

Desarrollando sistemas de información centrados en la calidad de datos

Developing information systems focused on data quality

Angélica Caro¹ Alejandra Fuentes¹ M. Antonieta Soto¹

Recibido 27 de diciembre de 2012, aceptado 14 de enero de 2013

Received: December 27, 2012 Accepted: January 14, 2013

RESUMEN

Los Sistemas de Información se pueden comparar con los sistemas de fabricación de productos en el sentido que estos últimos generan productos físicos a partir de materia prima, y los primeros generan productos de información a partir de datos puros. Esta analogía enfatiza la idea de que los productos, ya sean físicos de o información, tienen un valor que es transferible al consumidor. Si se considera que un Sistema de Información tiene como propósito entregar Productos de Información de alta calidad, entonces la Calidad de Datos o Calidad de Información es un factor importante de considerar en su producción. Efectivamente, la calidad de los datos suele definirse como “*datos apropiados para el uso*”, es decir, datos que sean de utilidad para los consumidores/usuarios en un contexto de uso específico. Considerando todo lo anterior, esta propuesta de investigación se centra en la idea de que la calidad de los datos debe ser abordada desde que comienza el desarrollo de un Producto de Información, es decir desde cuando se crea el Sistema de Información que lo generará. Así, en este artículo se presenta un método que guía a un equipo de desarrollo a construir Sistemas de Información centrándose en la Calidad de Datos. Concretamente se describe la primera versión de este método, orientada a la etapa de especificación de requisitos de un Sistema de Información con Calidad de Datos. Junto con esto, se presenta un caso de estudio en el que se muestra cómo se ha aplicado el método en el desarrollo de un Sistema de Información.

Palabras clave: Calidad de datos, calidad de información, producto de información, desarrollo de software, método.

ABSTRACT

Information systems can be compared with the manufacturing systems in the sense that the latter produce physical products from raw materials and the first generate information products from raw data. This analogy emphasizes the idea that products, whether physical or information, have a value that is transferable to the consumer. If we see an Information System as a system whose purpose is to deliver products of high quality information, then the Data Quality or Information Quality is an important factor to consider in its production. Indeed, the data quality is often defined as “data suitable for use”, i.e., data that are useful for consumers/users in a specified context of use. Considering the above, this research focuses on the idea that the quality of data must be addressed from the beginning of the development of an information product, i.e., from when creating the information system that will generate. Thus, this paper presents a method that leads to a development team to build information systems focusing on data quality. Specifically, the first version of this method, aimed at the requirements specification stage of an Information System with Data Quality. A case study showing how the method has been applied in the development of an information system is presented.

Keywords: Data quality, information quality, information product, software development, method.

1 Departamento de Ciencias de la Computación y Tecnologías de la Información. Universidad del Bío-Bío. Casilla 447. Chillán, Chile. E-mail: mcaro@ubiobio.cl; alejandra.fuentes.lagos@gmail.com; msoto@ubiobio.cl

INTRODUCCIÓN

Los Sistemas de Información (en adelante se abreviará como SI) existen en todo tipo de organizaciones y son fundamentales para que éstas puedan obtener, procesar, almacenar y gestionar su información [1]. Se puede decir entonces, que el propósito de los SI es proveer datos e información a quien la necesite dentro de la organización a quien sirve. Asimismo, un SI se puede ver como un sistema de procesamiento de datos que actúa sobre los datos puros para producir Productos de Información (en adelante se abreviará como PI) [2]. Un PI entonces es el resultado del procesamiento de un conjunto de datos. Por ejemplo, una factura es un PI y corresponde al resultado del procesamiento de las compras mensuales realizadas por un cliente.

Por otro lado, en los últimos 50 años la industria de desarrollo de software ha evolucionado mucho respecto de las técnicas y procesos que aseguren que los SI desarrollados poseen la calidad requerida por quienes los usarán. Sin embargo, a pesar de ello, las organizaciones y los usuarios de los SI dentro de ellas, no siempre están satisfechos con la información y datos que éstos les proveen [3].

Efectivamente, las organizaciones actuales están mejor preparadas que nunca para desarrollar SI que utilizan datos procedentes de una variedad de fuentes, pero desafortunadamente la mayoría de las bases de datos no están libres de errores [4]. Asimismo, es sabido que los problemas de datos pueden hacer que el rendimiento de un SI no sea el esperado.

Por otro lado, la calidad de los datos o calidad de información (suelen usarse indistintamente aunque sean conceptos diferentes) es un concepto multidimensional [5] y frecuentemente es definida como “datos apropiados para el uso” [4, 6]. Esto quiere decir, que el usuario es quien determina si un conjunto de datos, usados en una determinada tarea y en un contexto específico, pueden ser usados para el objetivo previsto. La norma ISO/IEC 25012 [7] define calidad de datos como “Grado en que las características de los datos satisfacen necesidades implícitas y establecidas cuando son usados en condiciones específicas”. Lo anterior, da un papel relevante a la participación del usuario a la hora de definir si un conjunto de datos es de calidad.

Recogiendo lo que afirma Pressman [8], que la garantía de calidad del software es una actividad de protección que se debe aplicar a lo largo de todo el proceso de desarrollo, y Wang [2], sobre incorporar aspectos de Calidad de Datos (DQ por las siglas en inglés de Data Quality) en las fases tempranas del desarrollo, el interés de esta investigación está centrado en proveer un método que complemente el desarrollo de software y lo centre en la DQ.

Este artículo presenta la primera versión del método llamado DeWIK (por las siglas en inglés de **Development With Information Quality**), para ser usado en la fase de especificación de requisitos. Algunos elementos básicos del método son: el concepto de PI y la norma ISO/IEC 25012. Asimismo, DeWIK recoge los principios de la metodología TDQM (Total Data Quality Methodology), propuesta en [2], cuyo objetivo es la mejora continua de los PIs en el contexto organizacional.

El resto de este artículo se organiza de la siguiente manera. La siguiente sección presenta los trabajos relacionados. En la tercera sección se describen los elementos básicos del método. Una descripción detallada del método DeWIK es presentada en la cuarta sección. La quinta sección describe un caso de estudio en el cual DeWIK fue aplicado. Finalmente, en la sexta sección se muestran las conclusiones y el trabajo futuro.

TRABAJOS RELACIONADOS

Diversas propuestas de la bibliografía abordan la calidad de los datos en un Sistema de Información. Algunas de ellas asumen la existencia de un proceso de generación de datos o de información, el cual entrega PIs que deben ser evaluados y mejorados como parte de un proceso permanente dentro de una organización [4, 9-12]. Sin embargo, pocas propuestas están centradas en cómo abordar la DQ desde dentro del proceso de generación de los PIs. Efectivamente, si se considera que un PI es un producto que se obtiene desde un Sistema de Información [2], resulta lógico pensar que durante el desarrollo de éste ya se debería prestar atención a la calidad de los PIs que se generarán.

Es decir, durante el proceso de desarrollo de un SI se deberían definir las características de DQ relevantes para cada uno de los PIs involucrados,

las estrategias de medición y evaluación que se aplicarán una vez que estén en circulación dentro la organización (o incluso entre organizaciones).

Entre las propuestas que apuntan en este sentido se pueden mencionar algunas centradas en las primeras etapas del desarrollo de un SI, cuando se realizan diversas actividades tendientes a definir los requisitos funcionales y no funcionales de la aplicación a desarrollar [2, 13-20]. Concretamente en trabajos como [2, 14-16] se proponen estrategias mediante las cuales se pueden identificar los objetos o PIs esenciales en una organización a un nivel conceptual, se definen las necesidades de DQ y las métricas de DQ que pueden ser aplicadas para conseguir el nivel de DQ deseado. Sin embargo, todas ellas abordan la DQ a nivel de toda la organización y no para un SI específico. Desafortunadamente, ninguna guía a un equipo de desarrollo en cómo centrar la creación de un SI en la DQ.

Respecto del modelado de los datos, trabajos como [13, 17, 18] proponen enriquecer semánticamente los modelos conceptuales introduciendo requisitos de DQ y estableciendo restricciones semánticas de modo que se controle la existencia de información correcta.

Por último, también se pueden mencionar propuestas como [19, 20], que extienden y/o crean perfiles UML de manera que se generan artefactos orientados a los aspectos de DQ que deben estar presentes en una aplicación. De estos trabajos hay que comentar que sus propuestas no están incluidas como parte de un método de desarrollo de software que oriente a un equipo de desarrollo respecto de cómo integrarlas durante todo el proceso de construcción de un SI, y por otro lado, no se orientan al PI.

ELEMENTOS BÁSICOS DE DEWIQ

Los tres elementos básicos del método DeWIQ son: el concepto de PI, la norma de calidad de datos ISO/IEC 25012, y la metodología TDQM. A continuación se describe brevemente cada uno de ellos.

Producto de Información

El concepto de PI nace de la analogía entre la fabricación de productos y la producción de información hecha por Wang en [2]. En el primer caso se tiene un sistema de fabricación que actúa

sobre materia prima para generar un producto tangible. De forma equivalente, en el segundo caso, se tiene un sistema de procesamiento de datos (SI) que actúa sobre datos puros para producir PIs.

En DeWIQ los PIs son los elementos básicos a identificar en un SI, viendo a un SI como un sistema cuyo propósito es entregar PIs de alta calidad para los usuarios finales o consumidores de información. En la Tabla 1 se muestran algunos ejemplos de PIs en un SI.

Tabla 1. Ejemplos de PIs.

Sistemas de Información	Productos de Información
Sistema de Ventas de Artículos	Factura, Boleta, Guía de Despacho, Registro de Existencia, Informe de Cuentas Corrientes de Clientes, etc.
Sistema de Administración de Empleados	Liquidación de Sueldo, Comprobante de Día Feriado, Informe de Comisión de Ventas, etc.
Sistema de Devoluciones	Notas de Crédito, Informe de Reposición de Artículos a Bodega, Informe de Motivos de Devolución, Listado de Productos Mermados, etc.

En un sistema de producción de información, un PI podría ser el insumo para la generación de un nuevo PI, por ejemplo un conjunto de facturas (cada una de ellas un PI) podrían ser el insumo para la generación de un informe de ventas por cliente (que a su vez es un PI). Estos insumos (PIs) pueden provenir de un mismo SI o bien de otros. Esto implica que un PI no sólo es la salida de un SI, sino que también puede ser una entrada. Y por tanto, un PI puede ser también el resultado de haber procesado varios PIs anteriores en otros SI.

El concepto de PI es introducido para enfatizar el hecho que éste tiene un valor que es transferible al usuario final o consumidor de datos. En un sistema de producción de información es posible identificar 3 roles [6], que en el método DeWIQ se han denominado roles de DQ:

- Consumidores de datos, quienes usan los datos en un contexto específico.
- Productores de datos, quienes generan los datos en un contexto específico.
- Custodios de los datos, quienes proveen y gestionan los sistemas y recursos computacionales para almacenar y procesar datos.

Esta primera versión de DeWIK está centrada en los roles de consumidor y productor de datos, pues son los que están más ligados a la especificación de requisitos en un SI. El otro rol deberá ser considerado en las siguientes etapas de desarrollo.

Norma ISO/IEC 25012

Esta norma presenta un modelo genérico de calidad de datos [7]. Plantea que la gestión y mejora de los datos es importante para abordar situaciones como:

- Adquisición de datos en organizaciones donde la calidad del proceso de producción de datos es desconocido o débil.
- Existencia de datos defectuosos que contribuyen a generar información insuficiente, que provoca resultados inutilizables y clientes insatisfechos.
- Dispersión de datos entre varios propietarios y usuarios. Lo que puede implicar la falta de una visión coherente e integrada, necesaria para garantizar la interoperabilidad y la cooperación.
- La coexistencia de sistemas heredados con sistemas modernos.
- SI donde los datos cambian con frecuencia y su integración con otros datos es relevante (por ejemplo, SI en la Web)

Teniendo en cuenta estas situaciones, y dado que el ciclo de vida de los datos es a menudo más largo que el ciclo de vida del software, el modelo de DQ propuesto por la ISO pretende responder a estas necesidades contribuyendo a:

- Definir y evaluar los requisitos de DQ en procesos de adquisición, producción e integración de los datos.
- Identificar los criterios de garantía de DQ, útiles también para la re-ingeniería, evaluación y mejora de los datos.
- Evaluar la conformidad de los datos con la legislación y/o requisitos.

El modelo identifica quince (15) características, que se agrupan desde dos puntos de vista [7], inherente y dependiente del sistema (véase la Tabla 2).

El *punto de vista inherente*, que se refiere al grado en el cual las características de DQ tienen el potencial intrínseco para satisfacer las necesidades implícitas y explícitas cuando los datos son usados en condiciones específicas. Es decir, la DQ está referida al grado en que los valores de los datos

pertencen a un dominio específico, cumplen con ciertas restricciones (por ejemplo, reglas del negocio), y respetan ciertas relaciones de los datos (por ejemplo, consistencia), entre otros.

El *punto de vista dependiente del sistema*, que se refiere al grado en el cual la DQ es enriquecida y preservada dentro de un sistema computacional cuando es usado bajo condiciones específicas. Es decir, la DQ depende del dominio tecnológico en que los datos son usados. Esto es, depende de las capacidades de los componentes de los sistemas computacionales como: hardware (por ejemplo, para proveer el acceso adecuado a los datos), software de sistemas (por ejemplo, para el respaldo de los datos y proveer la Recuperabilidad de los mismos), y otros tipos de software (por ejemplo, herramientas de migración para proveer Portabilidad).

Tabla 2. Modelo de DQ ISO/IEC 25012 [7].

Características	Puntos de vista de la DQ	
	Inherente	Dependiente del Sistema
Exactitud	X	
Completitud	X	
Consistencia	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

Como se ha destacado en la Tabla 2 (en negrita y con sombreado), algunas de las características de DQ deben ser abordadas desde ambos puntos de vista. A continuación en la Tabla 3 se presentan las definiciones que da la norma ISO/IEC 25012 a algunas de las características de DQ.

Es importante aclarar que en nuestro trabajo usamos como términos equivalentes “característica de DQ”, usado en la norma ISO/IEC 25012, y “dimensión de DQ”, usado ampliamente en área de DQ.

Tabla 3. Algunas definiciones de las características en la norma ISO/IEC 25012.

Característica	Descripción
Exactitud	El grado en el cual el dato tiene atributos que representan el valor correcto de un concepto o evento en un contexto específico de uso.
Completitud	El grado en el cual el dato asociado a una entidad tiene valores para todos los atributos esperados e instancias de entidad relacionadas, de acuerdo a un contexto específico de uso.
Consistencia	El grado en el cual el dato tiene atributos libres de contradicción y son coherentes con otros datos en un contexto específico de uso.
Credibilidad	El grado en el cual el dato tiene atributos considerados como verdaderos y creíbles por usuarios en un contexto específico de uso.
Actualidad	El grado en el cual el dato tiene los atributos que son del período correcto en un contexto específico de uso.
Accesibilidad.	El grado en el cual se puede acceder al dato en un contexto específico de uso.
Conformidad	El grado en el cual el dato tiene atributos que se adhieren a normas, convenciones o regulaciones vigentes y reglas relacionadas con la calidad de datos en un contexto específico de uso.
Confidencialidad	El grado en el cual el dato tiene los atributos que aseguran que éste es sólo accesible e interpretable por usuarios autorizados en un contexto específico de uso.
Eficiencia	El grado en el cual el dato tiene los atributos que pueden ser procesados y proporciona los niveles esperados de funcionamiento (desempeño) usando las cantidades y los tipos de recursos apropiados en un contexto específico de uso.
Precisión	El grado en el cual el dato tiene atributos que son exactos o que proporcionan su discriminación en un contexto específico de uso.
Trazabilidad	El grado en el cual el dato tiene atributos que proporcionan un rastro de auditoría de acceso a los datos y de cualquier cambio hecho a los datos en un contexto específico de uso.
Comprensibilidad	El grado en el cual el dato tiene atributos que le permiten ser leído e interpretado por usuarios, y es expresado en lenguajes apropiados, símbolos y unidades en un contexto específico de uso.

Disponibilidad	El grado en el cual el dato tiene atributos que le permiten ser recuperados por usuarios autorizados y/o aplicaciones en un contexto específico de uso.
Portabilidad	El grado en el cual los datos tienen atributos que les permiten ser instalados, substituidos o movidos de un sistema a otro conservando la calidad existente, en un contexto específico de uso.
Recuperabilidad	El grado en el cual el dato tiene atributos que le permiten mantener y conservar un nivel especificado de operaciones y calidad, aún en caso de falla, en un contexto específico de uso.

Enfoque TDQM

La metodología TDQM [2] tiene el propósito de entregar PIs de alta calidad a los consumidores de datos, facilitando la aplicación de políticas de calidad de datos globales en una organización a nivel de gestión y de alta dirección. Para lograr este propósito, TDQM propone un ciclo de mejora continua de los PIs, compuesto por 4 etapas: definir, medir, analizar y mejorar, como se muestra en la Figura 1.

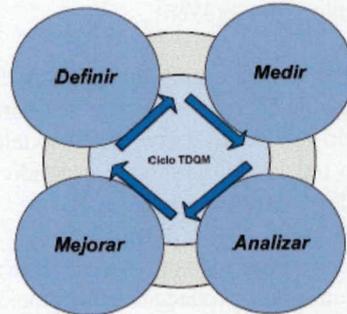


Figura 1. Etapas de TDQM [2].

Para la aplicación de TDQM, una organización debe desarrollar algunas actividades previas: (a) Definir los PI en términos del negocio, (b) Establecer un equipo de DQ, (c) Entrenar al equipo para la evaluación de la DQ y gestión de la DQ en los PIs y (d) Institucionalizar la mejora continua de los PIs.

Las etapas que contempla la metodología TDQM son las siguientes:

- **Definición de PIs:** Esta identifica adecuadamente los PIs que se usan en la organización. Como resultado, genera: (a) un modelo de calidad entidad-relación que define el PI y

sus requisitos de DQ, y (b) un sistema de producción de información que describe como el PI es producido, y las interacciones entre los proveedores de información, fabricantes, consumidores, y de gestión.

- **Medición de PIs:** Identifica y aplica las métricas que mejor describen cuantitativamente los problemas que se tienen con los niveles de calidad de los datos.
- **Análisis de PIs:** Permite identificar las causas que originan los problemas de DQ y estima el impacto de los datos de mala calidad.
- **Mejora de PIs:** Define técnicas para mejorar la calidad de los datos, identificando las áreas clave para las mejoras y alineamiento de los flujos de información con necesidades de los procesos del negocio.

Es fundamental para poder aplicar esta metodología, la premisa de que las organizaciones deben tratar la información como un producto que se mueve a través de un sistema de fabricación de información, muy similar a un producto físico, pero con las ventajas y limitaciones propias de la naturaleza de un PI.

TDQM ha demostrado ser eficaz para mejorar PIs, sobre todo cuando hay un fuerte compromiso desde el nivel gerencial expresado en su política organizacional.

Conjugando Elementos

En DeWIQ se usa el concepto de PI como el centro de su actividad, la norma ISO/IEC 25012 le proporciona el conjunto de características de DQ que pueden ser asociadas a los PIs y el enfoque TDQM sugiere la necesidad de que los PIs deben estar en un proceso de mejora continua, aunque en este trabajo esto se aplica en el contexto del desarrollo de un SI y no de manera global dentro de una organización.

DeWIQ: UN MÉTODO PARA EL DESARROLLO DE SISTEMAS DE INFORMACION CENTRADO EN LA CALIDAD DE DATOS

El objetivo de DeWIQ es guiar el desarrollo de software centrándose en la Calidad de los Datos. Puede ser aplicado por cualquier equipo de desarrollo de forma complementaria con el proceso de desarrollo de software que se utilice. En ésta

su primera versión, el método está orientado a la etapa de especificación de los requisitos de un SI, véase la Figura 2.

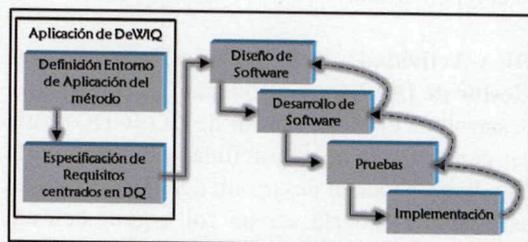


Figura 2. DeWIQ y el proceso genérico de desarrollo de software.

DeWIQ es iterativo y se recomienda aplicarlo mientras existan requisitos por definir y sea necesario establecer sus implicancias en cuanto a la DQ de los PIs involucrados con ellos. La Figura 3 presenta las distintas actividades que lo componen.

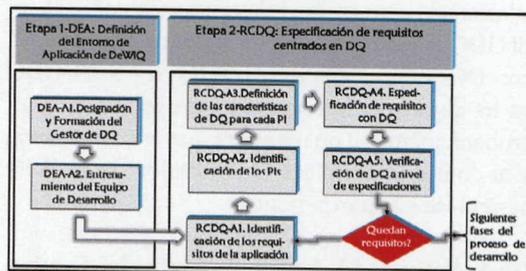


Figura 3. Etapas y actividades del Método DeWIQ.

Como muestra la Figura 3, DeWIQ consta de dos etapas. La primera etapa, compuesta por dos actividades, define el entorno de aplicación del método. La segunda etapa, compuesta de 5 actividades que se desarrollan de manera iterativa, permite obtener la especificación de requisitos del SI incluyendo requisitos de DQ. A continuación se describen las etapas y actividades de DeWIQ.

Etapla 1. DEA. Definición del Entorno de Aplicación de DeWIQ

Una vez decidido el desarrollo de un SI, se deben definir algunos aspectos necesarios para la aplicación del método. Esto, considerando que para la aplicación correcta de DeWIQ se deben definir los roles de (a) Gestor de Calidad de Datos (Responsable de la aplicación de DeWIQ), (b) Integrantes del Equipo de Desarrollo (por ejemplo, ingeniero de requisitos, analista, arquitecto, programador, etc.) y (c) Actor del SI (individuo que tiene un perfil específico

dentro de la organización y en el SI a desarrollar, es el futuro usuario del SI que está vinculado con alguno de los PI del SI). Las dos actividades que la componen son:

DEA-Actividad1. Designación y Formación del Gestor de DQ. Se debe designar a la persona que desarrollará el rol de Gestor de DQ (GDQ). Este rol concentra la responsabilidad de gestionar la DQ durante todo el desarrollo del SI. Idealmente, este gestor debería ser un rol organizacional preocupado permanentemente de la DQ en los SI existentes y por desarrollar. Además, este rol tendrá la responsabilidad de incorporar políticas de DQ, estrategias de formación y mejoramiento continuo en los SI implementados. Sin embargo, si el GDQ es designado por primera vez, se deben tener en cuenta las siguientes consideraciones: (a) El GDQ puede ser un integrante del Equipo de Desarrollo (ED) o no. En ambos casos, deberá trabajar estrechamente relacionado con el resto de los integrantes del ED, (b) El GDQ deberá instruirse en aspectos relacionados con DQ y en el método DeWIK, (c) Si el GDQ ya ha desarrollado este rol en proyectos previos, probablemente ya no requerirá mayor capacitación y al contrario mejorará la aplicación de DeWIK basándose en su experiencia.

DEA-Actividad2. Entrenamiento del Equipo de Desarrollo. El GDQ debe instruir al ED sobre DQ y DeWIK, de modo que pueda ser aplicado correctamente. Esta instrucción abarcará principalmente: (a) Definición de conceptos básicos sobre DQ y DeWIK (entre otros: PI, DQ, características de DQ, roles asociados a la DQ), (b) Modelo de DQ a aplicar (en principio la norma ISO/IEC 25012), (c) Asignación de roles de DQ a los actores del SI (se reclasificarán los roles de actores del SI en los roles clásicos de DQ como son: consumidor, custodio y productor de datos), y (d) Instrucción sobre DeWIK (funcionamiento, implementación y aplicación del método durante el proceso de desarrollo de un SI).

Etapas 2. (RCDQ). Especificación de requisitos centrados en DQ

Esta etapa tiene por objetivo capturar y definir los requisitos del SI, incorporando aspectos de DQ a considerar en el desarrollo. En esta etapa, las actividades que la componen, y que son descritas a continuación, se desarrollan de forma iterativa,

hasta conseguir la especificación completa de los requisitos del SI con DQ.

RCDQ-Actividad1. Identificación de los requisitos de la aplicación. Esta actividad consiste en la captura de requisitos del SI. Aquí el ED debe usar sus procedimientos habituales para interactuar con los usuarios, obtener y documentar los requisitos, objetivos y funcionalidades que ellos requieren respecto de la aplicación a desarrollar. El artefacto que captura cada uno de los requisitos se usará para especificar también los requisitos de DQ. Si bien DeWIK considera esta actividad como parte de su etapa 2, en realidad proporciona actividades complementarias a ésta, que es la que el ED debe desarrollar de la manera habitual. Las siguientes actividades son las que permitirán al ED abordar la DQ para los PI del SI que se pretende desarrollar.

RCDQ-Actividad2. Identificación de PIs. Una vez elicitados los requisitos de la aplicación, o al menos alguna parte de ellos, se deben identificar los PIs involucrados asociados a ellos. Para esto, y usando el entrenamiento previo, el ED, guiado por el GDQ, deberá analizar cada requisito en busca de PIs. Como estrategia de identificación de PIs el método propone usar listas de control, donde por cada uno de los requisitos se responden interrogantes que ayudan a determinar qué PIs tiene asociados. Se debe tener presente que un mismo PI puede estar involucrado con más de un requisito (por ejemplo, en uno puede ser generado y en otro poder ser usado y/o modificado). Una vez definidos los PIs, deben ser documentados en una plantilla como la que se muestra en la Tabla 4.

Tabla 4. Artefacto de Descripción de PIs.

Id PI	Nombre PI	Propósito	Descripción	Requisito	Actores del SI
PI-1					
PI-2					
...					
PI-n					

Este artefacto se registra para cada PI: un nombre que lo identifica de manera única, propósito u objetivo del PI, descripción del PI (datos que lo conforman) y los actores del SI relacionados con el PI. Además, este artefacto puede ser complementado con información adicional que sirva de documentación del desarrollo,

por ejemplo, fecha de actualización, número de versión, miembro del ED que lo actualizó, etc. Dado que DeWIK es iterativo, este artefacto y los demás pueden ser completados en cada iteración.

A continuación, se debe crear una matriz (véase Tabla 5), que refleja la relación entre los requisitos del SI y los PIs identificados.

Tabla 5. Artefacto Matriz Requisitos y PIs.

Requisitos	PI-1	PI-2	...	PI-n
Id Req-1	X			
Id Req-2	X	X		X
...	X	X		
Id Req-n		X		X

Este artefacto muestra la interacción que tienen los requisitos del SI con los PIs definidos. Esto es importante para: (a) determinar qué PIs pueden ser afectados al modificar un requisito, (b) definir qué PIs son críticos, cuando se relacionan con muchos requisitos y (c) definir requisitos críticos, correspondientes a aquellos que cruzan con mayor frecuencia a los PIs.

Una vez que los PIs estén asociados a los requisitos, se deberán asociar con los roles de DQ, para esto se debe usar el artefacto como el que se muestra en la Tabla 6.

Tabla 6. Artefacto de Interacción de PIs con actores del SI y roles de DQ.

Id PI	Actores del SI	Rol de DQ	
		Consumidor de Datos	Productor de Datos
PI-1	Usuario I del SI	X	
	Usuario II del SI		X
	...		
	Usuario n del SI		X
PI-2	Usuario I del SI	X	
	Usuario II del SI	X	
	...		
	Usuario n del SI		X
...			
PI-n	Usuario I del SI	X	X
	Usuario II del SI		
	...		
	Usuario n del SI	X	

Este artefacto (Tabla 6) muestra la relación entre actores del SI y los roles de DQ por cada PI. Así, a través de él, se puede visualizar: (a) Qué actores del SI son los que interactúan más frecuentemente con un PI y su rol de DQ (la participación de estos roles en el proceso de desarrollo e implementación será relevante), (b) Qué rol de usuario de DQ es el más utilizado (aportarán en función de su rol las necesidades de DQ para cada PI), (c) Qué PIs tienen mayor interacción con un actor del SI en función de su rol de usuario DQ (sobre ellos habrá que poner especial cuidado a la hora de definir las características de DQ que deben tener).

RCDQ-Actividad3. Definición de las características de DQ para cada PI.

En esta actividad se deben elaborar encuestas para obtener la importancia que los actores del SI asignan a cada característica de DQ, del modelo de DQ ISO/IEC 25012, para cada PI identificado en el SI. Los actores del SI, en su función de rol de DQ (consumidor o productor de datos), deben indicar en la encuesta cuánta importancia debe tener cada una de las características de DQ en cada PI analizado. Esta asignación de valores debe estar guiada de acuerdo a interrogantes lo suficientemente claras como para asegurar la comprensión del significado y ámbito de cada característica de DQ en los PIs consultados. Se ha definido una encuesta tipo que el ED puede usar como referencia para crear la suya, pero por razones de espacio no es incluida en este artículo.

Si la cantidad de PIs identificados en el SI es muy grande, y por escasez de tiempo y disponibilidad de recursos no se desea trabajar con todos ellos, el ED, GDQ y los actores del SI pueden seleccionar aquellos PIs más críticos y relevantes siguiendo algún criterio específico, como por ejemplo: aquellos PIs que están relacionados con muchos requisitos. Los artefactos mostrados en las Tablas 4, 5 y 6 ayudarán a determinar sobre qué PIs trabajar (sobre qué PIs encuestar a los actores del SI) y con qué actores del SI y roles de DQ. El siguiente paso es consolidar la información contenida en las encuestas y generar el artefacto mostrado en la Tabla 7.

Por cada PI, este artefacto registra la importancia promedio que cada rol de DQ da a cada característica de DQ. Este artefacto permite visualizar claramente cuáles características de DQ son más relevantes en cada PI y en qué grado. Cuando por cada rol de

Tabla 7. Artefacto con Características de DQ y su importancia para los Roles de DQ en el SI.

Id PI			
Característica de DQ	Consumidor de Datos	Productor de datos	Ponderación Promedio
Exactitud			
Compleitud			
Consistencia			
Credibilidad			
Actualidad			
Accesibilidad			
Conformidad			
Confidencialidad			
Eficiencia			
Precisión			
Trazabilidad			
Comprensibilidad			
Disponibilidad			
Portabilidad			
Recuperabilidad			

DQ se encuesta a más de un actor del SI se debe indicar el puntaje promedio de las valoraciones que ellos dieron por cada característica de DQ (columnas Consumidor de Datos y Productor de Datos). La tercera columna representa el valor de la ponderación promedio de cada característica de DQ para ambos roles de DQ, que se calcula usando la siguiente fórmula:

$$\text{Ponderación Promedio (Características DQ)} = \frac{(2CoD + PrD)}{3}$$

Donde, *CoD* corresponde al valor promedio de puntuación, dado por el rol de DQ “consumidor de datos” a la característica de DQ analizada. *PrD* corresponde al valor promedio de puntuación, dado por el rol de DQ “productor de datos” a la característica de DQ analizada. Como se observa en la fórmula, se da una mayor ponderación al rol de de DQ “consumidor de datos”, esto debido a que será quien más use los PIs generados por el sistema y por tanto puede determinar con más asertividad cuales son las características de DQ deseables para los PIs.

La idea de fondo de esta actividad es que luego se puedan determinar las características realmente relevantes para cada PI (no necesariamente las 15 propuestas por el modelo), con la idea de tomar

acciones desde el punto de vista del desarrollo sólo para ellas y no incrementar excesivamente el costo de un proyecto, por centrar el desarrollo en la DQ. En este sentido es importante señalar que un PI no necesariamente tiene que cumplir con todas las dimensiones de DQ. Esto porque, por un lado, el costo en tiempo y recursos puede ser considerado excesivo, y por otro lado, por la relación inversa que presentan algunas dimensiones (por ejemplo, Accesibilidad versus Confidencialidad, o Trazabilidad versus Eficiencia).

RCDQ-Actividad4. Especificación de requisitos incluyendo DQ. Cuando ya se conocen las características de DQ que cada PI debe tener, entonces se debe complementar la especificación de los requisitos del SI. Esto significa básicamente: (a) indicar dentro de la especificación el o los PIs involucrados y las características de DQ que éstos deben poseer, y (b) definir por cada PI acciones de desarrollo concretas que se deben asociar a la especificación de requisitos para poder asegurar que el PI alcanzará las características deseadas.

Para lo primero, es necesario analizar los valores de ponderación promedio calculados y elegir las características que tuvieron mayor puntuación de acuerdo a un puntaje de corte previamente establecido. Este puntaje de corte debe quedar definido en función de la escala de valores dada en la encuesta, y debe ser determinado en conjunto por el ED, el GDQ y el usuario. Así en la especificación de un requisito deberá agregar la identificación de cada PI involucrado en él y las características de DQ que éste debe tener.

En segundo lugar, se debe generar un artefacto con una estructura similar a la mostrada en la Tabla 8, por cada PI involucrado con el requisito.

Tabla 8. Artefacto que define acciones para lograr las características de DQ en los PIs del SI.

Id PI	Características de DQ	Valoración (sobre el puntaje de corte)	Acciones
PI-1			
PI-2			
...			
PI-n			

Así, por cada PI se tendrá un grupo de características de DQ relevantes y por cada una de ellas se especificarán algunas acciones de desarrollo tendientes a alcanzarlas. El artefacto debe ser anexo a la especificación de requisitos.

DeWIK provee un conjunto inicial de acciones a implementar en un SI para lograr cada característica de DQ. Las acciones se generaron a partir del desarrollo de un panel de expertos, formado por un grupo de 8 personas con experiencia en el desarrollo de software. De acuerdo con la definición de cada característica de DQ, se le pidió a cada uno acciones que se pudieran implementar para lograr cada característica de DQ en los datos de un SI. Se discutieron entre todos las acciones propuestas, y se seleccionó un grupo concreto de ellas para cada característica de DQ. En la Tabla 9 se muestran algunos ejemplos de posibles acciones.

Tabla 9. Ejemplo de acciones para lograr características de DQ.

Característica de DQ	Acciones Generales
Exactitud	<ul style="list-style-type: none"> Definir reglas de validación Acotar dominio de los datos de acuerdo a su tipo. Vista Previa de Información de ingreso.
Compleitud	<ul style="list-style-type: none"> Establecer niveles de ingreso. Por ejemplo: obligatorios, deseable, entre otros. Indicar aviso de datos que no cumplen con ciertas normas u otras restricciones de carácter legal o de tipo de procedimiento administrativo. No guardar un formulario hasta asegurar que esté completo.
Consistencia	<ul style="list-style-type: none"> Establecer reglas de consistencia propias del negocio o de información histórica. Validaciones de rango de acuerdo a reglas del negocio.
Credibilidad	<ul style="list-style-type: none"> Ingresar responsable o autor del dato. Ingresar fecha de ingreso o última actualización del dato.
Accesibilidad	<ul style="list-style-type: none"> Definir tipos de usuarios que determinan la accesibilidad a los datos. Por ejemplo: para personas minusválidas, no videntes, entre otros. PI tengan capacidad de ser leídos con acciones de "zoom".

Como resultado de las actividades desarrolladas hasta aquí, por cada requisito se debe obtener un documento con su especificación, indicando los PIs

relacionados y las características de DQ definidas para él. Adicionalmente a esto, se debe incluir el artefacto descrito en la Tabla 8, con las acciones para lograr dichas características de DQ por PI.

RCDQ-Actividad5. Verificación de DQ a nivel de especificaciones. En esta última actividad del ciclo se deben comprobar las especificaciones de requisitos. Básicamente esta evaluación apunta a determinar su completitud en los siguientes aspectos relacionados con la DQ:

- Cada especificación de requisitos debe contener una referencia a cada uno de los PIs asociados con el requisito.
- Por cada PI en la especificación, se deben indicar las características de DQ relevantes que éste debe poseer.
- Por cada PI, la especificación debe contener un anexo en el cual se indiquen acciones de desarrollo para el logro de las características de DQ en el PI.
- Todos los PIs del SI (o al menos los seleccionados) deben estar asociados al menos a una especificación de requisitos.

Como instrumento de evaluación se ha definido una lista de control, en la que el ED debe cotejar lo indicado anteriormente. En la Tabla 10 se muestran algunas de las interrogantes que esta lista contiene, las acciones asociadas a ellas no se incluyen, por razones de espacio.

Tabla 10. Interrogantes para verificar la DQ.

Interrogante
¿La especificación del requisito incluye PI's?
¿En la especificación del requisito se han definido las características de DQ asociadas al PI?
¿El PI se ha identificado con su información en el artefacto "Identificación de PI's"?
En el artefacto "Matriz de Requisitos de aplicación y sus PI's" ¿está el PI asociado a su requisito?
En el artefacto "Interacción de PI's con Actores del SI y DQ" ¿está el PI asociado a su Actores del SI y su rol de DQ?
¿Se aplicó la encuesta de características de DQ para el PI?
En el artefacto de "Características de DQ y su importancia en roles de usuario DQ", en el PI, ¿se han identificado características de DQ necesarias para su implementación?
¿Se han definido las acciones para el PI en el artefacto de listado de acciones en PI's?

Como resultado de la verificación de las especificaciones de requisitos y PIs del SI, las especificaciones serán clasificadas en uno de los siguientes dos niveles de madurez:

- Completa, lo que significa que la especificación del requisito está completamente definida, desde el punto de vista de la DQ, para pasar con él a la siguiente etapa de desarrollo de software.
- Por Completar, lo que significa que la especificación del requisito aún tiene pendiente alguna actividad, desde el punto de vista de la DQ, por lo cual debería volver a entrar en el ciclo de DeWIQ.

En las siguientes iteraciones de DeWIQ el ED debe repetir las distintas actividades de método, partiendo de la documentación generada en la última iteración y tratando de completar cada una de ellas de manera definitiva.

Una vez que las especificaciones de requisitos y PIs del SI hayan alcanzado el nivel de completitud esperado, el ED se encuentra en condiciones de avanzar a las siguientes fases del desarrollo de la aplicación. En dichas fases los requisitos de DQ, inmersos en los requisitos de la aplicación, deberán ser modelados, implementados, verificados y validados.

Finalmente, cada uno de los PIs identificados en el SI debe quedar debidamente documentado, ya que pasará a constituir parte del conjunto de PIs de la organización y que más adelante puede ser requerido como parte de un proceso de evaluación de la calidad de los datos.

CASO DE ESTUDIO

A continuación, se muestra la aplicación de DeWIQ en el desarrollo de un SI para la Gestión de la información de los egresados de la carrera de Ingeniería Civil de Informática de la Universidad del Bío-Bío. El objetivo es que la jefatura de carrera cuente con un registro actualizado con los datos de sus egresados. Este SI se desarrolló como un proyecto de fin de carrera.

A continuación se describe la aplicación del método en el proyecto y se muestran algunos de los artefactos generados.

Etapa 1. DEA. Definición del Entorno de Aplicación de DeWIQ

Por ser la primera aplicación del método se denominó como GDQ a una de las autoras del mismo. El ED fue constituido por una sola persona, el estudiante que debía desarrollar el proyecto de fin de carrera. Los actores del SI fueron definidos como: “Egresado”, “Jefe de Carrera”, “Docente” y “WebMaster”.

DEA-Actividad1. Como el GDQ designado era una de las autoras del método no fue necesaria su formación.

DEA-Actividad2. El GDQ instruyó al ED respecto de los conceptos básicos de DQ, el modelo de DQ a usar (norma ISO/IEC 25012) y el método DeWIQ. Asimismo, los actores de SI fueron clasificados en los roles de DQ (consumidor y/o productor de datos).

Etapa 2. RCDQ. Especificación de requisitos centrados en DQ

Para el desarrollo del SI el ED usó como metodología de desarrollo el Proceso Unificado. Por tanto utilizó los procedimientos establecidos en dicha metodología para la elicitación de los requisitos. Se usaron Casos de Uso como artefacto para la especificación de los requisitos.

RCDQ- Actividad1. Como resultado de esta actividad se generaron los requisitos del SI. En la Tabla 11 se presenta un listado con algunos de ellos.

Tabla 11. Algunos requisitos de software elicitados en el caso de estudio.

Requisito
CU-01 Crear Cuenta Egresado
CU-02 Editar Perfil Egresado
CU-03 Filtrar Egresados
CU-04 Crear Nuevo Aviso-Noticia
CU-05 Modificar Aviso-Noticia
CU-06 Eliminar Aviso-Noticia
CU-07 Ingresar Nuevo Docente
CU-08 Editar Datos Docente
CU-09 Eliminar Docente
CU-10 Agregar Nueva Carrera

RCDQ-Actividad2. Utilizando como estrategia las listas de control suministradas por el método, se identificaron los PIs asociados a los diferentes

requisitos de software elicitados previamente. En la Tabla 12 se muestran algunos de los PIs identificados en el SI.

Tabla 12. Ejemplo de algunos PIs identificados en el Caso de Estudio.

Id PI	Nombre PI	Propósito	Descripción	Actores del SI	Requisito
F P E R E	Ficha de Perfil Egresado	Ficha que posee los datos de perfil de un Egresado	Datos Personales Datos de Contacto Datos Laborales Datos Académicos Datos de Autenticación	Egresado Web Master Jefe de Carrera	CU-01, CU-02, CU-03
A V/ N O T	Aviso/Noticia	Información de un Aviso/Noticia	Autor Destinatarios Titulo Contenido	Egresado Jefe de Carrera Docente Web Master	CU-04, CU-05, CU-06
F P E R D	Ficha de Perfil Docente	Ficha que posee los datos del perfil Docente	Datos Personales Datos de Contacto Datos Académicos Datos de Autenticación	Egresado Jefe de Carrera Docente Web Master	CU-07, CU-08, CU-09
F I C C A R	Ficha Carrera	Ficha que muestra los datos de una Carrera	Nombre Código Descripción Jefe de Carrera.	Egresado Jefe de Carrera Docente Web Master	CU-10, CU-11

La descripción de los PIs, mostrados en la Tabla 12, es un resumen de ésta. En el caso de estudio esta descripción es mucho más detallada y especifica todos los datos que conformaban cada PI.

Una vez identificados los PIs del SI, se creó la matriz que los relaciona con cada requisito del SI. En la Tabla 13 se muestra dicha matriz para los PIs presentados previamente en la Tabla 12.

Como se puede ver en la Tabla 13, en general, la relación de requisito y PI es simple ya que básicamente cada requisito está asociado a un solo PI. Por otro lado, cada PI no está asociado con muchos requisitos.

Tabla 13. Artefacto Matriz de Relación de Requisitos de la aplicación y sus PIs del Caso de Estudio.

Requisitos	FPERE	AV/NOT	FPERD	FICCAR
CU-01	X			
CU-02	X			
CU-03	X			
CU-04		X		
CU-05		X		
CU-06		X		
CU-07			X	
CU-08			X	
CU-09			X	
CU-10				X
CU-11				X

A continuación, en la Tabla 14 se ingresa la información en el artefacto de Interacción de PIs con los actores del SI y sus Roles de DQ.

Tabla 14. Artefacto de Interacción de PIs con Actores del SI y su rol de DQ del Caso de Estudio.

Id PI	Actores del SI	Rol de DQ	
		Consumidor de Datos	Productor de Datos
FPERE	Egresado	X	X
	WebMaster	X	
	Jefe de Carrera	X	
	Docente	X	
LISTEGR	Jefe de Carrera	X	
	Docente	X	
AV/NOT	Egresado	X	X
	Jefe de Carrera	X	X
	Docente	X	X
	WebMaster	X	X
FPERD	Egresado	X	
	Jefe de Carrera	X	
	Docente	X	X
	WebMaster	X	X
FICCAR	Egresado	X	
	Jefe de Carrera	X	
	Docente	X	
	WebMaster	X	X

En la Tabla 14 podemos establecer, entre otras cosas, que los actores del SI preferentemente son consumidores de datos, qué actores tienen el rol de productores en cada PI, y que alguno de los PIs (en este caso AV/NOT) son usados de igual forma por consumidores como productores de datos.

RCDQ-Actividad3. Por cada PI en el SI, usando el patrón de encuesta provisto por el método y adaptado al contexto del caso de estudio, se consultó a cada actor del SI involucrado con él. En la encuesta se utilizó la escala de Likert con valores de 1 a 5, donde 1 significaba poco importante y 5 muy importante. A partir de esto fue posible determinar cuales eran las características de DQ aplicables a cada PI. Concretamente, se eligieron las características de DQ que tuvieron puntuación mayor o igual al puntaje de corte definido igual a 5, por cada PI del SI.

En la Tabla 15 se muestra como ejemplo el resultado para el PI FPERE (Ficha Perfil Egresado).

Tabla 15. Ejemplo de la valoración dada por los actores del SI a las características de DQ para el PI FPERE.

Característica de DQ	Consumidor de Datos	Productor de datos	Ponderación Promedio
Exactitud	5	N/A	5
Complejidad	5	N/A	5
Consistencia	4	N/A	4
Credibilidad	1	N/A	1
Actualidad	1	N/A	1
Accesibilidad	5	4	5
Conformidad	1	1	1
Confidencialidad	1	1	1
Eficiencia	4	4	4
Precisión	4	4	4
Trazabilidad	1	1	1
Comprensibilidad	4	5	4
Disponibilidad	1	1	1
Portabilidad	5	4	5
Recuperabilidad	1	1	1

Algunas de las características de DQ se consideraron no aplicables (N/A) en el caso del rol del productor de datos, ya que en la mayoría de los casos no les resultó cómodo realizar una valoración desde su perspectiva de productores.

RCDQ-Actividad4. En base a los resultados de la actividad anterior y por cada una de las características de DQ seleccionadas para cada PI, se seleccionaron las acciones a implementar en el SI para que cada PI cumpliera con las características de DQ especificadas por los usuarios para ellos.

En el ejemplo del PI FPERE, se consideraron como características de DQ deseables: Exactitud, Complejidad, Accesibilidad y Portabilidad. En la Tabla 16 se muestran las acciones seleccionadas para su implementación en el SI.

Tabla 16. Acciones de DQ a implementar para lograr las características de DQ en el PI FPERE.

Id PI	Características de DQ	Acciones
FPERE	Exactitud	<ol style="list-style-type: none"> 1. Ayuda para el ingreso de la información por medio de ejemplos. Los ejemplos y la ayuda se entregara por medio de la herramienta Tooltip Text. 2. Establecer niveles de ingreso de información. Por ejemplo: obligatorios, deseable, entre otros. 3. Organizar datos en el formulario, de tal forma que los datos obligatorios queden juntos (Esto permite identificar cuando hay datos incompletos de ingresar). 4. No crear la ficha de perfil hasta asegurar que al menos, todos los datos considerados como obligatorios fueron ingresados.
	Complejidad	<ol style="list-style-type: none"> 1. Se permitirá la búsqueda de la Ficha de Perfil del Egresado, por medio del ingreso de su apellido (puede ser el apellido exacto o una parte del apellido).
	Accesibilidad	<ol style="list-style-type: none"> 1. Dar la opción de exportar el Perfil de un Egresado, en formatos pdf y txt.
	Portabilidad	<ol style="list-style-type: none"> 1. Ayuda para el ingreso de la información por medio de ejemplos. Los ejemplos y la ayuda se entregara por medio de la herramienta Tooltip Text. 2. Establecer niveles de ingreso de información. Por ejemplo: obligatorios, deseable, entre otros. 3. Organizar datos en el formulario, de tal forma que los datos obligatorios queden juntos (Esto permite identificar cuando hay datos incompletos de ingresar). 4. No crear la ficha de perfil hasta asegurar que al menos, todos los datos considerados como obligatorios fueron ingresados.

A continuación, se complementó la especificación de requisitos del SI, con la información pertinente a los PIs involucrados con ellos. En particular, en este caso de estudio se incluyó una sección en el artefacto que detallaba cada caso de uso, indicando el o los PI involucrados en el requisito, con las características de DQ consideradas relevantes. Junto al artefacto, también se anexó un listado detallado con las acciones de DQ para cada PI especificado en él. En la Tabla 17 se presenta el caso de uso Crear Cuenta Egresado.

Tabla 17. Caso de Uso Crear Cuenta Egresado.

Caso de Uso: Crear Cuenta Egresado
ID: CU-01
Breve Descripción: Este Caso de Uso permite al actor, crear un perfil en el Sistema, el que registrará sus datos personales, académicos, laborales y de Autenticación para la cuenta.
Actores Principales: Egresado
Actores Secundarios: No existen actores secundarios
Productos de Información: 1. FPERE (<i>Ficha de Perfil Egresado</i>): Posee los datos de perfil de un Egresado (Datos Personales, de Contacto, Laborales, Académicos y de Ingreso al Sistema). • Características de Calidad de Datos: Exactitud, Completitud, Accesibilidad y Portabilidad.
Precondiciones: 1. La persona que desee crear una cuenta mediante este Caso de Uso, debe haber egresado de alguna de las carreras pertenecientes al Departamento de Ciencias de la Computación y Tecnologías de la Información.
Flujo Principal: 1. El Caso de Uso comienza cuando el actor selecciona la opción "Crear Cuenta". 2. El Sistema muestra un contrato que muestra las Condiciones y Reglas del Registro en el Sistema. 3. El actor acepta los términos y procede con el registro. 4. El Sistema despliega el formulario de registro, el cual consta de cinco secciones: Datos Personales, Datos de Contacto, Datos Académicos, Datos Laborales y Datos de Ingreso al Sistema en el Sistema. El detalle de cada sección se muestra a continuación: • Datos Personales: Nombre, Rut, Fecha de Nacimiento, Sexo. • Datos De Contacto: Dirección, Ciudad y Región de Residencia, Correo Electrónico y Teléfono (opcional). • Datos Académicos: Carrera de Egreso, Año de egreso. • Datos Laborales: Lugar de Trabajo (Empresa), Cargo que ocupa y Tecnologías que utiliza en el cargo. • Datos de Ingreso al Sistema en el Sistema: Nombre de Usuario y Contraseña (Con confirmación). 5. El actor ingresa los datos marcados como obligatorios, y los opcionales que desee rellenar, y presiona en Registrar. 6. El Sistema valida los datos ingresados. 7. La validación de los datos es exitosa y el Sistema crea la cuenta del Egresado con los datos ingresados. 8. Se finaliza el Caso de Uso, y el Egresado puede ingresar con los datos de Autenticación que determino para su cuenta.

Post-Condiciones:

1. El egresado crea su cuenta en el Sistema.
2. El egresado puede ingresar a su perfil, mediante los datos de acceso que el mismo eligió.

Flujos Alternativos:

- 3.1. El actor no acepta las Condiciones y Reglas del Registro en el Sistema, y no continua con el proceso de registro.
- 3.2. El Sistema retorna a la página de inicio, y el Caso de Uso termina.
- 7.1 La validación de los datos es errónea, y no se puede completar el proceso de registro. Esto requiere que el actor reingrese los datos verificando los datos erróneos.
- 7.2. El Caso de Uso retorna al paso 4.

RCDQ-Actividad5. En esta última actividad del ciclo, se realizó un seguimiento, usando la lista de control propuesta por el método, para verificar que las especificaciones de requisitos relacionados con PIs, contengan las características de DQ.

Como resultado de esta tarea, algunas especificaciones quedaron en estado Pendiente. Por lo tanto, se realizó una segunda iteración que incluyó nuevas especificaciones de requisitos y PIs definidos con sus correspondientes características de DQ. En la etapa de verificación de requisitos de la segunda iteración se determinó que todas las especificaciones de requisitos quedaron centradas en la DQ.

A partir de este caso de estudio se obtuvo la siguiente experiencia:

- Al realizar dos iteraciones de DeWIK, la segunda iteración fue realizada en un tiempo considerablemente más reducido que la primera iteración. La etapa de verificación de DQ de la primera iteración permite identificar claramente las consideraciones faltantes, por lo tanto, su aplicación es bastante más fluida.
- Dado que las encuestas para considerar las características de DQ son extensas, lo más apropiado es que el GDQ o el líder de desarrollo las supervise personalmente.
- Los usuarios, al ser encuestados, indicaron que al tener claro el alcance del PI, es más fácil identificar las características de DQ que se desea incorporar.

Asimismo, la implementación de DeWIK generó la oportunidad de pensar en algunos ajustes al método para mejorar su funcionamiento y para cumplir con sus objetivos, entre ellos:

- Definir nombres para los artefactos que permitan referenciarlos con mayor facilidad.
- Estudiar la posibilidad de agregar un nuevo artefacto que permita determinar la relación entre los PIs, esto con el objetivo de identificar el impacto que puede tener un PI al ser modificado y sus acciones involucradas para incluir las características de DQ.
- Incluir en la encuesta que se realiza a los actores del SI para identificar las características de DQ, la descripción del PI encuestado. Esto permitirá clarificar el contexto del PI consultado.
- Estudiar la participación del usuario productor de datos sólo en aquellas características que son dependientes del sistema por tener una participación más directa en ellas.
- La aplicación del método permite que tanto usuarios como desarrolladores tengan de forma temprana consciencia de la necesidad de calidad de datos de los PIs que luego les proveerá un SI.
- En el largo plazo, el método posibilita que una organización tenga identificados sus PIs relevantes y de ese modo tenga una base sólida para emprender un proceso de evaluación y mejora continua de los mismos.

La concreción de los ajustes al método se realizará una vez que se disponga de los resultados de otros casos de estudio que están aún en desarrollo.

CONCLUSIONES Y TRABAJO FUTURO

En este artículo se ha presentado una propuesta para abordar durante el proceso de desarrollo de un SI, la implementación de mecanismos que garanticen en el futuro uso del SI el nivel adecuado de Calidad de los Datos. En particular, cuando recién se comienzan a definir los PIs involucrados con él, y se pueden definir de manera temprana los requisitos de Calidad de Datos que éstos deben cumplir de acuerdo a las necesidades de los usuarios que los usarán.

Concretamente, se propone el método DeWIK que guía a un equipo de desarrollo de software a incorporar la DQ como parte de su proceso tradicional de desarrollo. Este método se fundamenta en el concepto de PI, la norma de calidad de datos ISO/IEC 25012 y el enfoque TDQM. DeWIK es iterativo, y cuenta con un conjunto de actividades que deben ser desarrolladas en torno a la especificación de los requisitos de un SI. Asimismo se ha presentado un caso de estudio que permite clarificar su aplicación.

Las contribuciones y beneficios de esta propuesta pueden ser resumidos en:

- DeWIK puede ser utilizado en conjunto con cualquier proceso de software orientado a los requisitos y que utilice cualquier artefacto para documentar la especificación de éstos.

Es importante acotar, que al igual que cualquier proceso de software, DeWIK puede ser adaptado y ajustado por cada equipo de desarrollo a su propio sistema de trabajo.

El trabajo futuro, a corto plazo, se orienta a la obtención de resultados de algunos casos de estudio que en los cuales se está aplicando DeWIK. Esto con el objeto de obtener la realimentación necesaria para realizar los ajustes que fueran convenientes. Además, por otro lado, es de interés realizar estimaciones relativas a la incidencia del uso del método en los recursos requeridos por un proyecto de desarrollo. DeWIK se ha aplicado hasta ahora en 4 casos de estudio y se espera contar prontamente con los resultados de todos ellos.

REFERENCIAS

- [1] A. Dedeke. "A conceptual framework for developing quality measures for information systems". In International Conference on Information Quality, pp. 126-128. Cambridge, Massachusetts, USA. 2000.
- [2] R. Wang. "A Product Perspective on Total data Quality Management". Communications of the ACM. Vol. 41, Issue 2, pp. 58-65. 1998.
- [3] C. Fisher, E. Lauria, S. Chengalur-Smith and R. Wang. "Introduction to information quality: M.I.T.". Information Quality Program. 2006.
- [4] D. Ballou and R. Wang. "Modeling information manufacturing systems to determine information product quality". Management Science. Vol. 44, pp. 462-484. 1998.
- [5] B.D. Klein. "User perceptions of data quality: Internet and traditional text sources". Journal of Computer Information Systems. Vol. 41, pp. 9-18. 2001.

- [6] D. Strong, Y. Lee and R. Wang. "Data Quality in Context". *Communications of the ACM*. Vol. 40, Issue 5, pp. 103-110. May, 1997.
- [7] ISO/IEC-25012. "ISO/IEC 25012: Software Engineering - Software Quality Requirements and Evaluation (SQuaRE) - Data Quality Model". 2008.
- [8] R. Pressman. "Software Engineering: a Practitioner's Approach". Sixth ed.: McGraw-Hill. 2006.
- [9] G. Tayi and D. Ballou. "Examining data quality". *Communications of the ACM*. Vol. 41, Issue 2, pp. 54-57. 1998.
- [10] L. Pipino, Y. Lee and R. Wang. "Data quality assessment". *Communications of the ACM*. Vol. 45, Issue 4, pp. 211-218. 2002.
- [11] S. Madnick, R. Wang, L. Yang and H. Zhu, "Overview and Framework for Data and Information Quality Research". *Data and Information Quality*. Vol. 1, Issue 1, pp. 1-22. 2009.
- [12] A. Koronios, S. Lin and J. Gao. "A Data Quality Model For Asset Management in Engineering Organisations". In *International Conference on Information Quality*, pp. 27-51. Cambridge, Massachusetts, USA. 2005.
- [13] V. Storey and R. Wang. "Modeling quality requirements in conceptual database design". In *International Conference on Information Quality*, pp. 64-84. Cambridge, Massachusetts, USA. 1998.
- [14] A. Schmidt and B. Otto. "A Method for the Identification and Definition of Information Objects". In *International Conference on Information Quality*, pp. 214-228. Cambridge, Massachusetts, USA. 2008.
- [15] M. Mielke. "IQ Principles in Software Development". In *International Conference on Information Quality*, pp. 358-369. Cambridge, Massachusetts, USA. 2005.
- [16] E. Pierce. "Developing, implementing and monitoring an information product quality strategy". In *International Conference on Information Quality*, pp. 13-26. Cambridge, Massachusetts, USA. 2004.
- [17] L. Orman, V. Storey and R. Wang. "Systems Approaches to Improving Data Quality". In *International Conference on Information Quality*, pp. 117-126. Cambridge, Massachusetts, USA. 1996.
- [18] L. Jiang. "Data Quality By Design: A Goal-Oriented Approach". In *International Conference on Information Quality*, pp. 249-263. Cambridge, Massachusetts, USA. 2007.
- [19] C. Guerra-García, I. Caballero, I. De Guzmán, y M. Piattini. "Modelado de Requisitos de Calidad de Datos en el Proceso de Desarrollo de Portales Web". En *Taller de Desarrollo de Software Dirigido por Modelos en JISBD*. San Sebastian, España, pp. 124-133. 2009.
- [20] M. Scanapiecco, B. Pernici and E. Pierce. "IP-UML: Towards a methodology for quality improvement based on the IP-MAP framework". In *International Conference on Information Quality*, pp. 279-291. Cambridge, Massachusetts, USA. 2002.

Copyright of INGENIARE - Revista Chilena de Ingeniería is the property of Universidad de Tarapaca and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.