO/IEC 25000. El uso de dicho método está orientado en el contexto de la

Metodología Ágil SCRUM y pretende ser un aporte en pos de garantizar la calidad de datos de los productos de información, cuyo valor es directamente transferible a los usuarios de los SI.

Palabras clave: Requisitos de Calidad de Datos – Productos de Información - Procesos de Desarrollo Ágil – SCRUM

CONTEXTO

El presente trabajo se enmarca dentro de los siguientes proyectos de Investigación:

- Ingeniería de Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto N° P-03-2020. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se ha logrado un importante vínculo con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el programa de Incentivos.

- Gestión de Proyectos de Software: Los Modelos de Calidad como Soporte a los Procesos y Productos Software. Este proyecto fue financiado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional, y homologado bajo el código SIPPBSF0008185.

1. INTRODUCCIÓN

En la actualidad las organizaciones gestionan sus procesos de negocio a través de una variedad de SI, los cuales no sólo permiten llevar adelante el desarrollo, seguimiento y control de sus procesos y actividades, sino que también permiten brindar información de valor para tomar decisiones oportunas. El propósito principal de todo SI es brindar información a los usuarios dentro de una organización, partiendo del procesamiento de datos de entrada en bruto, para obtener finalmente como salida “Productos de Información”. Según (R. Wang, 1998), un Producto de Información puede definirse como el resultado del procesamiento de un conjunto de datos, el cual tiene un valor que es directamente transferible a los usuarios. Por ejemplo, una factura es un PI y corresponde al resultado que se obtiene del procesamiento de las compras mensuales realizadas por un cliente (A. Caro et. al, 2013).

Dada la importancia de poder obtener productos de información de los SI que tengan el nivel de calidad deseado (Y. Álvarez, 2015), que no sólo apoyen los procesos de negocio de la organización, sino también la toma de decisiones (Z. Houhamdi, B. Athamena, 2019), es necesario garantizar la calidad de los datos, los cuales al ser procesados se transformarán en información de calidad. Según ISO/IEC 25012:2008, la calidad de datos puede definirse como el “Grado en

que las características de los datos satisfacen necesidades implícitas y establecidas cuando son usados en condiciones específicas”.

Si bien existen autores que abordan la calidad de los datos desde diferentes perspectivas (Wang & Strong, 1996; Pipino et. al, 2002), existe la Familia de Normas ISO/IEC 25000, también conocida como SQUARE (Systems and software engineering-Systems and software Quality Requirements and Evaluation), la cual propone un marco para la definición de requerimientos de la calidad de software/datos y evaluación de la calidad del software/datos, apoyados por un proceso de medición de la calidad de software/datos. El uso de esta serie de normas está enfocado hacia adquirentes, desarrolladores, entre otros. En lo que respecta a la definición de requisitos de calidad, SQUARE provee la Norma ISO/IEC 25030:2019, la cual ofrece un marco que permite guiar la definición de Requisitos de Calidad de Datos, por medio de un proceso de definición y análisis, en el que se utilizan respectivamente la Norma ISO/IEC 25012:2008, que permite definir un modelo general con características de calidad de datos, y la Norma ISO/IEC 25024:2015, que ofrece las medidas o métricas para cada una de las características del modelo de calidad de datos.

Según lo antes mencionado, para poder obtener productos de información de calidad, resulta importante considerar la incorporación de aspectos relacionados con la calidad de los datos en fases tempranas del proceso de desarrollo de los SI, más específicamente en fase de desarrollo de requisitos. Si bien existen trabajos que consideran la inclusión de aspectos de calidad de datos en etapas tempranas del proceso de desarrollo de software (A. Caro et. al 2013; A.

Rodríguez y A. Caro, 2013; A. Rodríguez, 2012; C. Guerra-García, 2012; Y. Alonso, 2020), se han encontrado pocas propuestas que aborden la definición de requisitos de calidad de datos en fase de requisitos en un proceso de desarrollo ágil.

Lo descrito anteriormente, nos motivó a iniciar una línea de investigación que permita realizar un aporte desde la calidad de los datos, hacia los productos de información resultantes de los SI, los cuales se desarrollen en el contexto de metodologías ágiles. Aunque actualmente existen varias de ellas en el mercado (Scrum, Extreme Programming, Kanban), se optó por la metodología ágil Scrum, ya que es la más reconocida y utilizada a nivel mundial, y la misma consiste en un conjunto de prácticas y roles que permiten el trabajo de entregas incrementales de un producto. Se basa en una estructura de desarrollo incremental, esto es, cualquier ciclo de desarrollo del producto y/o servicio se desgrana en pequeños proyectos divididos en distintas etapas: análisis, desarrollo y testing. En la etapa de desarrollo se llevan adelante interacciones del proceso, es decir, entregas regulares y parciales del producto final.

De acuerdo a lo expuesto anteriormente, uno de los objetivos principales de la línea de investigación consistió en desarrollar un método basado en SQUARE y aplicado en el contexto del desarrollo de software utilizando la metodología ágil SCRUM, el cual le pueda servir de soporte al rol del Dueño del Producto (DP), para guiar la definición y especificación temprana de Requisitos de Calidad de Datos para Productos de Información.

Para la definición de Requisitos de Calidad de Datos se toma como punto de partida el concepto de Productos de Información, que están asociados a

Requisitos Funcionales especificados a través de Historias de Usuario. Para obtener una especificación de requisitos de calidad de datos para los productos de información, se propone un formato de documento portátil (pdf), el cual fue elegido por su alta compatibilidad.

El método desarrollado está compuesto por 4 actividades principales:

Actividad 1. Definición de Perfiles de Usuarios: El DP debe definir los distintos perfiles de usuarios que participarán en las historias de usuario definidas.

Actividad 2. Definición de Historias de Usuarios: El DP debe definir las historias de usuarios que se consideran importantes o relevantes y que van a estar relacionadas a los productos de información. Cada historia de usuario está asociada con un perfil de usuario (definido en la actividad anterior) y con un rol con respecto a la calidad de datos (consumidor de datos / productor de datos).

Actividad 3. Definición de Productos de Información: el DP debe definir cada uno de los productos de información, los cuales deben tener relación con las historias de usuario definidas. Por cada producto de información se define el nombre, una descripción, las historias de usuarios relacionadas y los atributos que son propios del producto de información.

Actividad 4. Definición de Requisitos de Calidad de Datos para Productos de Información: el DP deberá identificar y definir los requisitos de calidad de datos que son pertinentes para cada producto de información o atributos del producto de información. Los requisitos de calidad de datos se obtienen a partir del uso de las Normas ISO/IEC 25030, ISO/IEC 25012 e ISO/IEC 25024.

Otro de los objetivos propuestos y que está contemplado dentro de la línea de investigación, es el desarrollo de una herramienta, que será implementada en un

entorno web, y que permitirá facilitar el uso y aplicación del método, ofreciendo al DP un asistente que permita llevar adelante cada una de las actividades que propone el método, hasta lograr la definición y especificación de requisitos de calidad de datos para los productos de información. Cabe acotar, que por cada producto de información definido, la herramienta generará una especificación de requisitos de calidad de datos en un documento en formato pdf, el cual luego podrá ser adjuntado a cada una de las historias de usuario definidas.

La principal contribución de esta línea de investigación es ofrecer desde el ámbito de la calidad de datos, un método sistematizado y automatizado, que está basado en SQUARE, y que permite definir y especificar requisitos de calidad de datos, en pos de garantizar la calidad de los productos de información en contexto de procesos de desarrollos ágiles.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

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

- Revisión Sistemática de la Literatura asociada a la temática calidad de datos en general y en contexto de procesos de desarrollo de software en particular.
- Estudio en profundidad de las Normas ISO/IEC 25030, ISO/IEC 25012 e ISO/IEC 25024.
- Estudio en profundidad de las metodologías ágiles existentes en el mercado.

- Elaboración de un método que permita definir en forma temprana requisitos de calidad de datos para productos de información.
- Desarrollo de una herramienta que permita automatizar el método para facilitar su uso y aplicación.

3. RESULTADOS OBTENIDOS/ESPERADOS

Resultados Obtenidos

- Informe con resultado de la revisión sistemática de la literatura sobre la temática calidad de datos y asociada a procesos de desarrollo ágiles.
- Informe con resultado del proceso de revisión y estudio de las normas ISO/IEC 25030, ISO/IEC 25012 e ISO/IEC 25024.
- Informe con resultado de la revisión y estudio de las metodologías ágiles utilizadas actualmente en el mercado.
- Método para definir requisitos de calidad de datos para productos de información en el contexto de desarrollo ágil.

Resultados Esperados

- Desarrollo de una herramienta que permita automatizar el método propuesto.

4. FORMACIÓN DE RECURSOS HUMANOS

Esta línea de investigación se trabaja desde el Grupo de Investigación "Calidad de Software y Datos", que pertenece a la UTN Facultad Regional San Francisco, Córdoba.

El equipo de trabajo está constituido por

8 docentes investigadores (un Director de Proyecto, un Co-Director y seis docentes investigadores), todos de la especialidad Sistemas de Información. A su vez, se está trabajando en el Proyecto de Tesis de la Maestría en Calidad de Software (Plan Ord. 017/09-CD) de la Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis, de los Ing. Claudio Carrizo y Javier Saldarini.

5. BIBLIOGRAFÍA

- A. Caro, A. Fuentes, A. Soto. “Desarrollando sistemas de información centrados en la calidad de datos”, Revista Chilena de Ingeniería, vol.21, pp. 54-69, 2013.
- 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.
- 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.
- 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.
- Extreme Programming. Disponible en <https://www.agilealliance.org/glossary/xp>
- ISO/IEC 25000. Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models.
- ISO/IEC 25030:2019. Systems and software engineering — Systems and software quality requirements and evaluation (SQuaRE). Quality requirements framework.
- ISO/IEC 25012:2008. Software engineering -- Software product 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 25012:2008. Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE). Data quality model.
- Kanbanize. Qué es Kanban: Definición, Características y Ventajas. Disponible en <https://kanbanize.com/es/recursos-de-kanban/primeros-pasos/que-es-kanban>.
- La Metodología ágil más usada: SCRUM. Disponible en <https://www.nextu.com/blog/que-es-scrum/>.
- Leo L. Pipino, Y.W.L. and Richard Y. Wang, “Data Quality Assessment, in COMMUNICATIONS OF THE ACM”, 2002, p. 211 – 218.
- R. Wang, “A Product Perspective on Total data Quality Management”. Communications of the ACM. Vol. 41, Issue 2, pp. 58-65. 1998.
- ScrumAlliance. Disponible en <https://www.scrumalliance.org/>
- Y. Fernanda Torres Alonso, “Especificación de requerimientos de software con un enfoque de calidad de datos”, Universidad Autónoma de San Luis, Potosí, 2020.
- Y. Álvarez, “Análisis de la calidad de datos en fuentes de la suite ABCD”, Trabajo de Diplomatura. Universidad Central “Marta Abreu” de las Villas, Facultad de Matemática, Física y Computación, Carrera de Licenciatura en Ciencias de la Computación, Santa Clara, Cuba, 2015.

Wang & Strong, D.M. (1996), "Beyond accuracy: What data quality means to data consumers". *Journal of management information systems*, 5-33.

Z. Houhamdi, B. Athamena, "Impacts of information quality on decision-making". *Global Business and Economics Review*, Vol. 21, No. 1, 2019.