



Universidad del Bío-Bío
Facultad de Ciencias Empresariales
Departamento de Ciencias de la Computación y Tecnologías de la
Información

Tesis de Magíster en Ciencias de la Computación

Desarrollo de un Método para Definir
Requisitos de Software Centrados
en la Calidad de los Datos

Chillán (Chile), Octubre 2011

Alumna: Lucía Alejandra Fuentes L.

Directora de Tesis: María Angélica Caro G.

Co-Directora de Tesis: María Antonieta Soto Ch.

Dedicatoria

Agradecimientos

Índice General

1	Introducción	15
1.1	Planteamiento y Justificación del Trabajo	17
1.2	Hipótesis y Objetivos	18
1.2.1	Objetivo Principal	18
1.2.2	Objetivos Específicos	18
1.3	Organización de la Tesis	19
2	Metodología de Trabajo	21
2.1	Método de Investigación Acción	23
2.1.1	Aplicación del método de trabajo en esta tesis	25
2.2	Etapas de trabajo en esta tesis	26
2.3	Resultados de la RSL	27
3	Estado del Arte	30
3.1	Calidad de Datos	32
3.1.1	Sistemas de Información y Problemas de DQ	33
3.2	Desarrollo de Sistemas de Información y DQ	34
3.2.1	Especificación de Requisitos	35
3.2.2	Diseño de Software	38
3.3	Norma ISO/IEC 25012	44
3.4	Metodología TDQM	48
3.4.1	Múltiples usos de TDQM	52
3.5	Conclusiones	58
4	Desarrollo con Calidad de Información	61
4.1	Elementos básicos del método	64
4.1.1	Producto de Información	64
4.1.2	Norma ISO/IEC 25012	64
4.1.3	Enfoque TDQM	65
4.1.4	Conjugando Elementos	65

4.2 El método DeWIQ	65
4.2.1 Etapa 1: Definición del entorno de aplicación de DeWIQ	66
4.2.2 Etapa 2: Definición de los requisitos de la aplicación con DQ	68
4.2.3 Próximas Iteraciones	79
4.3 Conclusiones	82
5 Casos de Estudio	84
5.1 Caso de Estudio 1: Aplicación Web de Gestión de Información de egresados de la Carrera de Ingeniería Civil en Informática de la Universidad del Bío-Bío	86
5.1.1 Etapa 1. Definición del entorno de aplicación de DeWIQ	87
5.1.2 Etapa 2. Definición de los requisitos de la aplicación de DQ	87
5.1.3 Análisis de la Aplicación	93
5.2 Caso de Estudio 2: Sistema de Gestión Administrativa para Grupo de Robótica de la Universidad del Bío-Bío, sede Chillán, incorporando Calidad de Datos	94
5.2.1 Etapa 1. Definición del entorno de aplicación de DeWIQ	94
5.2.2 Etapa 2. Definición de los requisitos de la aplicación de DQ	95
5.2.3 Análisis de la Aplicación	101
5.3 Caso de Estudio 3: Gestor de Pacientes de un Nutricionista	101
5.3.1 Etapa 1. Definición del entorno de aplicación de DeWIQ	102
5.3.2 Etapa 2. Definición de los requisitos de la aplicación de DQ	102
5.3.3 Análisis de la Aplicación	109
5.4 Caso de Estudio 4: Sistema de Facturación a Proveedor	109
5.4.1 Etapa 1. Definición del entorno de aplicación de DeWIQ	109
5.4.2 Etapa 2. Definición de los requisitos de la aplicación de DQ	110
5.4.3 Análisis de la Aplicación	114
5.5 Lecciones aprendidas	115
5.6 Ajustes de DeWIQ	116
6 Conclusiones	119
6.1 Análisis de los objetivos propuestos/cumplidos	121
6.2 Principales Aportes	122
6.3 Contraste de resultados	123
6.4 Trabajos Futuros	123

Anexos	126
Referencias	166

Resumen

Actualmente, existe una necesidad de trabajar con datos que tengan niveles apropiados de calidad para múltiples propósitos, siendo el más importante la toma de decisiones en una organización. Dada la responsabilidad que recae en los Sistemas de Información, respecto de la calidad de los datos que proveen a sus usuarios, la comunidad de profesionales involucrados en el desarrollo de software, deben preocuparse de incorporar la calidad de datos como un elemento central.

Por otro lado, 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 o de información, tienen un valor que es transferible al consumidor. Si vemos a un Sistema de Información como un sistema cuyo propósito es 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 tesis 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 nuestro trabajo, se ha desarrollado la primera versión de un método, denominado **DeWIQ**, que guía a un equipo de desarrollo a construir Sistemas de Información centrándose en la Calidad de Datos. Concretamente, exponemos la primera versión de este método, orientada a la etapa de especificación de requisitos de un proceso de desarrollo de un Sistema de Información con Calidad de Datos.

Finalmente, se presenta la experiencia y los resultados de la aplicación de DeWIQ en cuatro casos de estudio relacionados con el desarrollo de diversos Sistemas de Información.

Abstract

Currently, there is a need to work with appropriate quality data for multiple purposes, one of the most important is the decision-making in organizations. Because the quality of the data provided to users is responsibility of Information Systems, the community of professionals involved in the information systems development, should be concerned to incorporate data quality as a central issue.

On the other hand, 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 pure 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 high quality information products, then the data quality or information quality is an important factor to consider in their production. Indeed, the data quality is often defined as "data fitness for use", i.e. data that are useful for consumers/users in a specified context of use.

Considering the above, this thesis focuses on data quality must be addressed from the beginning of the development of an information product, i.e. when the information system that generates that information product starts to be developed. Thus, in our work, we have developed the first version of a method, called **DeWIQ** (Development With Information Quality), which leads a development team to develop information systems focusing on data quality. Specifically, this first version of the method, aims at the requirements specification stage with data quality.

Finally, we present the results of applying DeWIQ in four study cases, each consists of an information system development.

Capítulo 1

Introducción

1.1 Planteamiento y justificación del trabajo

Durante los últimos años los sistemas de información (en adelante SI) han evolucionado considerablemente, constituyéndose como sólidas fuentes estructurales y lógicas para el desarrollo de cualquier organización y en un eje central de avance empresarial. Este impacto positivo en las organizaciones, lleva a incorporar los avances tecnológicos de la información, con el fin de ser efectivas en un mercado que cada día ofrece mayor competencia con eficiencia, eficacia y optimización de procesos.

Un SI se puede definir como un conjunto de elementos, ordenadamente relacionados entre sí de acuerdo con unas ciertas reglas, que aporta al sistema objeto (es decir a la organización a la cual sirve y que le marca las directrices de funcionamiento) la información necesaria para el cumplimiento de fines, para lo cual tendrá que recoger, procesar y almacenar datos, procedentes, tanto de la misma organización como de fuentes externas, facilitando la recuperación, elaboración y presentación de los mismos” (De Miguel y Piattini, 1993).

Por otro lado, dada la incorporación de la información¹ en las directrices del funcionamiento, e incluso en la misión y visión de cada organización (Hess y Talburt, 2004), se hace vital contar con un nivel de calidad en los datos suficiente para llevar a cabo estos propósitos y prevenir la presencia de problemas de calidad de datos al momento de la utilización de los mismos.

En un contexto de desarrollo de SI, Pressman sostiene que algunos desarrolladores de software continúan creyendo que la calidad es algo en lo que empiezan a preocuparse una vez que se ha generado el código (Pressman, 2005). Sin embargo, la garantía de calidad del software es una actividad de protección que debe aplicarse a lo largo de todo el proceso de desarrollo (Pressman, 2005). De igual forma debe ocurrir con los datos, definiendo criterios de calidad desde las fases tempranas de desarrollo de software (Wang et al., 1992).

La calidad de datos² (en adelante DQ, de las siglas en inglés Data Quality) suele ser definida como “datos apropiados para el uso” (fitness for use) (Wang y Strong, 1996; Cappiello et al., 2004). Esto quiere decir, que un usuario estima la calidad de un conjunto de datos usados para una determinada tarea realizada en un contexto específico, según un conjunto de criterios o dimensiones de calidad de datos (Adelakun, 1997).

Dado que la cantidad de datos manejada por los sistemas informáticos crece día a día, la DQ pasa a ser un componente clave en la utilidad de la información derivada de esos datos y, por consiguiente, la mayoría de los procesos de negocios son dependientes de este factor. Adicional a

¹ En esta tesis utilizaremos el término de “información” y “datos” como sinónimos, por lo tanto se usarán en forma indistinta.

² Los términos “calidad de datos” y “calidad de la información” serán considerados como sinónimos, por lo tanto se usarán en forma indistinta.

esto, uno de los requisitos comunes a todos los proyectos de tecnología de información es el nivel de DQ que se intercambian entre los SI y los usuarios (Ballou, 1989; ISO/IEC-25012, 2008).

Concretamente, esta tesis aborda la DQ desde las primeras fases del desarrollo de un SI, es decir aquellas que tienen como resultado la generación de los requisitos de la aplicación. Para esto, se identificarán las etapas del proceso de desarrollo de un SI cuyas actividades estén orientadas a la obtención de requisitos y se propondrá un método para abordarlas, de modo que su desarrollo esté centrado en la DQ. Este método recoge los principios de la metodología Total Data Quality Management (en adelante TDQM) (Wang, 1998), uno de los enfoques más prestigiosos y usados para mejorar la DQ en el contexto de un SI en una organización, utilizará como modelo de calidad de datos el contenido en la reciente norma ISO/IEC 25012 (ISO/IEC-25012, 2008) y usará como elemento central para su aplicación el concepto de Producto de Información.

1.2 Hipótesis y Objetivos

La principal hipótesis de esta tesis de Magíster es:

Es posible crear un método complementario al proceso de desarrollo de Sistemas de Información, que aborde la calidad de los datos desde sus fases tempranas de desarrollo, concretamente en la fase de especificación de requisitos.

1.2.1 Objetivo Principal

Con el propósito de comprobar la hipótesis planteada, se ha definido el siguiente objetivo general para esta tesis:

Desarrollar un método para definir los requisitos de un Sistema de Información centrados en la calidad de datos.

1.2.2 Objetivos Específicos

Para alcanzar el objetivo principal de esta tesis, se han definido los siguientes objetivos específicos:

- Estudiar propuestas existentes en la literatura que aborden la incorporación de aspectos de DQ como parte del proceso de desarrollo de software.
- Estudiar en detalle la metodología TDQM, con el fin de definir su aplicación en el proceso de desarrollo de software.

- Estudiar la norma ISO/IEC 25012 con el propósito de identificar las características de calidad de datos que pueden ser relevantes para el proceso de desarrollo de software.
- Definir un método, basado en los principios de TDQM y la norma ISO/IEC 25012, que permita abordar la DQ en aquellas etapas del proceso de desarrollo de software que estén relacionadas con la definición de requerimientos de un SI.
- Evaluar el método propuesto mediante la implementación en casos de estudio representativos.

1.3 Organización de la Tesis

Esta tesis de Magíster está organizada en siete capítulos y cuatro anexos. Los contenidos del resto del documento son:

- **Capítulo 2. Metodología de Trabajo:** En este capítulo se presentan los métodos de trabajado empleados en el desarrollo de esta tesis.
- **Capítulo 3. Estado del Arte:** Este capítulo corresponde a una descripción resumida de las propuestas existentes en la literatura en relación a la DQ durante el proceso de desarrollo de software.
- **Capítulo 4. Desarrollo con Calidad de Información:** Descripción del método desarrollado.
- **Capítulo 5. Casos de estudio:** Presenta la implementación del método en cuatro casos de estudio de proyectos abocados al desarrollo de diversos sistemas de información.
- **Capítulo 6 Conclusiones:** En este capítulo se presentan los principales aportes de este trabajo, los resultados obtenidos y las líneas de investigación que quedan abiertas para el futuro.
- **Anexos.** Los anexos incluidos ayudan a comprender de mejor forma algunos de los aspectos presentados en los capítulos anteriores. Los anexos incluidos en esta tesis son los siguientes:
 - I. Resultados de Revisión Sistemática de la Literatura.
 - II. Inducción del Método DeWIQ.
 - III. Panel de Expertos, acciones que ayudan a lograr un nivel adecuado de DQ.
 - IV. Encuesta tipo de DQ en Productos de Información.

Capítulo 2

Metodología de Trabajo

En este capítulo se presentan los métodos del trabajo utilizados para el desarrollo de esta tesis. En la primera sección, el método de Investigación-Acción y como fue abordado en el desarrollo de esta tesis. En la segunda sección se presentan las etapas de aplicación del método de trabajo. En la tercera sección se presentan los resultados obtenidos en la revisión de estudios realizada conforme al objetivo de esta tesis.

2.1 Método de Investigación-Acción

Es un método de investigación cualitativa que asocia la investigación y la práctica, de modo que la investigación informa a la práctica, y la práctica informa a la investigación.

Es propuesto por Kurt Lewin (Lewin, 1947), en el contexto de las ciencias sociales. Su idea era enlazar el enfoque experimental de la ciencia social con programas de acción social que responden a los problemas sociales principales.

Mediante la investigación-acción se pretende tratar de forma simultánea conocimientos y cambios sociales, uniendo la teoría y la práctica.

En realidad Investigación-Acción no se refiere a un método de investigación concreto, sino a una clase de métodos que tienen en común las siguientes características (Baskerville, 1999).

- Orientación a la acción y al cambio.
- Focalización en un problema.
- Un modelo de proceso “orgánico” que engloba etapas sistemáticas y algunas veces iterativas.
- Colaboración entre los participantes.

Por no ser un método concreto, ha sido definido de muchas maneras, entre ellas:

- La forma que tienen los grupos de personas de preparar las condiciones necesarias para aprender de sus propias experiencias, y hacer estas experiencias accesibles a otros. (McTaggart, 1991).
- La participación de todas las partes involucradas en la investigación, examinando la situación existente (que sienten como problemática), con los objetivos de cambiarla y mejorarla (Wadsworth, 1998).
- El proceso de recopilar de forma sistemática datos de la investigación acerca de un sistema actual en relación con algún objetivo, meta o necesidad de ese sistema; de alimentar de nuevo con esos datos al sistema; de emprender acciones por medio de variables alternativas

seleccionadas dentro del sistema, basándose tanto en los datos como en las hipótesis; y de evaluar los resultados de las acciones, recopilando datos adicionales (French y Bell, 1996).

A partir de estas definiciones podemos deducir que la investigación-acción tiene una doble finalidad: a) generar un beneficio al “cliente” de la investigación y b) generar “conocimiento de investigación” relevante. Por lo tanto, la investigación acción es de carácter colaborativo, que busca unir teoría y práctica entre investigadores y profesionales mediante un proceso cíclico. Además, está orientado a la producción de nuevo conocimiento útil en la práctica obtenida mediante el cambio y/o búsqueda de soluciones a situaciones reales que ocurren a un grupo de profesionales.

En (French y Bell, 1996) se proponen cuatro variantes que dependen principalmente de las características del proyecto de investigación:

- **De diagnóstico:** el investigador se adentra en una situación problemática, la diagnostica y realiza recomendaciones al grupo crítico de referencia, pero sin que haya un control posterior de sus efectos.
- **Participativa:** el grupo crítico de referencia pone en práctica las recomendaciones realizadas por el investigador, compartiendo con él sus efectos y resultados.
- **Empírica:** el grupo crítico de referencia realiza un registro amplio y sistemático de sus acciones y sus efectos.
- **Experimental:** consiste en evaluar las diferentes opciones que existen para conseguir un objetivo. El principal inconveniente de esta variante reside en la dificultad de poder medir objetivamente las diversas opciones, ya que por lo general serán bien aplicadas en distintas organizaciones con distintas características que enturbian los resultados de la investigación, o bien en una sola organización pero en distintos momentos, con los que el entorno experimental habrá variado.

Wadsworth (Wadsworth, 1998) identifica cuatro tipos de roles en este método:

- **El investigador:** individuo o grupo que lleva a cabo de forma activa el proceso investigador.
- **El objeto investigado:** el problema a resolver.
- **El grupo crítico de referencia:** aquél para quien se investiga en el sentido de que tiene un problema que necesita ser resuelto y que también participa en el proceso de investigación (en forma menos activa que el investigador). En él hay tanto personas que saben que están participando en la investigación, como otras que participan sin saberlo.

- **El beneficiario:** aquél para quien se investiga en el sentido de que puede beneficiarse del resultado de la investigación, aunque no participa directamente en el proceso. Puede ser el receptor de documentos, informes, etc. En este grupo, por ejemplo, caben tanto las empresas que se benefician de un nuevo método para resolver problemas en tecnologías de la información, como los técnicos que aplican dichas tecnologías.

Un proceso de investigación que emplea investigación-acción, debe seguir los siguientes pasos:

1. **Planificación:** identificar las cuestiones relevantes, que guiarán la investigación, que deben estar directamente relacionadas con el objeto que se está investigando y ser susceptibles de encontrarles respuesta. En esta actividad se buscan caminos alternativos, líneas a seguir o reforzar algo existente. El resultado es que se define otros problemas o situaciones a tratar.
2. **Acción:** variación de la práctica, cuidadosa, deliberada y controlada. Se efectúa una simulación o prueba de la solución. Es cuando el investigador interviene sobre la realidad.
3. **Observación:** recoger información, tomar datos, documentar lo que ocurre. Esta información puede proceder prácticamente de cualquier sitio. También se conoce como evaluación.
4. **Reflexión:** compartir y analizar los resultados con el resto de los interesados, de tal manera que se invite al planteamiento de nuevas cuestiones relevantes. También se conoce como “especificación del aprendizaje”.

Con estas características, el proceso definido por investigación-acción es iterativo, de forma que se va avanzando en soluciones cada vez más refinadas mediante el seguimiento de ciclos, en cada uno de los cuales se ponen en marcha nuevas ideas, que son puestas en práctica y comprobadas en el ciclo siguiente.

2.1.1 Aplicación del método de trabajo en esta tesis

Considerando los objetivos de este trabajo que es el de “Desarrollar un método para definir los requisitos de un Sistema de Información centrados en la calidad de datos”, se decide usar el método de Investigación-Acción.

Para la creación del método se ha aplicado la variante participativa del método de Investigación-Acción, para lo cual se han considerado los siguientes participantes:

- **Investigador:** el grupo de investigación conformado por la alumna tesista de este trabajo y sus profesoras guías de la Universidad del Bío-Bío sede Chillán.
- **Objeto investigado:** la calidad de datos abordada en un proceso de desarrollo de software.

- **Grupo crítico de referencia:** para realizar la validación del método propuesto en esta tesis, se implementaron casos de estudio en la Universidad del Bío-Bío en proyectos de título de alumnos de la carrera de Ingeniería Civil en Informática y en organizaciones del sector privado. Por esto, el grupo crítico de referencia está formado por los miembros pertenecientes a los equipos asociados a los casos de estudio.
- **Beneficiarios:** organizaciones que consideran que la DQ es necesaria en sus SI, usuarios que necesiten SI con DQ incorporada.

Los resultados conseguidos con la aplicación del método fueron:

- Se desarrolló un método de que abordará la calidad de los datos durante el proceso de desarrollo de software que fue aceptado por el grupo crítico de referencia.
- La alumna tesista trabajó para que los beneficios científicos sean para el investigador y práctico para los beneficiarios.
- El conocimiento adquirido puede ser aplicado en forma inmediata.
- La investigación fue desarrollada en proceso cíclico e iterativo combinando teoría y práctica.

2.2 Etapas de trabajo en esta tesis

En el marco de la metodología de investigación-acción (Baskerville, 1999), la presente propuesta contempla las siguientes etapas de trabajo:

- Realizar una RSL (Kitchenham, 2004) con el objetivo de revisar aquellas propuestas existentes que aborden la incorporación de aspectos de DQ como parte del proceso de desarrollo de software.
- Realizar una RSL de documentos relacionados con TDQM, para estudiar esta metodología en detalle.
- Identificar las características de calidad de datos que pueden ser relevantes para el proceso de desarrollo de software, una vez estudiada en detalle la norma ISO/IEC 25012.
- Elaborar un método, aplicando TDQM y conceptos relacionados con la norma ISO/IEC 25012, para abordar la DQ en el proceso de desarrollo de software.
- Ajustar el método propuesto en las etapas del proceso de desarrollo de software en las etapas relacionadas con la especificación de requerimientos.
- Aplicar el método en casos de estudio y luego evaluarlo.

2.3 Resultados de la RSL

De acuerdo al objetivo propuesto, que es el desarrollar un método para definir los requisitos de un Sistema de Información centrados en la calidad de datos, ha sido necesario analizar diferentes propuestas que aportan como base del objetivo de esta tesis.

Para esto se realizó una revisión sistemática de la literatura (ver Anexo I) de acuerdo a la metodología definida en (Kitchenham, 2004), seleccionando los aportes más relevantes de acuerdo al objetivo de esta tesis.

Como parte del objetivo, se revisaron estudios que abordan la DQ en las fases de desarrollo de software además de incluir la metodología TDQM para mejorar la calidad de los datos en un producto de información.

En la Tabla 2.1 se muestra un resumen de los resultados obtenidos de la RSL efectuada, clasificados según las fuentes usadas con mayor frecuencia.

Fuentes de Búsqueda	Encontrados	Relevantes	Primarios
Actas del ICIQ	361	73	30
Google Scholar	316	33	11
Scirus	30	7	7
Portal ACM	274	15	7

Tabla 2.1: Resumen de estudios encontrados en la RSL.

En particular, los trabajos seleccionados en esta RSL son los que se presentan a continuación en la Tabla 2.2.

Autor	Año	Dominio
A. Schmidt, B. Otto	2008	Calidad de Datos en Etapas de Desarrollo de Software
K. Hüner, B. Otto, H. Österle	2009	Calidad de Datos en Etapas de Desarrollo de Software
M. Mielke	2005	Calidad de Datos en Etapas de Desarrollo de Software
L. Orman, V. Storey, R. Wang	1996	Calidad de Datos en Etapas de Desarrollo de Software
V. Storey, R. Wang	1998	Calidad de Datos en Etapas de Desarrollo de Software
L. Liu, L. Chi	2002	Calidad de Datos en Etapas de Desarrollo de Software
L. Jiang	2007	Calidad de Datos en Etapas de Desarrollo de Software
E. Pierce	2004	Calidad de Datos en Etapas de Desarrollo de Software
C. Guerra-Garcia, C. Caballero, I. De Guzman, M. Piattini	2009	Calidad de Datos y la Norma ISO/IEC 25012
A. Koronios, S. Lin, J. Gao	2005	Desarrollo de Software y TDQM
M. Heravizadeh, J. Mendling, M. Roseman	2008	Desarrollo de Software y TDQM
K. Kerr, R. Stockdale, T. Norris	2004	Desarrollo de Software y TDQM
P. Kaomea	2005	Desarrollo de Software y TDQM
V. Sessions	2007	Desarrollo de Software y TDQM
K. Wende, B. Otto	2007	Desarrollo de Software y TDQM
Z. Gackowski	2007	Desarrollo de Software y TDQM
S. Madnick, R. Wang, L. Yang, Z. Hongwei	2009	Desarrollo de Software y TDQM
M. Mielke	2005	Desarrollo de Software y TDQM
S. Baskarada, A. Koronios, J. Gao	2006	Desarrollo de Software y TDQM
J. Wright, G. Vesonder	2004	Desarrollo de Software y TDQM
R. Boelens	2006	Desarrollo de Software y TDQM

Tabla 2.2: Estudios Primarios considerados en la síntesis de los datos

Para un mayor detalle de cómo se hizo esta revisión se encuentra en la Anexo I. La síntesis de estos resultados se presenta en la siguiente sección.

Capítulo 3

Estado del Arte

De acuerdo al objetivo planteado, que es el desarrollar un método para definir los requisitos de un Sistema de Información centrados en la calidad de datos, ha sido necesario analizar diferentes propuestas de investigación que lo abordan de manera relacionada.

En la primera sección se expondrán los resultados relacionados con el concepto de calidad de datos. En la segunda sección se presentará la revisión de la literatura relacionada con la calidad de datos en el desarrollo de sistemas de información. En la tercera sección, se presentará el resultado de investigación relacionado con la norma ISO/IEC 25012. Finalmente, en la cuarta sección, se expondrán los resultados de la revisión de estudios relacionados con TDQM.

3.1 Calidad de Datos

En la literatura relacionada con nuestra área, se suelen usar como términos equivalentes Calidad de Datos y de Calidad de Información, al igual que Característica de Calidad de Datos, Atributo, Aspecto o Dimensión. Por lo tanto, en esta tesis, utilizaremos Calidad de Datos y Calidad de Información en forma indistinta.

Entre las múltiples definiciones de DQ , una de la más utilizadas es “fitness for use” (Wang y Strong, 1996; Tayi y Ballou, 1998; Cappiello et al., 2004), en donde se destaca su adecuación al uso y su naturaleza subjetiva. Este término se utiliza con referencia a un conjunto de datos que debe poseer ciertas características, las cuales deben ser suficientes para cumplir con el objetivo que tiene definido para su uso.

Desde el punto de vista del consumidor, se puede decir que la calidad de datos es "el conjunto de características que deben satisfacer o superar las expectativas del consumidor" (Kahn et al., 2002) y es “el conjunto de características que hacen que la información tenga más valor para él” (Lesca y Lesca, 1995). Este conjunto de características puede ser denominado dimensiones de calidad de datos, referido como las propiedades o características de calidad que debe poseer la información (Lee et al., 2006).

En efecto, una de las estrategias más comunes para entender este concepto consiste en descomponer la percepción global de la DQ de un conjunto de datos en una serie de “subcalidades” denominadas dimensiones o características de DQ y establecer determinados criterios de aceptación para cada una de ellas. Dichos criterios de aceptación deben estar definidos en función de los posibles rangos de los resultados de medición. Así, si tras medir el nivel de calidad de un conjunto de datos para cada una de las características de DQ, se obtienen resultados fuera de los rangos de aceptación preestablecidos, entonces se considerará como “defectuoso” y, por tanto, se puede decidir no usarlo, o al menos saber que existe un riesgo al usarlo.

Para cada contexto (entendiéndose como tal el ámbito de una tarea) es necesario identificar aquellas características que mejor representen los requisitos de DQ de los usuarios. De esta forma,

al obtener un conjunto de características de DQ aplicables en un contexto determinado y los niveles de DQ requeridos, se puede aplicar una metodología, un modelo de DQ, marcos de trabajo y/o técnicas y herramientas para mejorar situaciones en que la DQ es insuficiente (Scannapieca et al., 2002). En el ámbito de esta tesis, el propósito es recoger las características de la DQ, definir y aplicar un método para abordar y mejorar los requisitos de un sistema de información que tenga en consideración dichas características.

Uno de los marcos de trabajo más utilizados es el propuesto por (Wang y Strong, 1996; Strong et al., 1997) , el cual presenta una clasificación de las dimensiones de DQ en cuatro categorías definidas, desde el punto de vista de los consumidores de datos, ver Tabla 3.1.

Categorías DQ	Dimensiones DQ
DQ Intrínseca	Exactitud, Objetividad, Credibilidad, Reputación
DQ Accesibilidad	Accesibilidad y Seguridad
DQ Contextual	Reputación, Valor Agregado, Oportunidad, Completitud, Cantidad de Información
DQ Representacional	Interpretabilidad, Fácil entendimiento, Representación Concisa, Representación Consistente

Tabla 3.1: Categorías y Dimensiones de DQ (Strong et al., 1997)

Cuando nos referimos a que la calidad de datos es intrínseca, se indica que la información tiene características de DQ de manera inseparable y que no dependen de otro factor. Las características de calidad de datos de tipo contextual dependen de la situación en la cual se maneja la información. Las características de calidad de datos de tipo representacional dependen de la forma en que la información sea interpretable y fácil de entender. Finalmente, las características de calidad de datos de accesibilidad dependen de la importancia del rol de los SI en aspectos relacionados con la accesibilidad y seguridad de los datos.

Existen investigaciones que afirman que la calidad no puede evaluarse independiente de los consumidores del producto (Deming, 1986). Desde el punto de vista de la información, podemos decir que la DQ no puede ser evaluada sin tener el punto de vista de los consumidores de los datos. Ellos cuentan con más opciones de control sobre su entorno y de la información que utilizan, por lo tanto, sus evaluaciones adquieren cada vez mayor importancia.

3.1.1 Sistemas de Información y problemas de DQ

Cuando se habla de problemas de DQ, se refiere a cualquier dificultad encontrada en una o más características de DQ, de acuerdo al contexto en que el dato es utilizado.

Para resolver problemas de DQ a nivel organizacional, hay que considerar más allá de la categoría intrínseca, es decir, se debe incorporar el punto de vista contextual, de accesibilidad y

representacional. Junto con ello, incluir los datos que se encuentran almacenados en historia y producción y en la utilización de los procesos (Strong, 1997).

Los SI pueden influir de manera favorable en la resolución de problemas de DQ en las categorías descritas. Hay enfoques convencionales de DQ que emplean técnicas de control (como editar controles, implementar restricciones de integridad de base de datos, y definir control de cambios en la base de datos) que han mejorado sustancialmente la DQ intrínseca, especialmente la característica exactitud (Orman et al., 1996; Liu y Chi, 2002). Sin embargo, es necesario aplicar técnicas de mejora orientado a los procesos que producen esta información (Dedeke y Kahn, 2002).

Desde el punto de vista de la accesibilidad, los consumidores de datos perciben las barreras de acceso como un problema. Más allá de lo técnico, la facilidad con que se puede manipular el acceso a estos datos resulta clave para satisfacer sus necesidades. Por ejemplo, los datos codificados son técnicamente accesibles en formato texto, pero los consumidores de datos los ven como inaccesibles porque no pueden interpretar su codificación. Otro caso es el de grandes volúmenes de datos que son técnicamente accesibles, pero los consumidores los ven como datos inaccesibles debido al tiempo excesivo que deben ocupar para acceder a ellos.

Los SI profesionales deben entender la diferencia entre la accesibilidad técnica y los problemas de acceso amplio para los consumidores de datos. Una vez aclarada la diferencia, se pueden hacer esfuerzos para aminorar estos problemas, como por ejemplo disminuir los datos menos relevantes y mejorar la facilidad de las interfaces gráficas para ayudar al acceso (Pipino et al., 2002).

En cuanto a las categorías representacional y contextual, los consumidores de datos evalúan la DQ respecto de sus tareas, es decir, los mismos datos pueden ser necesarios en múltiples tareas con una DQ diferente en cada actividad, además estas características pueden cambiar dependiendo de sus requisitos. Por lo tanto, proporcionar datos de alta calidad implica un seguimiento de un objetivo que se encuentra en constante movimiento (Wang et al., 2001).

En resumen, los consumidores de datos pueden realizar diferentes tareas y las especificaciones de información pueden cambiar regularmente. Por lo tanto, es primordial proporcionar una alta calidad a los datos, dándoles valor a sus características y utilidad en relación al contexto de trabajo.

En definitiva, la idea es diseñar sistemas de información flexibles que sean capaces de adaptarse a las especificaciones tomando ciertas medidas de control en el uso de los recursos involucrados (Strong et al., 1997) y el objetivo de esta tesis apunta a esto.

3.2 Desarrollo de Sistemas de Información y DQ

El proceso de desarrollo de un sistema de información tiene como propósito la producción eficiente de un producto de software que reúna los requisitos del cliente. Dicho proceso, en términos

globales, tiene como resultado final información que es útil de acuerdo a la especificación de los requisitos entregados por las personas interesadas en el producto (Pressman, 2005).

Partiendo de un modelo de desarrollo de Sw genérico, como el presentado en (Pressman, 2005), a continuación se describe como ha sido abordada hasta ahora la DQ en el desarrollo de SI, y en particular en la etapa de especificación de Requisitos.

3.2.1 Especificación de Requisitos

En la literatura existen variadas investigaciones realizadas en organizaciones del mundo real y que, a partir de su experiencia, elaboran un método que incluye la calidad de los datos antes de desarrollar un SI. Tal es el caso de un enfoque metodológico que identifica y describe objetos de información esenciales en una empresa (Schmidt y Otto, 2008). La especificación completa del método incorpora una descripción precisa de cada actividad, roles responsable y técnicas para lograr los resultados requeridos.

Entre las actividades exigidas para este enfoque se pueden identificar: a) distinción entre la definición e identificación de los objetos de información, obtenidos de procesos iterativos que contempla rondas de entrevistas y talleres y b) la combinación de enfoques top-down y bottom-up que incluye el análisis de las entidades involucradas desde la perspectiva de los negocios y del desarrollo de un SI para garantizar su coherencia y común entendimiento.

Este enfoque es una herramienta adecuada para el almacenamiento de la información y el mantenimiento de las definiciones de los objetos de información. La Figura 3.1 muestra un esquema completo de este enfoque:

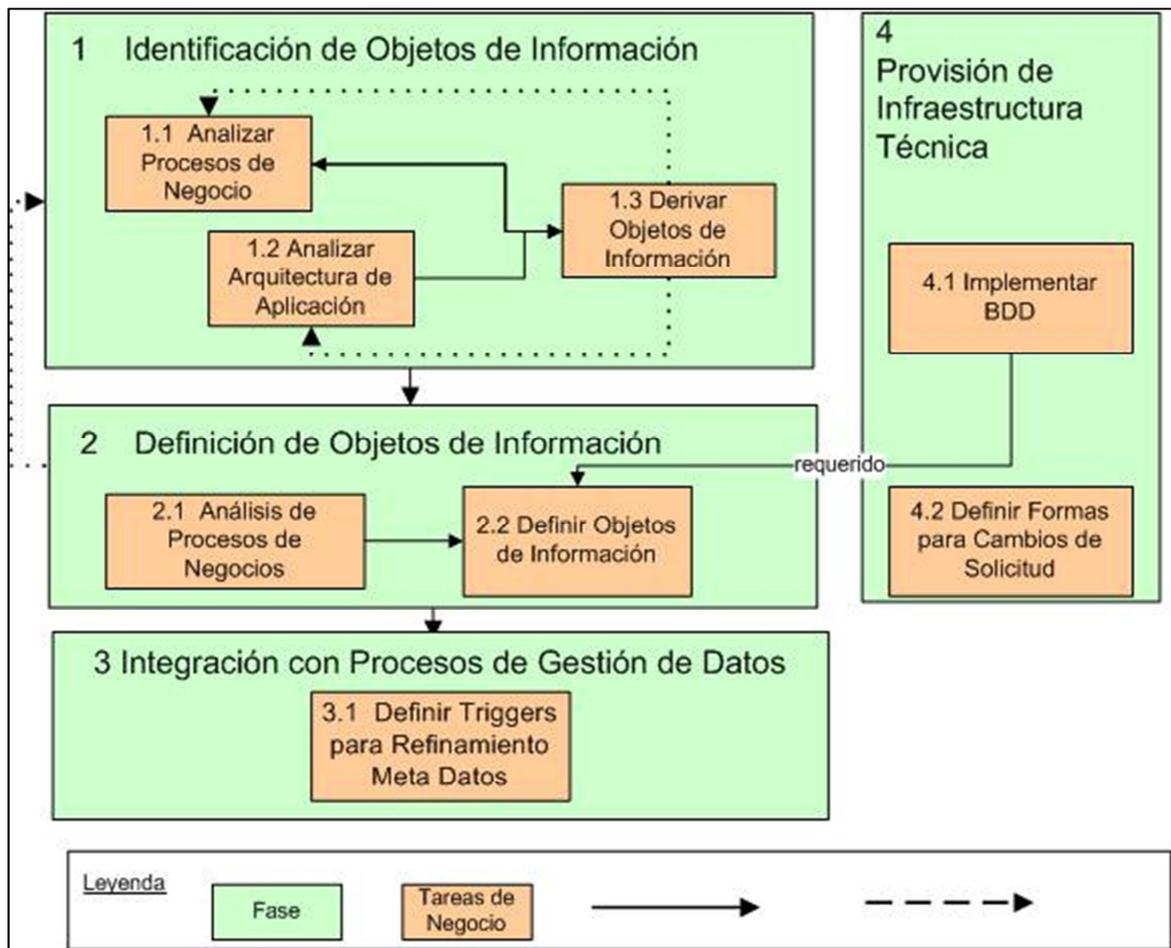


Figura 3.1: Modelo de Procedimiento General (Schmidt y Otto, 2008)

Con esta metodología los atributos de DQ quedarán libres de ambigüedades, permitiendo una mejor descripción de los objetos de información.

Otro método apunta a mirar el flujo de información a través de los procesos de negocio, y determinar qué información tendrá el SI que se construirá a través de una metodología definida en (Mielke, 2005). Este análisis debe contar con ciertas bases, tales como: a) una buena documentación que especifique los procesos de negocios b) los datos e información deben ser presentadas en función de las características de DQ requeridas c) el valor de la información de un proceso de negocio debe estar en función del riesgo de perder los datos e información y d) los objetos de información son usados principalmente para la toma de decisiones, los cuales deben tener la capacidad de adaptarse a los cambios que pueden imponer los procesos de negocio.

Al tener claro estos principios en una organización, se puede estimar el riesgo de tener malos datos y evitar trabajo adicional para corregir errores posteriores a la implementación de un SI.

La metodología propuesta, mostrada en la Figura 4, señala incluir cuestionarios a clientes que estimen las características de DQ más importantes en un objeto de información, la participación activa de los usuarios involucrados en los puntos críticos de los procesos de negocio, asignar

valores con una escala definida a las características más relevantes y ajustar el objeto de información de acuerdo a los requisitos recolectados y las características de DQ con más alta puntuación.

Para obtener las características de DQ más importantes se aplicó una matriz de relaciones de dimensiones de DQ definida en (Wang et al., 2001).

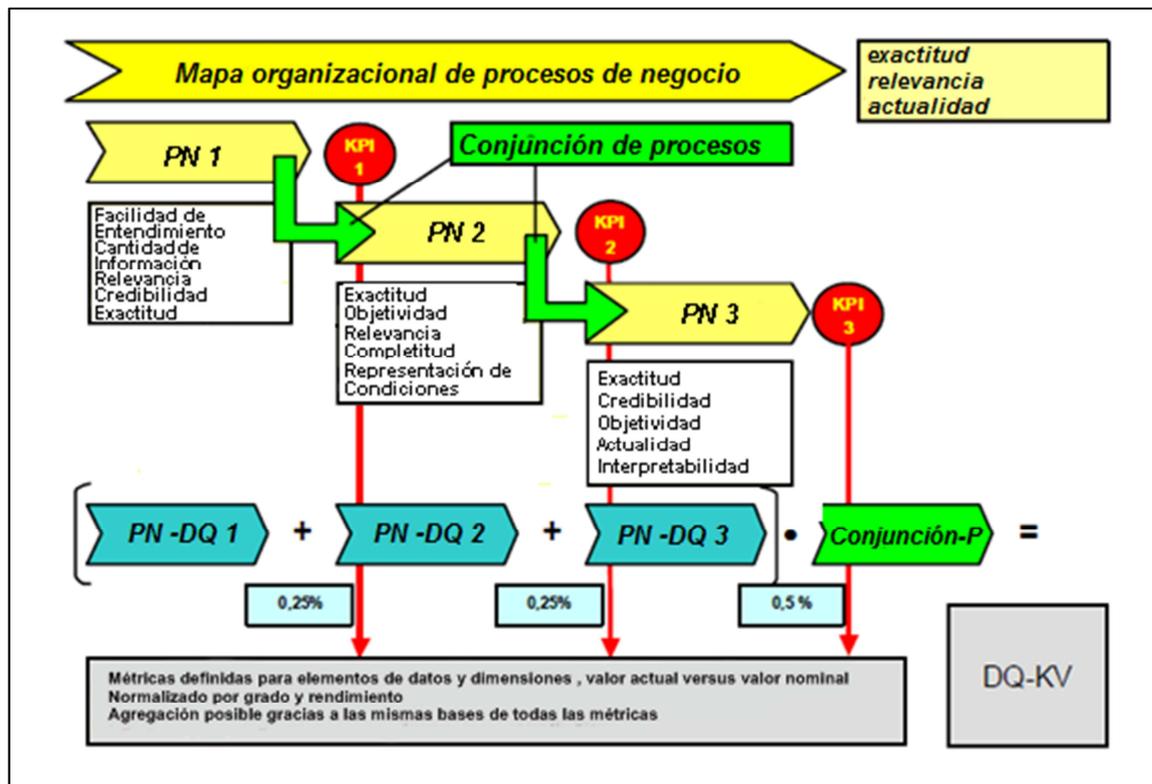


Figura 3.2: Metodología que integra procesos de negocio y DQ (Mielke, 2005).

La metodología mostrada en la Figura 3.2 ayuda a clarificar las necesidades de información y permite tener una mejor visión de la organización y alcance de los objetos de información. Con métricas claramente establecidas, se facilita la tarea de medir los requisitos de DQ, se ahorra en costo de corrección de errores en objetos de información ya establecidos y optimiza la inclusión de las características de DQ realmente necesarias.

En (Pierce, 2004) se describe una estrategia que señala cómo una empresa puede beneficiarse de técnicas y herramientas existentes para generar información con DQ. Su propósito es mejorar los servicios y la información presente en la organización para otorgar un mejor servicio al cliente y cumplir con las exigencias impuestas referente a gestión de datos y las necesidades existentes. Se requiere de la participación de la alta dirección de la empresa con el aporte de los responsables de la ejecución de los programas que apoyen la misión declarada.

Esta estrategia cuenta con las siguientes etapas:

- **Desarrollo de la estrategia:** Inspira la necesidad de crear una misión de calidad de información que exprese una visión general de la calidad de información de acuerdo a la cultura y valores comunes de la organización, seguido de una definición clara de los objetos de información que deben precisar de calidad. La estrategia de DQ debe documentar los mecanismos que se utilizarán para lograr este objetivo. A continuación se debe identificar los clientes directamente relacionados para realizarles entrevistas o cuestionarios cortos a fin de obtener las dimensiones de DQ más importantes. Finalmente, en el diseño de la estrategia, se establecen las prioridades de DQ, de acuerdo a los resultados obtenidos en las tareas anteriormente descritas.
- **Implementación de la estrategia:** En esta etapa se planifica paso a paso la introducción de la estrategia, siempre teniendo presente la forma de satisfacer las especificaciones de los clientes y sus expectativas. La estrategia debe formar parte del análisis de requisitos que se utilizan para desarrollar o modificar el SI y debe complementar las etapas posteriores, por ejemplo en el diseño del modelo conceptual, especificaciones de las bases de datos, selección de hardware, pruebas de aceptación para garantizar la calidad de la información, y por último, la operación y mantenimiento del SI.
- **Seguimiento de la Estrategia:** Durante la operación y la fase de mantenimiento de un SI, hay que controlar continuamente el rendimiento del sistema con el fin de evaluar la DQ en base a las especificaciones y expectativas de los clientes. Si el gerente o el líder del desarrollo del SI detecta diferencias entre la calidad obtenida y los criterios de los consumidores, entonces esto indica que ya es hora de revisar la estrategia e implementar cambios si es necesario. Para esta tarea se debe contar con el apoyo de la documentación existente y medir nuevamente el nivel de DQ para detectar el foco del problema y corregir.

Se garantiza el buen funcionamiento de esta estrategia, cuando hay una evaluación continua que debe satisfacer las cambiantes expectativas de los consumidores. Las organizaciones deben revisar continuamente la información generada por los SI para mantenerse al día con las innovaciones en tecnología de información y el cambio de las necesidades.

3.2.2 Diseño de Software

A fines de la década de los 90, ya se hablaba de construcción de modelos semánticos para bases de datos, en (Orman et al., 1996) se refiere a las dimensiones de DQ introducidas en el diseño de bases de datos para enriquecer semánticamente los modelos de bases de datos. Las dimensiones utilizadas en este modelo son la exactitud, completitud, precisión, actualidad y la usabilidad, las cuales son aplicadas en el modelo de datos combinando sus características para dar un mejor uso a la base de datos, elemento clave al momento de implementar el objeto de información.

Este enfoque señala limpiar datos que no cumplan con una semántica correcta, asegurar la completitud con datos faltantes y cuantificar el registro de DQ en el modelo de datos. Además de reforzar el diseño de base de datos y definir procedimientos para restringir el uso y aplicación del sistema de información.

Dado que los datos deben ser mantenidos independientes del SI, este enfoque apunta a tener un modelo que se ajuste a necesidades de calidad de datos y, por consiguiente, sus bases de datos que almacenan dicha información deben ajustarse también a estas necesidades.

En los mismos años, Storey y Wang en (Storey y Wang, 1998) hacen una propuesta para incorporar requerimientos de DQ en el diseño lógico de un SI. Clasifica los requerimientos en: a) requerimientos de aplicación, b) requerimientos de calidad de aplicación y c) requerimientos de DQ. El objetivo de esta propuesta es que el usuario no se olvide de incluir los requerimientos de calidad en la especificación de requerimientos y decidir en qué medida se utilizarán. Esta propuesta se representa a través de un esquema descrito en la Figura 3.3.

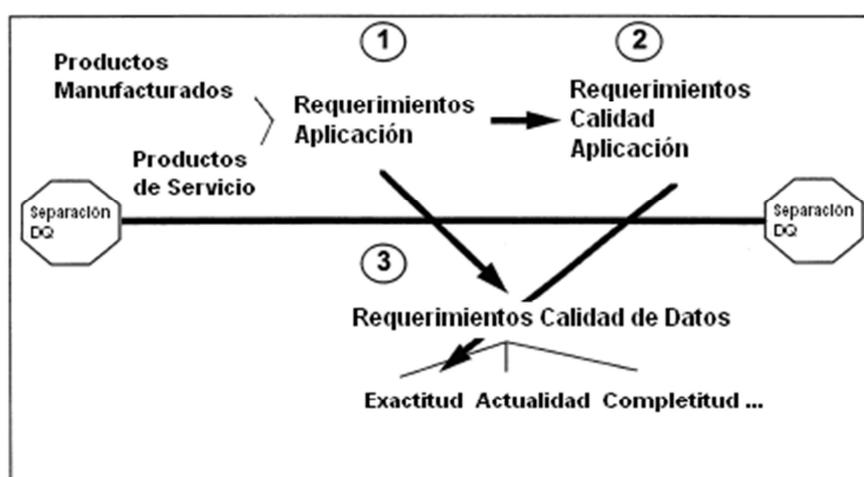


Figura 3.3: Esquema para incorporar DQ en un SI (Storey y Wang, 1998)

La propuesta destaca el trabajo del diseñador con el usuario es vital para elegir qué dimensiones de DQ son las más relevantes y aportar con su experiencia en SI de similar dominio. Para trabajar con las dimensiones de DQ se utiliza el framework descrito por Strong y Wang en (Strong et al., 1997).

Los pasos 1 y 2 mostrados en la Figura 5 corresponden a la identificación de los requerimientos de aplicación, lo cual precede la identificación de los requerimientos de DQ indicada en el paso 3.

Los requerimientos son modelados y representados a través de entidades, relaciones y atributos. Los requerimientos de DQ son identificados para medir la calidad de los valores posibles de los atributos y son almacenados en una base de datos que representa el mundo real.

Con esto se puede elaborar un esquema conceptual que incorpora estos tres tipos de requerimientos, el cual permitirá aplicar mecanismos de restricción según cada tipo.

Posteriormente en el modelado, las entidades deben ser representadas de acuerdo a entidades de aplicación, entidades de calidad de aplicación y entidades de DQ. Sus atributos deben ser correctamente representados y asociar las entidades de acuerdo a sus relaciones.

Finalmente, el modelo relacional obtenido quedará representado en una base de datos, considerando un uso eficiente de recursos al incorporar entidades y relaciones que no están en un modelo tradicional.

Otro estudio pertinente al área de interés, es presentado en (Liu y Chi, 2002). Se plantea un concepto de calidad de datos (DQ) desde una perspectiva teórica, que sostiene que la DQ se refiere al grado en que los datos satisfacen las necesidades y especificaciones de una teoría de evolución.

Los datos son capturados a través de la observación de los procesos del mundo real, la medición de los objetos tangibles e intangibles y la percepción de estímulos del mundo real. Luego, los datos se organizan y almacenan en archivos simples o en BD. Posteriormente, los datos se procesan, re-interpretan, resumen, y representan en vistas. Por último, los datos son utilizados para lograr un cierto propósito de aplicación, lo que a su vez conduce a la captura de más datos. A esta secuencia se le denomina el ciclo de datos evolutivo (ver Figura 3.4).

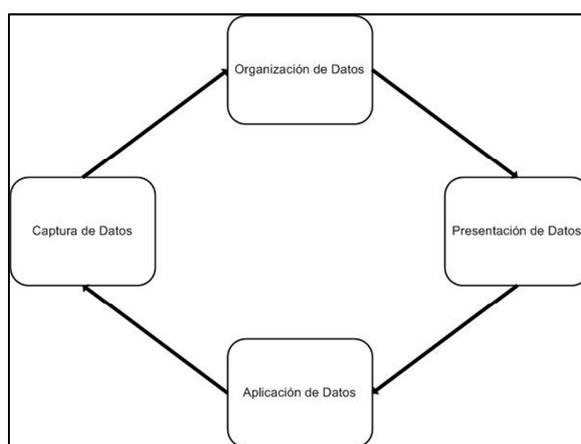


Figura 3.4: Ciclo de datos evolutivo (Liu y Chi, 2002)

La Figura 3.4 muestra el ciclo en el cual el dato va evolucionando de acuerdo a cada etapa descrita. Las transformaciones en general, hacen que el dato utilizado sea diferente de lo que se presentó, de lo que se almacena, de cómo está organizado y de cuando fue capturado. Además, cada transformación introduce errores propios de cada etapa de un proceso de desarrollo de SI. Ejemplos de tales errores son problemas en la medición durante la recolección de datos, errores de entrada de datos durante su organización, y los sesgos de interpretación para la presentación de datos.

En consecuencia, la calidad de los datos capturados no puede ser la misma que la de los datos organizados, de los datos presentados y de los datos utilizados.

En base a estos argumentos, este estudio propone un concepto de DQ evolutiva que consta de tres componentes. En primer lugar, dicho concepto enfatiza el uso de diferentes definiciones para medir la DQ en las diferentes etapas del ciclo. En lugar de un único concepto universal de DQ, se plantean cuatro puntos de vista jerárquicos de DQ, calidad en la captura, calidad en la organización, en la presentación y en la aplicación, las que se aplican para medir la calidad de los datos capturados, almacenados, presentados y utilizados, respectivamente. En segundo lugar, el concepto sugiere la naturaleza evolutiva de los cuatro puntos de vistas de DQ, es decir, la calidad de los datos en las primeras etapas del proceso evolutivo de los datos contribuye positivamente a su calidad en las etapas posteriores.

Al seleccionar un modelo adecuado para la evaluación de DQ, hay que tener en cuenta estos cuatro niveles. En cuanto a exactamente cómo se deben considerar los atributos de DQ, este estudio no da una respuesta fácil, sólo indica que las pruebas para la evaluación deben dejar una buena trazabilidad.

En esta misma etapa de desarrollo, Jiang en (Jiang, 2007) plantea, a partir de un enfoque GORE (Goal-Oriented Requirements Engineering) en donde se definen metas de calidad de alto nivel en requisitos técnicos y de operación, un enfoque orientado a definición de procesos en el contexto del diseño lógico de un SI llamado GODB (Goal-Oriented Databases). En él se aborda las especificaciones de la aplicación y la garantía de DQ. Los procesos expuestos acompañan al esquema en tiempos de diseño, y sirven como una receta que necesita ser seguida en tiempo de ejecución y que además ayuda a la documentación y en actividades que garantizan DQ.

El esquema conceptual de una base de datos normalmente es visto como una especificación de requisitos formales. Esta propuesta describe un proceso de diseño de calidad para apoyar los datos de alta calidad durante los requisitos de análisis y diseño del esquema conceptual. Un primer paso necesario, es la capacidad de diseñar un esquema común de partida conceptual de los objetivos específicos de la aplicación, realizando una representación propia y operativa en requisitos de DQ, y tener a disposición un estándar de los procedimientos operativos.

El proceso de análisis de requisitos de calidad comienza con una etapa de modelado, donde los atributos de la aplicación son identificados y organizados en un esquema conceptual. Con esto, se identifican los parámetros de calidad y se asocian con la aplicación de ciertos atributos en el esquema. Por último, cada parámetro se "refina" en uno o más indicadores de calidad. Los requisitos de calidad, una vez analizados y documentados, deben ser incorporados en un esquema conceptual durante la etapa de diseño. Esta secuencia de pasos está descrita en la Figura 3.5.

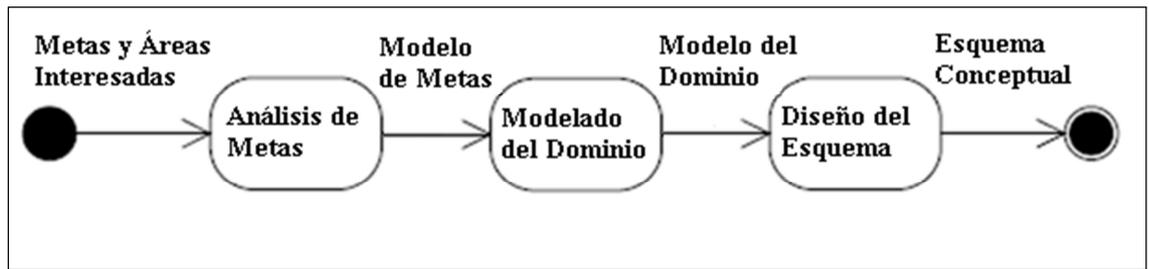


Figura 3.5: Enfoque GODB (Jiang, 2007)

GODB se divide en dos etapas: el modelado del dominio de aplicación y un esquema conceptual. El modelo de dominio se construye mediante la extracción de conceptos, relaciones y atributos de los objetivos en la alternativa de diseño elegida, la que describe la comprensión necesaria de una parte del mundo real y facilita la comunicación de los conocimientos de dominio entre los desarrolladores, usuarios finales y otros interesados. Un esquema conceptual, por el contrario, representa la semántica de los datos reales en la base de datos propuesta, su diseño se centra en asuntos específicos de datos que son relevantes en el modelado de dominio. DQ es uno de esos temas de diseño importante.

En consecuencia, este trabajo propone un enfoque de transformación del modelo de dominio a un esquema conceptual, utilizando una secuencia de operaciones de diseño.

Este proceso se inicia con un conjunto de metas de DQ y los requisitos de datos de aplicación específica obtenidos de acuerdo a lo descrito anteriormente. El último paso crea una alternativa de diseño nuevo en el modelado para satisfacer los objetivos del nivel superior, pero con una mejor atención a las metas de calidad.

En resumen, el enfoque GODB se usa para obtener un esquema conceptual de base de datos común de los objetivos específicos de la aplicación, sin tener en cuenta aspectos de calidad. Cuando las metas de calidad en el modelo son incorporadas, se pueden agregar al modelo otras variantes de este proceso, proporcionando alternativas para cumplir con diferentes niveles de apoyo a estas metas.

Por último, pero no menos importante, en (Guerra-García et al., 2009) se propone un perfil UML que modela requisitos de DQ para sistemas de información orientados a portales web. Dado que estos tipos de sistemas, se han establecido como uno de los principales recursos de información en Internet, una gran cantidad de personas acceden a todo tipo de datos con un nivel de aceptación de DQ.

Este perfil recoge aspectos considerados básicos para incorporar en la especificación y modelado del sistema, cuyas características fueron consideradas de acuerdo a lo establecido por la norma internacional ISO/IEC 25012 (ISO/IEC-25012, 2008) y las funcionalidades genéricas de un portal Web.

El objetivo de esta propuesta es permitir que los analistas y diseñadores de portales puedan especificar de forma clara qué dimensiones de DQ se deben implementar en el sistema, obtenidas a partir de la especificación de los requisitos de usuarios, y aplicadas a cada una de las funcionalidades de un portal Web. Una consecuencia de satisfacer estos nuevos requisitos de DQ, es que su incorporación va a modificar tanto el modelo de datos como el de procesos en los SI. Por lo tanto, todos los requisitos funcionales y no funcionales (incluyendo DQ), deberán ser considerados desde las etapas iniciales del desarrollo, permitiendo al diseñador del sistema reflejarlos mediante las extensiones convenientes ofrecidas por el lenguaje UML. El objetivo es especificar una notación común para definir de una forma expresiva e intuitiva ambos tipos de requisitos.

El perfil de UML desarrollado especifica cómo los conceptos de las funcionalidades de un portal y las dimensiones de calidad están relacionados y representados, haciendo uso de los estereotipos del lenguaje UML. El paquete que contiene los estereotipos definidos dentro del perfil representado con un diagrama de clases extendido, se presenta en la Figura 3.6.

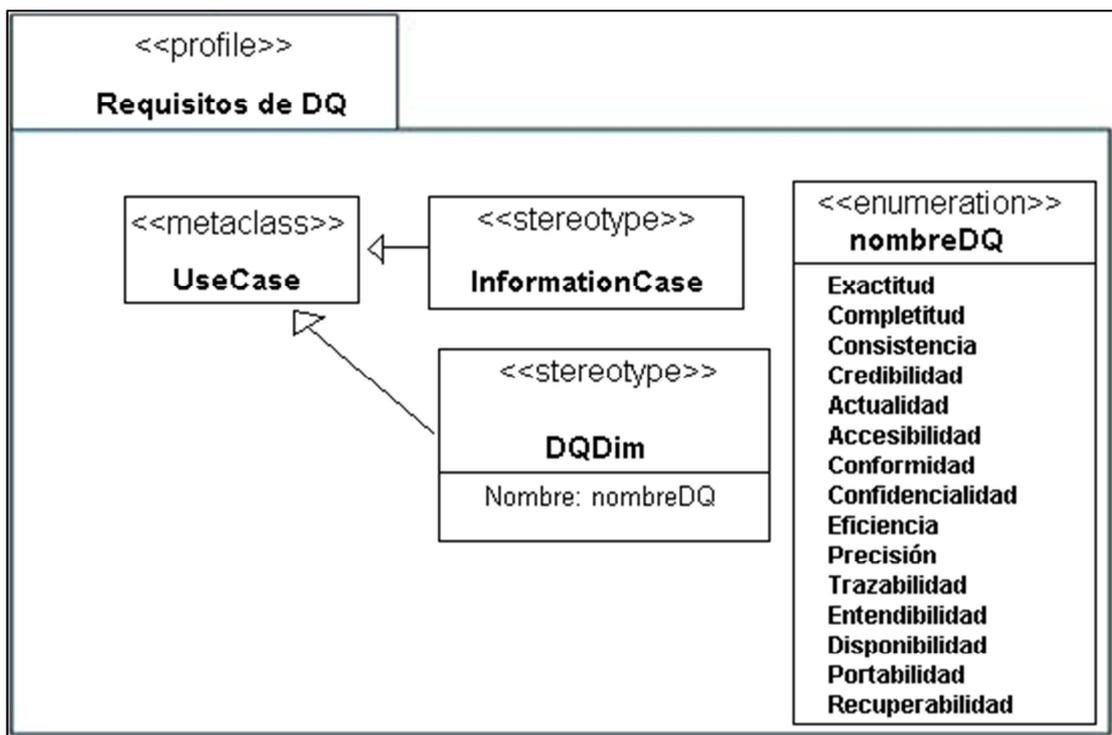


Figura 3.6: Perfil para la especificación de requisitos de DQ de un portal Web.

En la Figura 3.7 se describen los estereotipos para la especificación de requisitos de DQ, cada estereotipo contiene: nombre, clase base, descripción, restricciones y valores etiquetados (opcional).

Nombre	InformationCase
Clase Base	UseCase
Descripción	Necesario para representar los casos de información que proveerán las distintas funcionalidades del portal Web. Los casos de información a diferencia de los casos de uso, además de proveer una cierta funcionalidad, su principal función es gestionar y administrar datos. Estos datos van a ser objeto de los requisitos específicos de calidad de datos representado mediante las correspondientes dimensiones de DQ que se les asocia; se considera que la mejor forma de enlazarlos es mediante una relación de tipo "include" con los casos de uso que implementarán, de modo que se puedan satisfacer dichas dimensiones de calidad.
Restricciones	Este tiene que estar relacionado al menos con un actor.
Valores Etiquetados	No tiene.
Nombre	DQDim
Clase Base	UseCase
Descripción	Necesario para la especificación de las dimensiones de calidad de datos que se asociarán a cada uno de los casos de información de un portal Web.
Restricciones	-Este tiene que estar relacionado (tipo "include") a un caso de información. -Al menos una dimensión de calidad de datos debe ser especificada, a través de un comentario.
Valores Etiquetados	Nombre: nombreDQ (indicará el nombre específico de cada dimensión de DQ asociada.)
Nombre	nombreDQ
Clase Base	Classifier::DataType::Enumeration
Descripción	Contiene los valores posibles en relación a las diferentes dimensiones de DQ.
Valores	{ "Exactitud", "Completitud", "Consistencia", "Credibilidad", "Actualidad", "Accesibilidad", "Conformidad", "Confidencialidad", "Eficiencia", "Precisión", "Trazabilidad", "Entendibilidad", "Disponibilidad", "Portabilidad", "Recuperabilidad" }

Figura 3.7: Descripción de estereotipos para especificación de requisitos de DQ.

De esta forma, todos los casos de uso estereotipados "DQDim" que modelan las dimensiones de DQ relacionadas a los casos de información de las funcionalidades del portal Web, deberán ser tomadas en cuenta para su posterior implementación durante el proceso de desarrollo del portal.

Con este enfoque, un desarrollador de aplicaciones desearía implementar un sistema completo con una simple herramienta de desarrollo de software, gestionando tanto los aspectos de calidad de datos requeridos por los usuarios, como los aspectos propios del desarrollo, en la especificación de requisitos, el diseño, las pruebas, hasta la generación automática de código.

3.3 Norma ISO/IEC 25012

Hace pocos años surgió la norma ISO/IEC 25012 (ISO/IEC-25012, 2008), cuyo objetivo es definir un modelo genérico y que a la vez presenta de manera formal un conjunto de definiciones de conceptos relacionados con la DQ.

Según lo que indica la norma, la gestión y mejora de la DQ es importante para solucionar problemas como:

- La adquisición de datos de organizaciones en que la DQ del proceso de producción es desconocido o débil.
- La existencia de datos defectuosos que contribuyen a una información insuficiente provoca resultados inutilizables y clientes insatisfechos.

- La dispersión de estos datos entre los sistemas dueños de la información y los usuarios, a menudo carecen de una visión coherente e integrada, necesaria para garantizar la interoperabilidad y la cooperación.
- La dificultad presente en la coexistencia de la arquitectura de sistemas heredados con los sistemas diseñados y distribuidos en una organización.
- La existencia de SI de tipo web, en donde los datos cambian con frecuencia y su integración pasa a ser una situación compleja de manejar.

La detección de errores o ineficiencias de los datos, da lugar a la mejora e intervenciones correctivas a los datos producidos (análisis, limpieza, enriquecimiento, etc.), a componentes del sistema en el que residen los datos (software y hardware) y los procesos de negocio (formación de usuarios, mejora de los procesos, etc.).

Teniendo en cuenta que el ciclo de vida de los datos es a menudo más largo que el ciclo de vida del Sw, el modelo de DQ definido en esta norma pretende contribuir a:

- Definir y evaluar los datos de los requisitos de calidad de la adquisición, producción y procesos de integración de los datos.
- Identificar los criterios de garantía de DQ, además de ser útil para la re-ingeniería, evaluación y mejora de los datos.
- Evaluar la conformidad de los datos con la legislación y/o requisitos.

La Tabla 3.2 muestra las características de calidad de datos clasificados en dos puntos de vista. El símbolo X indica la relevancia de las características en el modelo de calidad de los datos requeridos o a evaluarse, a partir de los puntos de vista inherente y sistema dependiente, sin dejar de mencionar que algunas características tienen relevancia en ambos.

Características	CALIDAD DE DATOS	
	Inherente	Sistema Dependiente
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

Tabla 3.2: Características de un Modelo de Calidad presentado en la Norma ISO/IEC 25012

La norma presenta un modelo genérico de calidad de datos, categorizando la DQ en 15 características o dimensiones consideradas desde dos puntos de vista:

- Inherente: Referido al grado en el cual las características de DQ tienen el potencial intrínseco para satisfacer las necesidades, usado bajo condiciones específicas. Desde el punto de vista inherente, la calidad de los datos se refiere a los datos en sí, en particular, a los valores de datos del dominio y sus posibles restricciones. Por ejemplo, reglas de negocio que regulen la calidad requerida para las características de una determinada aplicación.
- Dependiente del sistema: Referido al grado en el cual la DQ es enriquecida y preservada dentro de un sistema cuando es usado bajo condiciones específicas. Desde este punto de vista, la DQ depende del dominio tecnológico en el que los datos se utilizan y que es alcanzado por las capacidades de los componentes de los SI, tales como: dispositivos de hardware (por ejemplo, para que los datos estén disponibles o para obtener la precisión requerida), software de sistema (por ejemplo, software de copia de seguridad para lograr valorización) y otros programas (por ejemplo, herramientas de migración para lograr portabilidad).

La Norma ISO/IEC 25012 define cada una de las características presentadas en la Tabla 3.3 y ejemplos que clarifican cada una de ellas.

Dimensión	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. Ejemplo 1: Cuando hay un bajo grado de exactitud sintáctica, la palabra María se almacena como Marj. Ejemplo 2: Cuando hay un bajo grado de exactitud semántica, el nombre de Juan se almacena como George. Ambos nombres son sintácticamente correctos, por el dominio de referencia en el que residen, pero George es un nombre diferente relacionado con otra persona.
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. Ejemplo: Para una base de datos de empleados, disminuye la completitud, si los registros de algunos empleados no contienen el dato de número de teléfono en el caso de llamar por una emergencia.
Consistencia	El grado en el cual el dato tiene los atributos libres de contradicción y son coherentes con otros datos en un contexto específico de uso. Ejemplo: Fecha de nacimiento de un empleado no puede ser posterior a su fecha de contrato.
Credibilidad	El grado en el cual el dato tiene atributos considerados como verdaderos y creíbles por usuarios en un contexto específico de uso. Ejemplo: Los datos de un certificado de una organización independiente y de confianza deben ser creíbles.
Actualidad	El grado en el cual el dato tiene los atributos que son del período correcto en un contexto específico de uso. Ejemplo: La programación de una estación de ferrocarril se debe actualizar con la frecuencia necesaria para permitir a los pasajeros tomar un tren, aunque haya cambios en la hora prevista o cambios de plataforma.
Accesibilidad.	El grado en el cual se puede acceder al dato en un contexto específico de uso. Ejemplo 1: La gente que necesita soporte de tecnología o una configuración especial debido a alguna inhabilidad (incapacidad). Ejemplo 2: Los datos que son manejados en un monitor de pantalla no deben ser almacenados como una imagen para evitar problemas con su acceso posterior.
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. Ejemplo: Un crédito bancario debe cumplir con las leyes y normas específicas.
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. Ejemplo: Datos relacionados con la información personal o confidencial como los de salud, deben ser accedidos sólo por usuarios autorizados, o deben ser registrados en forma codificada (cifrada).
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. Ejemplo: El uso de más espacio que el necesario para almacenar datos puede causar pérdida de almacenamiento, memoria y tiempo.
Precisión	El grado en el cual el dato tiene atributos que son exactos o que proporcionan la discriminación en un contexto específico de uso. Ejemplo: Una precisión de 5 decimales permite diferentes funcionalidades en lugar de una precisión de dos decimales.
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. Ejemplo: Las administraciones públicas deben mantener una bitácora con información de acceso de usuarios y de transacciones importantes como lo es el ingreso de datos confidenciales.
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. Ejemplo: Para representar a un estado (dentro de un país), un acrónimo estándar es más comprensible que un código numérico.
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. Ejemplo: Los datos deben estar disponibles tanto para la operación, dentro de una organización, como en una copia de seguridad.
Portabilidad	El grado en el cual el dato tiene los atributos que le permiten ser instalado, substituido o movido de un sistema a otro conservando la calidad existente, en un contexto específico de uso. Ejemplo: disponer de funcionalidades de importación y exportación de la información que opera en un sistema.
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. Ejemplo: Cuando un dispositivo de almacenamiento tiene una falla, los datos almacenados en el dispositivo pueden ser recuperados.

Tabla 3.3: Descripción de Dimensiones de Calidad según Norma ISO/IEC 25012

3.4 Metodología TDQM

Como se ha visto, el área de la calidad de la información ha experimentado importantes avances durante su historia relativamente breve. Se han hecho aportes para resolver problemas de DQ desde la definición, la medición, análisis y mejora de herramientas, métodos y procesos.

En este contexto nace la idea de una metodología llamada TDQM (de las siglas en inglés Total Data Quality Management), cuyo propósito es entregar productos de información de alta calidad a los consumidores de datos, facilitando la aplicación de datos globales en una organización a nivel de gestión y de alta dirección (Wang, 1998).

Dentro de un contexto organizacional, es posible señalar que una organización puede utilizar ciertas directrices para implementar un proyecto de calidad de información, identificar procesos, focos de riesgo y desarrollar procedimientos y métricas para continuar el análisis y la mejora. Las limitaciones surgen de la naturaleza de las materias primas, mirados desde el punto de vista de los datos y de los insumos tradicionales.

Para crear TDQM, se realizaron variadas investigaciones basadas en la experiencia de los procesos productivos tradicionales y los ciclos de producción de información con el objetivo de analizar las características dadas en la naturaleza de una materia prima utilizadas en la fabricación de la información, manejando sus limitaciones y potenciando sus ventajas (Firth y Wang, 1996).

Para lograr el propósito de TDQM, se propone realizar un ciclo de mejora continua a los productos de información, basándose en el principio teórico que incluye el ciclo de Deming adaptado a la fabricación de la información, es decir, **Planificar-Hacer-Comprobar-Actuar** (Deming, 1986). El ciclo de TDQM fue adaptado para: definir, medir, analizar y mejorar, en un ciclo continuo, a fin de mantener productos de información de alta calidad.

Al aplicar TDQM, una organización debe tener en cuenta lo siguiente (Wang, 1998):

- Establecer un equipo de DQ que utilizará TDQM. Si el equipo está formado por las personas adecuadas, la comunicación mejora notablemente y el éxito de implementar la metodología se hace bastante probable.
- Enseñar una evaluación de la DQ y las habilidades de gestión de los datos al equipo de DQ. Para poder poner en práctica la metodología, se debe dar la orientación que se necesita, comprobar con ellos la efectividad de implementar la metodología y dar a conocer los problemas prácticos y las posibles deficiencias de la misma.
- Definir claramente el producto de información (PI) en términos de negocio. Cuando un atributo específico del PI está mejorado con TDQM, en su primera fase de definición se describen los datos del producto con mayor claridad. Esto mejora el conocimiento sobre los atributos del PI constituyendo una base sólida para su posterior análisis.

- Institucionalizar la mejora continua de los productos de información. Una vez que la aplicación de TDQM haya finalizado, habiendo demostrado ser valioso para la mejora de la DQ, la organización debe institucionalizar el método diseñado y el equipo de DQ, para asegurar una mejora continua de sus datos.

La Figura 3.8 muestra un esquema de la metodología TDQM. Las tareas que incluye se realizan en un proceso iterativo.

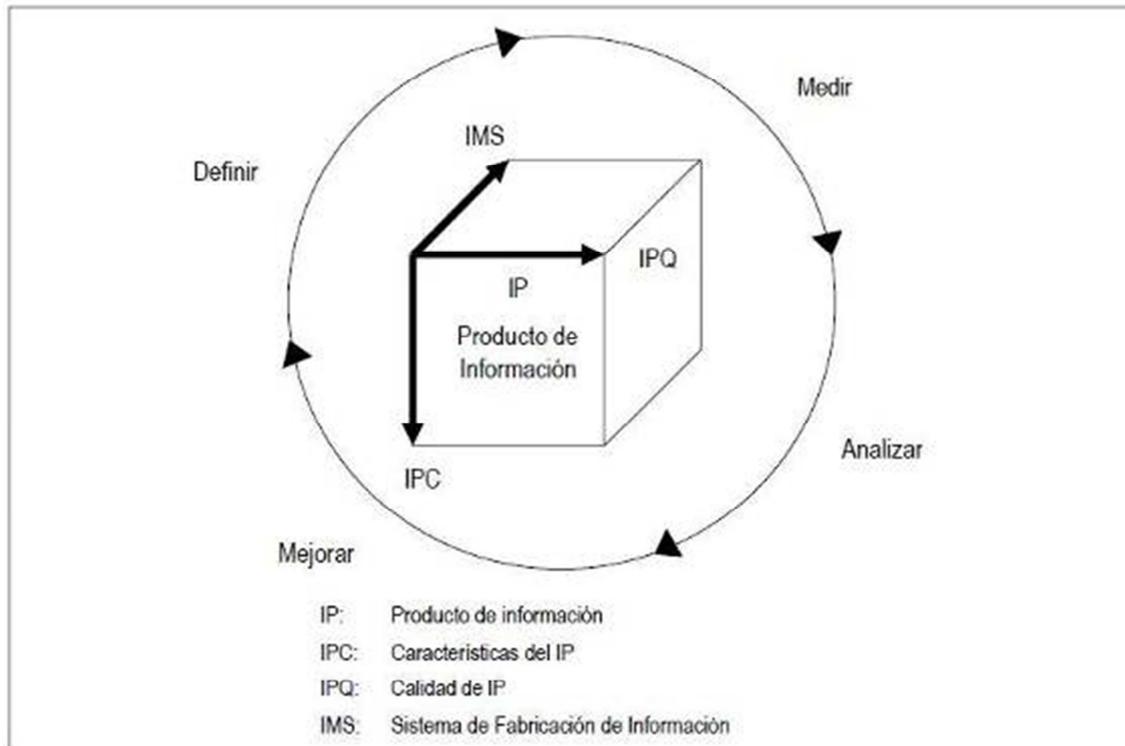


Figura 3.8: Un esquema de la metodología TDQM (Wang, 1998).

A continuación se explican las etapas que componen esta metodología:

- **Definición de Productos de Información:** Wang se refiere a un PI como el resultado que produce un sistema de información. Este concepto es incorporado para destacar el hecho que la información de salida producida por un sistema es transferible al consumidor (Wang et al., 2001).

De acuerdo a las definiciones dadas por Wang, es posible indicar algunos ejemplos comunes de productos de información en una organización.

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, Libro de Ventas.
Sistema de administración de empleados	Liquidación de Sueldo, Comprobante de Día Feriado, Informe de Cotizaciones en AFP, Informe de Comisión de Ventas.
Sistema de devoluciones de artículos.	Archivo de Stock, Notas de Crédito, Informe de Reposición de Artículos a Bodega, Informe de Motivos de Devolución, Listado de Productos Mermados.

Tabla 3.4: Ejemplos de Productos de Información

Como se puede observar en la Tabla 3.4 los PI's pueden ser muy variados, con distintos atributos dependientes del SI que se da como ejemplo.

En esta etapa de definición, se identifican las dimensiones de DQ importantes y los correspondientes requisitos de DQ (Boelens, 2006):

- Define lo que significa DQ en un PI y el nivel requerido por los usuarios.
- Define las funcionalidades de los PI's.
- Define las características de los PI's y las etapas que definen su fabricación.
- Define la unidad mínima de información, componentes y sus relaciones para los PI's a través de un modelo entidad-relación.
- Define DQ desde la perspectiva de proveedores de datos, desarrolladores y consumidores.
- Prioriza las dimensiones de DQ para PI. Esto debe incluir un ranking o una ponderación para decidir qué dimensiones son de más alta prioridad.

Para resumir, esta fase produce dos resultados clave: (1) un modelo de calidad entidad-relación que define cada PI y sus requisitos de DQ, y (2) un sistema de producción de información que describe cómo el PI es producido y las interacciones entre los proveedores de información, fabricantes, consumidores, y usuarios de gestión.

- **Medición de Productos de Información:** Identifica las métricas más eficientes para medir DQ (Guimaraes et al., 2004). Una vez obtenidas las dimensiones de DQ, estas métricas pueden obtenerse con la utilización de los datos existentes en los repositorios o con reglas de negocios previamente definidas que necesitan ser monitoreadas.

Estos indicadores pueden ser las medidas básicas de DQ (Wang, 1998) o a un nivel más complejo. Para el monitoreo de las reglas de negocio, el equipo de personas a cargo del PI debe desarrollar un método de registro que permita relacionar información orientada a métricas de PI's. Por ejemplo, formulando las siguientes interrogantes:

- ¿Qué área de la organización ha hecho la mayoría de los cambios en el sistema la semana pasada?
 - ¿Cuántos accesos de seguridad no autorizados se han intentado?
 - ¿Quiénes realizan consultas de datos, directas al repositorio de información?
- **Análisis de Productos de Información:** Permite identificar las causas que originan los problemas de DQ y calcula el impacto de los datos de mala calidad. Los métodos y herramientas útiles en esta tarea pueden ser tan simples como una inspección visual o más complejos como el control estadístico de un proceso, reconocimiento de patrones o análisis de gráficos de las dimensiones de DQ en el tiempo (Wang, 1998).

Las siguientes interrogantes ayudan en esta etapa:

- ¿De qué forma las métricas de DQ relacionan los factores críticos?
 - ¿Qué tan representativas e integrales son estas métricas?
 - ¿Estas métricas son el conjunto adecuado?
- **Mejora de Productos de Información:** Proporciona técnicas para mejorar la calidad de los datos. Los componentes se aplican a lo largo de las dimensiones de DQ de acuerdo con los requisitos especificados por el consumidor. Para esto se necesita identificar las áreas claves para las mejoras y alineamiento de los flujos de información con necesidades de los procesos del negocio.

Es fundamental para esta metodología, contar con 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.

Los consumidores tienen más probabilidades de encontrar problemas con la información que ellos utilizan, sobre todo del contexto de DQ. Sin embargo, no es el ideal que ellos detecten estas situaciones y las resuelvan. El equipo de PI, de manera proactiva, debe mejorar la calidad del PI de forma continua. Con este objetivo, los fabricantes de información, así como los proveedores de información necesitan ampliar sus conocimientos acerca del cómo y por qué los consumidores usan la información. Por otra parte, los consumidores necesitan entender de qué forma la información es producida y mantenida para que la comunicación entre los diferentes roles pueda ser efectiva.

TDQM ha demostrado ser eficaz para mejorar un PI, sobre todo cuando hay un fuerte compromiso a nivel gerencial. Las empresas actuales deben aprovechar el potencial de sus datos con el fin de ganar ventajas competitivas, oportunidades en el mercado y alcanzar sus objetivos estratégicos.

3.4.1 Múltiples usos de TDQM

De acuerdo a nuestra propuesta, no se encontraron estudios de cómo abordar la DQ en etapas de captura y análisis de requisitos. Sin embargo se consideró el estudio de (Boelens, 2006) por presentar una descripción completa e integral de cómo aplicar TDQM en una organización. A continuación se describirán las etapas de TDQM con sus tareas respectivas de acuerdo a la experiencia descrita por el autor.

3.4.1.1 Definición de Producto de Información

A continuación se describen cada una de las tareas relacionadas con esta etapa de TDQM:

- **Selección de equipo de DQ y atributos de PI:** El establecimiento de un equipo de calidad de datos es el primer paso para usar TDQM. Indica que el equipo debe consistir en:
 - Un experto TDQM, de preferencia un alto ejecutivo
 - Un ingeniero de calidad de datos que esté familiarizado con la metodología TDQM
 - Proveedores de datos
 - Fabricantes de datos
 - Consumidores de datos
 - Gerentes de PI

- **Definición de características de atributos de PI:** Para ayudar a determinar estas características, se deben responder preguntas tales como: ¿cuáles son las funciones de los atributos del producto?, ¿qué hacer con estas?, ¿dónde se utilizan?, ¿sólo disponemos de datos en el sistema de base de datos, o hay otras fuentes de datos conectadas a los atributos? (planillas excel, archivos planos, contratos o declaraciones, ¿quiénes son los proveedores de datos, los datos provistos y los usuarios de datos?

Una vez que estas preguntas son respondidas, todo el equipo es consciente de las principales características de los atributos de un PI y este conocimiento puede servir de base para el resto del ciclo.

- **Definición de las dimensiones más importantes de DQ:** Para determinar los requisitos de DQ más importantes, se debe utilizar el modelo de calidad de datos. Esto significa, representar el PI en un modelo entidad-relación cuyo propósito es identificar y describir los

diferentes atributos que componen el PI, de modo que para cada atributo se puedan determinar sus requisitos en los próximos pasos del método.

La segunda parte del método consiste en determinar las dimensiones de DQ para cada uno de los atributos descritos en el modelo ER. El propósito es determinar la calidad percibida y esperada en cada dimensión que es importante para los clientes y proveedores de información

Estas dimensiones son determinadas de acuerdo a entrevistas a los miembros del equipo, identificando aspectos importantes de DQ de un gran listado de criterios de calidad que tendrá que ser determinado.

De los resultados obtenidos, se deben elegir las dimensiones a medir, por ejemplo, la dimensión de baja puntuación en importancia a los ojos del proveedor y alta puntuación en importancia a los ojos de los clientes.

- **Definición del proceso de manufacturación de PI:** Este es el proceso en el que los datos son manufacturados. Aquí se describe todo el flujo de los datos desde el proveedor hasta el usuario.

El conocimiento de todo el proceso puede conducir a una mejor comprensión de por qué algunos pasos son necesarios y otros no, considerando los riesgos adicionales de datos de baja de calidad.

3.4.1.3 Medición de calidad de la información

El equipo de DQ debe elegir un sistema de medición. A continuación se detallan algunas formas de medir cada dimensión de calidad, usados en (Pipino et al., 2002).

- **Precisión:** representa los datos libres de error. Si se cuentan las unidades de datos con error, la métrica se puede definir como el número de unidades de datos con error dividido por el número total de unidades de datos.
- **Credibilidad:** es la medida en que los datos son considerados como confiables. Puede reflejar una evaluación de la credibilidad de la fuente de datos, la comparación con un estándar ampliamente aceptado y la experiencia previa.
- **Objetividad y reputación:** es la medida en que los datos son imparciales y sin prejuicios. La reputación es la medida en que los datos están muy bien considerados en términos de su origen o contenido. La indicación de la procedencia de los datos para algunas empresas puede ser suficiente para determinar la reputación de los datos o su objetividad.
- **Oportunidad:** es la medida en que los datos son lo suficientemente actuales para la tarea a desarrollar. La puntualidad está definida por los clientes. Ellos determinan cómo los datos

son actuales para lo que se necesita hacer. Se debe tener clara la diferencia entre disponibilidad de información en los sistemas y disponibilidad de información a los clientes. Las métricas de oportunidad deben reflejar esta última, porque es cuando los clientes pueden acceder a él para su procesamiento.

- **Compleitud:** es el grado en que los datos no faltan y cuentan con la suficiente amplitud o profundidad que requiere la tarea en cuestión. En un nivel abstracto, se puede definir el grado en que las entidades y los PI's no deben faltar en un esquema con datos específicos.
- **Interpretabilidad:** es la medida en que los datos se expresan en un lenguaje apropiado, con símbolos, unidades, y definiciones claras. Es difícil medir la interpretabilidad de un PI, las normas ISO ayudan en esta tarea estableciendo reglas que deben cumplir los PI's para asegurar su interpretabilidad.
- **Facilidad de comprensión:** es el grado en que los datos son fácilmente comprendidos. El uso de las normas ISO aumentan la facilidad de entendimiento de los PI's.
- **Representación consistente:** es el grado en que los datos se presentan en el mismo formato. Esta dimensión puede ser vista desde varias perspectivas, una de ellas es la coherencia de los valores de datos a través de tablas que contienen sus valores posibles.
- **Accesibilidad:** es el grado en que los datos son disponibles, fácilmente recuperables y rápidamente accedidos. La métrica se basa en la variable "tiempo de accesibilidad" y se mide desde la solicitud realizada por el usuario hasta su entrega, dividido por el intervalo de tiempo desde su solicitud hasta cuando los datos ya no son útiles.
- **Seguridad de acceso:** es la medida en que el acceso a los datos está restringido apropiadamente para mantener su seguridad. Una forma de medir esta dimensión está determinada por el segmento de usuarios que no deberían poder acceder a los datos.
- **Relevancia, valor agregado:** La relevancia es el grado en que los datos son aplicables y útiles para la tarea en cuestión. El valor agregado es la medida en que los datos son beneficiosos y proporcionan ventajas por su uso.
- **Cantidad de datos, representación concisa:** la cantidad de datos es el grado en que el volumen de datos es apropiado para la tarea en cuestión. La representación concisa es la medida en que los datos son representados de forma compacta. Una métrica que se puede aplicar sirve para encontrar un equilibrio entre dos razones simples: la relación entre el número de unidades de datos previstas para el número de unidades de datos necesarios, y la proporción entre el número de unidades de datos necesarios para el número de unidades de datos proporcionados.

Una vez que las mediciones hayan finalizado, se deben mostrar sus resultados. TDQM no da ejemplos de cómo mostrar los resultados de una medición, pero se puede contar con una planilla electrónica que proporcione varios tipos de gráficos para mostrar los resultados.

3.4.1.3 Análisis

A continuación, se explican las tareas que componen esta etapa:

- **Descripción de problemas específicos:** De acuerdo al estudio experiencia en desarrollo e interacción sistemas de información, entre los problemas principales se pueden encontrar:
 - Las fuentes múltiples de la misma información y que producen valores diferentes. Las dimensiones de calidad que están involucrados en esta causa son la coherencia y credibilidad.
 - Errores sistémicos que produce pérdidas de información. Las dimensiones de calidad que están involucradas con esta causa son la corrección, la integridad y la relevancia.
 - Grandes volúmenes de información almacenada hacen que sea difícil acceder a la misma en un tiempo razonable. Las dimensiones de calidad que están involucradas en esta causa son la representación concisa, la actualidad, el valor agregado y la accesibilidad.
 - Los sistemas distribuidos heterogéneos conducen a definiciones, formatos y valores inconsistentes. Las dimensiones de calidad que están involucradas en esta causa son la representación coherente, la actualidad y el valor agregado.
 - El cambio en el clima organizacional, la información que es pertinente y las modificaciones necesarias. Las dimensiones de calidad que están involucradas en esta causa son: la pertinencia, el valor agregado y la integridad.
 - La facilidad de acceso a la información pueden entrar en conflicto con los requisitos de seguridad, privacidad y confidencialidad. Las dimensiones de calidad que están involucradas en esta causa son: la seguridad, la accesibilidad, y el valor agregado.
- **Análisis de problemas:** El objetivo principal de esta fase es encontrar las causas profundas de los problemas en las diferentes dimensiones. Las causas profundas se pueden encontrar con tres métodos explicados en (Doggett, 2006). Los cuales son a) el diagrama de causa y efecto, b) esquema de interacción y c) árbol de realidad concurrente.

3.4.1.4 Mejora

A continuación se describirán las tareas de esta etapa:

- **Generación de soluciones:** Las posibles soluciones son ideas de los miembros del equipo que evitarán que el problema se siga produciendo.
- **Selección de la solución:** Al seleccionar las mejores soluciones, es necesario saber qué impacto podría tener la solución y la contribución que se espera en la mejora de la DQ. Por lo tanto, el método de asignación de causas utiliza varios criterios que deben considerarse al elegir una solución, entre estos están el costo de diseñar la solución, el costo de implementar la solución y la reducción del riesgo.
- **Plan de acción:** En esta parte la solución seleccionada se transforma en acciones concretas que deben adoptarse. Un plan de acción puede ser configurado para realizar un seguimiento de todas las acciones y estados involucrados.
- **Verificación de progreso:** Se deben evaluar los resultados de la solución implementada, haciendo seguimiento de su proceso y verificando que los resultados sean los esperados.

La forma en que está estructurado este método permite una mayor comunicación para resolver las situaciones de DQ y claridad en la asignación de las tareas presentes en cada fase de TDQM. Es posible obtener un buen plan de acción, con una rápida visualización del avance en la realización de las tareas.

La Figura 3.10 muestra un esquema visual de cómo se aplica TDQM.



Figura 3.9: Un enfoque práctico de 12 pasos para aplicar TDQM.

En la Tabla 3.5, se muestra un resumen de algunos estudios, clasificados de acuerdo a su contexto, en que se ejemplifica la aplicación de la metodología TDQM.

Contexto	Autores	Aplicación de TDQM
Procesos de Negocios	Heravizadeh M., Mendling J., Roseman, M. (Heravizadeh et al., 2008)	TDQM es una metodología que se adapta al contexto organizacional en el cual se quiere aplicar, y permite ampliar su uso agregando más fases que estén relacionadas con la integración de los procesos de negocio. Dentro de este ámbito, previo al paso de identificar las dimensiones de DQ se propone definir los objetivos del proceso de negocio, y al pasar a la siguiente fase, se logran identificar más claramente los requisitos de DQ. Este estudio propone una serie de pasos que contribuirán a este objetivo, los cuales son: definir un modelo de procesos de negocio, un modelo de DQ, un modelo de metas, un modelo de correlación entre las metas definidas, un modelo de medición e implementación de dichos modelos. Con esto se busca determinar la raíz de posibles problemas de DQ con el fin de prevenirlos.
Datos en el área de la Salud, Modelamiento de Datos	Kerr K., Norris T., Stockdale R. (Kerr et al., 2004) Kaomea P. (Kaomea, 2005)	Adicionalmente, se han hecho propuestas posteriores al modelamiento de TDQM, respecto de una estrategia de DQ para distintas áreas existentes en la actualidad. Particularmente, en el área de salud los requisitos de DQ aumentan en la medida que haya un mayor número de datos disponibles y de tecnología para extraer los datos. Un marco de trabajo de DQ, complementado con los principios de TDQM, es el paso inicial hacia una organización de datos de una amplia estrategia de calidad que se alinea con el sector de la salud, políticas y estrategias organizacionales. Una buena estrategia determina el propósito y contexto de negocio. Un marco de trabajo adecuado define las funciones clave como la adquisición de datos, conversión, recolección de datos en el futuro, el diseño, creación y mantenimiento. Se debe contar con un modelo conceptual para el análisis y la gestión de DQ, y los dominios de uso, cuyas funcionalidades deben centrarse en el significado, pertinencia y la disponibilidad que se ha desarrollado y demostrado en resultados que muestran que el marco de trabajo es útil para llamar la atención a las áreas claves que necesitan mejora, indicando las acciones que deben ser adoptadas para hacer realidad sus beneficios.
Datos de Gobierno Diseño Lógico	Sessions V. (Sessions, 2007) Wende, K., Otto B. (Wende y Otto, 2007)	Dado que TDQM debe ser adaptada de acuerdo al uso de la información, se han propuesto herramientas de aplicabilidad de esta metodología que son definidas de acuerdo al área que pretende ser mejorada. Para datos del gobierno, hay propuestas de herramientas de aplicabilidad de TDQM que establece una serie de pasos aplicados en el diseño lógico en un SI. Los datos de gobierno especifican el marco de los derechos de decisión y rendición de cuentas como parte de los datos a nivel corporativo en la gestión de DQ. Se ha ideado ampliar TDQM para contribuir a la responsabilidad de los datos en todo gobierno y al nivel de contingencias del día a día presentes en cada gobierno en forma particular haciendo la información muy flexible. Las áreas de decisión fundamentales y principales actividades de TDQM se pueden estructurar de acuerdo a lo estratégico, de organización y de aspectos técnicos.
Gestión de Datos	Gackowski Z., Madnick S., Wang R., Yang L., Hongwei Z. (Gackowski, 2007; Madnick et al., 2009) Mielke M., Baškarada S., Koronios A., Gao J. (Mielke, 2005; Baškarada et al., 2006)	Un modelo de gestión de datos ayuda a las empresas en la estructuración de sus responsabilidades en DQ. Sobre la base de la de funciones propuestas, las áreas de decisión y responsabilidades, pueden utilizar el modelo de gestión de datos con la interacción de TDQM en sus actividades y decisiones específicas. En este mismo contexto, la gestión de DQ de un SI de un área en particular puede carecer totalmente de un parámetro de comparación con otras áreas, y por consecuencia la trayectoria de los datos también es distinta. Por consiguiente, un buen modelo de gestión de datos orientados a la aplicación de los principios de TDQM es ideado para usarse de acuerdo al contexto y área que será utilizado y el flujo de procesos previos a la disposición de la información para asegurar y mantener datos de alta calidad en toda la organización.
Minería de Datos	Wright J., Vesonder G. (Wright y Vesonder, 2004)	Dentro del ámbito de minería de datos, se debe tener presente que los cambios son constantes y que la información a gran escala está presente en todo momento. Estos factores constituyen un gran desafío para realizar mejoras en DQ. Se disponen de herramientas y metodologías para aplicar TDQM en este contexto. Se da énfasis en las etapas previas a la minería de datos, es decir, poner atención en la definición y manejo de restricciones en una planificación de SI. Se agrega el conocimiento de dominio a un sistema a la planificación, lo que produce resultados sorprendentemente efectivos, aun cuando se ocupen simples algoritmos de planificación.

Tabla 3.5. Cuadro resumen de estudios más relevantes que abordan TDQM.

3.5 Conclusiones

Hoy en día las organizaciones que disponen de variados sistemas y aplicaciones, tienen conceptos de objetos del mundo real representados en una forma no consistente. Esto conduce a una multitud de sinónimos y conceptos relacionados que hacen disminuir la transparencia y coherencia en los datos. Para resolver esta situación, se debe consolidar la variedad de información y obtener objetos de datos y sistemas libres de redundancia. El método que se definirá en esta tesis, intenta resolver esta situación, identificando objetos de información a partir de los requisitos de un SI e incorpora aspectos de DQ antes de ser desarrollado e implementado.

En la literatura relacionada con estudios que abordan la calidad de los datos en la etapa de especificación de requisitos, se han encontrado métodos y estrategias con etapas claramente definidas, cuyo buen funcionamiento depende del compromiso de las personas involucradas en la generación y manejo de la información obtenida de los SI, desde el líder de desarrollo del producto hasta el consumidor de los datos. Los usuarios, ojalá con experiencia en interacción con SI, eligen los atributos de DQ, y los desarrolladores, quienes les pueden dar una visión más acertada acerca de incorporar aspectos de DQ en tiempos de diseño.

En la mayoría de los estudios revisados, relacionados con incorporar elementos de DQ en la fase de desarrollo de software, se trabaja con herramientas, técnicas y teorías de modelado ya existentes, a los cuales se les agregan aspectos de DQ. Además, sus autores señalan la conveniencia de contar con requisitos de DQ desde etapas previas al desarrollo del SI relacionadas con la especificación de sus requisitos.

Si bien en esta etapa de diseño de un SI, el hecho de incorporar DQ demanda más recursos, las etapas posteriores del proceso de desarrollo de software se verán claramente beneficiadas con una cantidad inferior de correcciones y ajustes realizados en caso de que la información obtenida de los SI no cumpla con los requisitos de DQ definidos. Al identificar claramente los errores posibles en el manejo de los datos, permite implementar con mayor facilidad el tipo de acciones que ayudarán a alcanzar determinadas características de calidad de datos.

De acuerdo al estado del arte descrito, nuestro método propuesto apunta a incorporar aspectos de DQ en la etapa de especificación de requisitos, a través de un ciclo iterativo, de acuerdo a los criterios de un equipo de personas especializadas en el método. Ellos deben contar con el apoyo de herramientas especificadas y de artefactos en apoyo de una mejor comprensión en la aplicación del método.

En esta tesis, el objetivo es guiar el desarrollo de una aplicación centrando la especificación de sus requisitos en los aspectos de DQ que sean relevantes para los usuarios. Nuestra intención es que el método pueda ser usado por cualquier equipo de desarrollo y con cualquier proceso de desarrollo que contemple una fase de especificación de requisitos.

En esta tesis, se propone que la DQ puede ser abordada desde antes de disponer de un grupo de PI's identificado. Es decir, desde el momento en que surge una necesidad de automatizar información, se puede diseñar y modelar una solución informática, incorporando el concepto de PI y DQ en ellos.

A continuación se presenta una tabla resumen de los estudios revisados y clasificados de acuerdo a nuestra área de interés, ver Tabla 3.8.

TRABAJO	ORGANIZACIÓN	PI	ISO/IEC 25012	REQUISITOS SW	BD	DESARROLLO SW
Wang, 1998	X	X				
Koronios et al, 2005	X	X				
Mielke, 2005	X	X				
Pierce, 2004	X	X				
Storey & Wang, 1998					X	
Orman, 1996					X	
Jiang, 2007					X	
Guerra-García et al., 2009			X	X		X
Scannapieca et al., 2002				X		X

Tabla 3.8. Resumen de estudios más relevantes que abordan DQ

Capítulo 4

Desarrollo con Calidad de Información

En los umbrales del siglo XXI, las organizaciones se vienen enfrentando a nuevos escenarios, caracterizados por la globalización e internacionalización de sus negocios. Este nuevo contexto impone a las empresas la necesidad vital de disponer de sistemas de información adecuados que les permitan obtener los datos necesarios para conseguir un conocimiento real y suficiente del entorno que interactúa con sus procesos. De aquí, la calidad ha adquirido un rol esencial en todos los planos en que se desenvuelven las organizaciones y, por consiguiente, la calidad de los datos se vuelve un tema esencial.

Consecuentemente con lo anterior, nuestra investigación tiene por objetivo la creación de un método, llamado DeWIQ (por las siglas en inglés de Development With Information Quality) que permita abordar la DQ desde el proceso de desarrollo de software. Concretamente, la idea es que cuando se decide emprender el desarrollo de una aplicación de software (típicamente un Sistema de Información³) en cada una de las etapas de su desarrollo, se considere la DQ de los productos de información que este generará.

En esta tesis se presenta la primera versión de DeWIQ, versión enfocada a la etapa de especificación de requisitos. En ésta, el objetivo es guiar el desarrollo de una aplicación centrando la especificación de sus requisitos en los aspectos de DQ que sean relevantes para los usuarios. Nuestra intención es que DeWIQ pueda ser usado por cualquier equipo de desarrollo (DT por las siglas en inglés de Development Team) y con cualquier proceso de desarrollo que contemple una fase de especificación de requisitos.

A diferencia de otras propuestas de la literatura (Ballou et al., 1998; Wang, 1998; Wang et. al, 2001; Schmidt y Otto, 2008), DeWIQ propone que la DQ puede ser abordada desde antes de disponer de un grupo de PI's identificados en la organización. Es decir, desde el momento en que surge una necesidad de automatizar información, se puede diseñar y modelar una solución informática, incorporando el concepto de PI y DQ en ellos. El uso de este método puede darse en un ámbito bastante amplio, no quedando condicionado a un tipo de aplicación de software, metodologías de desarrollo u otros factores gravitantes al momento de su desarrollo.

La Figura 4.1 muestra, de forma gráfica, una visión del proceso genérico de desarrollo de software (Pressman, 2005) destacando dónde es aplicable DeWIQ en esta primera versión.

³ Hacemos referencia a los Sistemas de Información como uno de los tipos de aplicaciones de software que utilizan de forma masiva y global las organizaciones y personas y que hacen uso intensivo de datos.

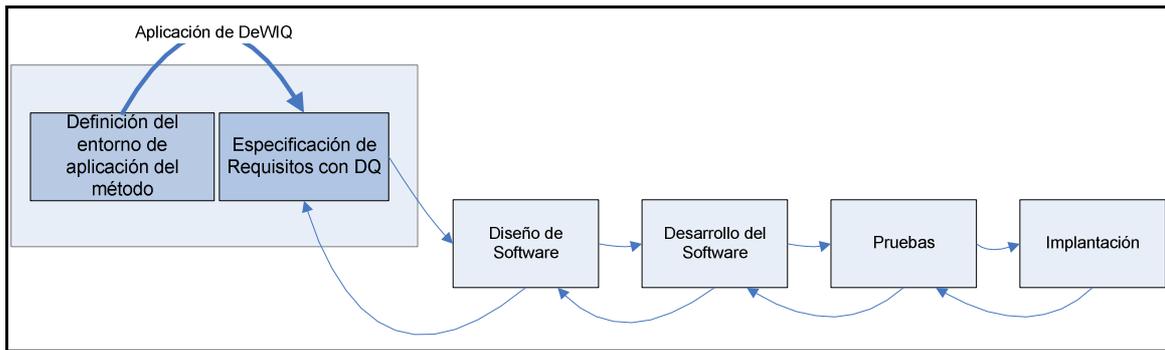


Figura 4.1: DeWIK (primera versión) y el proceso genérico de desarrollo de software.

4.1 Elementos básicos del método

DeWIK considera en su definición tres elementos básicos que son señalados a continuación:

4.1.1 Producto de Información

El concepto de Producto de Información nace de la analogía hecha entre la fabricación de productos y la fabricación de información (Wang, 1998), en el primer caso tenemos un sistema de fabricación que actúa sobre materia prima con la cual se genera un producto tangible. Análogamente, en el segundo caso, tenemos un sistema de procesamiento de datos (SI) que actúa sobre datos puros para producir productos de información.

En DeWIK, los PI's son los elementos básicos a identificar en un SI, viendo a un SI como un sistema cuyo propósito es entregar PI's de alta calidad para los usuarios finales o consumidores de información.

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.

4.1.2 Norma ISO/IEC 25012

Esta norma presenta un modelo genérico de calidad de datos, explicado en detalle en el capítulo 3 de esta tesis (sección 3.3).

DeWIK usa como referencia el modelo de calidad de datos proporcionado por la norma, utilizando sus características de DQ definidas a fin de ser incluidas en las especificaciones de requisitos de un SI a desarrollar.

DeWIK no restringe su uso sólo a este modelo, es decir, puede utilizarse perfectamente otro modelo que defina un marco de referencia de calidad de daos.

4.1.3 Enfoque TDQM

La metodología TDQM (Total Data Quality Methodology) (Wang, 1998) tiene el propósito de entregar PI's de alta calidad a los consumidores de datos, en proceso de mejora continua, facilitando la aplicación de políticas de calidad de datos globales en una organización a nivel de gestión y de alta dirección.

En el capítulo 3 en su sección 3.4, se encuentra más información de esta metodología.

4.1.4 Conjugando Elementos

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

4.2 El método DeWIQ

DeWIQ es un método iterativo que se aplicará mientras existan requisitos por definir y sea necesario determinar las necesidades de DQ en los PI's involucrados en ellos.

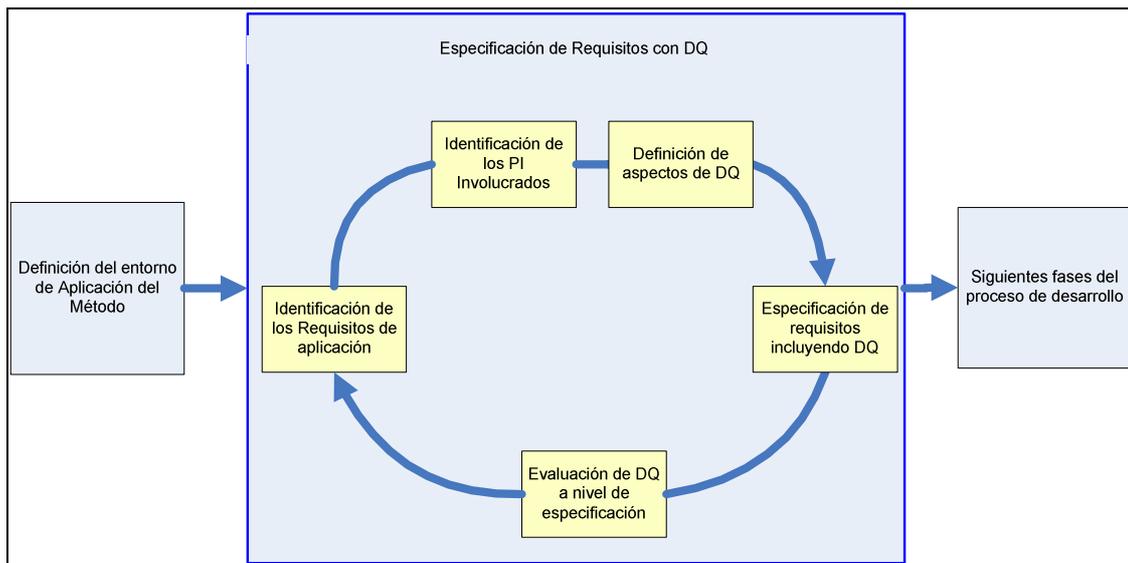


Figura 4.2: Aplicación de DeWIQ en la etapa de Especificación de Requisitos.

A continuación se detallan cada una de las etapas del método y las actividades involucradas en ellas.

4.2.1 Etapa 1: Definición del entorno de aplicación de DeWIQ

Esta etapa, que no es parte del proceso iterativo, consiste en definir y establecer las condiciones necesarias para la aplicación del método. En ella se determina quién dirigirá o será responsable de la aplicación de DeWIQ y las actividades de capacitación sobre DQ y DeWIQ necesarias para los involucrados en la especificación de los requisitos de la aplicación. Contempla tres actividades: a) Denominación y Formación del Gestor de Calidad de Datos, b) Definición de Roles, y c) Entrenamiento del DT en aspectos relacionados con DQ y el método DeWIQ. A continuación se describe cada una de ellas:

a) *Denominación y Formación del Gestor de Calidad de Datos:* El Gestor de Calidad de Datos (DQM por las siglas en inglés de Data Quality Manager) es el responsable de gestionar la aplicación de DeWIQ a lo largo de cada una de las etapas del proceso de desarrollo de software empleado, siguiendo las tareas y directrices impuestas por el método, o bien adaptándolas a la realidad concreta del proyecto en desarrollo.

Se debe designar a la persona que será la encargada de gestionar la DQ durante el desarrollo del SI. Para esto, se pueden dar las siguientes modalidades:

- El DQM, puede ser un integrante del DT o no. En ambos casos, deberá trabajar estrechamente con el resto de los integrantes del DT.
- Si el DQM ha sido designado por primera vez para desarrollar este rol, deberá instruirse en aspectos relacionados con DQ y en el método DeWIQ.
- Si el DQM ya ha desarrollado este rol en proyectos previos, probablemente ya no requerirá mayor capacitación y al contrario, mejorará la aplicación de DeWIQ en base a su experiencia previa en el desarrollo de otros SI centrados en la DQ.

Esta última modalidad es la deseable, es decir, contar con una persona, dentro de la organización o de la unidad de desarrollo, que esté preocupada permanentemente de la DQ en los SI a desarrollar. Esto permitirá incorporar políticas de calidad de datos y estrategias de formación y mejoramiento continuo en los SI implementados en una organización, tareas coordinadas por el DQM.

b) *Definición de Roles:* Esta actividad consiste en determinar las personas que estarán involucradas, directa o indirectamente, en las actividades relacionadas con la DQ durante todo el desarrollo de la aplicación y el rol que éstas tendrán. Los roles relevantes son:

- *Integrante del DT:* Encargado de tareas propias del desarrollo como puede serlo el ingeniero de requisitos, analista, arquitecto de software, programador, tester, etc. En las diversas fases de aplicación de DeWIK, cada uno de los integrantes del DT deberá desarrollar tareas relacionadas con DQ.
- *Usuario de la Aplicación:* Individuo que tiene un perfil específico dentro de la organización y en la aplicación a desarrollar, ya sea proporcionando información, operando directamente u ocupando los servicios de información que el SI entregará.

c) *Entrenamiento del DT en aspectos relacionados con DQ y el método DeWIK:* Consiste en dar a conocer al DT temas relacionados con DQ y una inducción de DeWIK para que sea aplicado de manera adecuada. Las actividades de esta etapa las realiza el DQM e incluye lo siguiente:

- Definición de los conceptos básicos que el equipo debe dominar respecto de DQ, por ejemplo: producto de información, calidad de datos, característica de calidad de datos y roles asociados a la calidad de datos.
- Definición del Modelo de DQ a aplicar: Se sugiere que se use la norma ISO/IEC 25012, la cual puede ser complementada con otro modelo que se ajuste de mejor forma al tipo de SI que se desarrollará. El objetivo de lo anterior es que se identifiquen las características de DQ que serán de interés considerar en el SI a desarrollar.
- Asignación de roles de usuario DQ: Consiste en reclasificar los roles de usuario de la aplicación a los roles clásicos de DQ como son consumidor, custodio y productor de datos. Para la aplicación del método en sus actividades posteriores, se considerará el rol de usuario DQ consumidor de datos por ser quien usa la información y el de productor de datos quien usará las reglas incorporadas a la producción de la información con calidad y que además requiere, en ocasiones, recuperar un PI para alimentar como información de entrada a otro PI, entonces el PI consumido debe considerar ciertas características de DQ específicas de acuerdo al contexto. El rol de usuario DQ custodio de datos no se considerará, en esta etapa de especificación de requisitos, por no tener influencia directa en la DQ. En una etapa posterior a la especificación de requisitos se re-evaluarán la participación de los roles de usuario DQ.
- Inducción del método DeWIK: Esta actividad tiene como propósito proporcionar conocimientos relacionados con el funcionamiento y modo de implementación y aplicación

del método durante el proceso de desarrollo del SI. El Anexo II proporciona la documentación necesaria para realizar esta inducción.

4.2.2 Etapa 2: Definición de los requisitos de la aplicación con DQ

Esta etapa tiene por objetivo abordar la fase de definición de requisitos de la aplicación incorporando los aspectos de DQ, asociados a los PI's involucrados con los requisitos, que deberán ser considerados en el desarrollo. Esta etapa se compone de 5 actividades que se desarrollan de forma iterativa hasta conseguir la especificación completa de los requisitos de la aplicación con DQ. Las actividades son las siguientes:

a) *Identificación de los requisitos de la aplicación:* Aquí el DT debe usar sus procedimientos habituales para interactuar con los usuarios y captar los requisitos, objetivos y funcionalidades que ellos requieren respecto de la aplicación a desarrollar.

En general, los requisitos suelen ser especificados mediante un artefacto que registre los detalles relevantes asociados a él. DeWIQ no está asociado a ningún artefacto particular, por lo cual puede ser usado utilizando cualquiera, lo importante es que el artefacto pueda ser adaptado para incorporar los aspectos de DQ relacionados con el requisito. En este documento, usaremos Casos de Uso para ejemplificar el método. La Figura 4.3, muestra la estructura típica de un artefacto de Casos de Uso, como se puede apreciar en otro color, este se ha ampliado para incluir información sobre los PI's involucrado con el Caso de Uso que documenta.

Caso de Uso: <<Nombre de caso de uso>>
ID: <<Identificador de Caso de Uso>>
Breve Descripción: <<Descripción máxima de dos líneas>>
Actores Principales: <<Actores Principales implicados en el caso de uso>>
Actores Secundarios: <<Actores Secundarios implicados en el caso de uso>>
Productos de Información (PI's): <<Nombre y Descripción>>
Precondiciones: <<Estado del sistema antes de que el caso de uso pueda empezar>> 1. 2.
Flujo Principal: <<Pasos del caso de uso>> 1. El Caso de Uso comienza ... 2.
PostCondiciones: <<Estado del sistema después de que el caso de uso ha sido terminado>> 1. 2.
Flujos Alternativos: <<Lista de alternativas al flujo principal>>

Figura 4.3. Caso de Uso.

En caso de que el DT use otro artefacto para la especificación de requisitos, sólo debe adaptarlo para incorporar la información de los PI's, como en el caso anterior, y de acuerdo a cómo se detalla más adelante.

b) *Identificación de los PI's involucrados con los requisitos de la aplicación:* Una vez que se han determinado los requisitos, o al menos alguna parte de ellos, corresponde identificar los PI's relacionados con cada uno de ellos. Para esto, y usando el entrenamiento previo, el DT, guiado por el DQM, deberá analizar cada requisito en busca de PI's. Como estrategia de identificación de PI's, DeWIQ provee una lista de control (ver Figura 20), donde por cada uno de los requisitos se responden interrogantes que ayudan a identificar los PI's asociados.

Lista de Control para identificar PI's				
Interrogante	Confirmación		Negación	
	Si	Acción	No	Acción
¿La implementación del requisito implica la generación de información de salida no analizada aún en esta lista de control?	<input type="checkbox"/>	Pase a la siguiente interrogante.	<input type="checkbox"/>	Finalice la determinación de PI's en el requisito analizado, pase al siguiente requisito y aplíquese esta lista.
¿Hay información de salida que es un informe?	<input type="checkbox"/>	Considere el informe como un PI. Vuelva a la primera interrogante.	<input type="checkbox"/>	Pase a la siguiente interrogante.
¿Hay información de salida que es un formulario?	<input type="checkbox"/>	Considere el formulario como un PI. Vuelva a la primera interrogante.	<input type="checkbox"/>	Pase a la siguiente interrogante.
¿Hay información de salida que es un archivo de texto?	<input type="checkbox"/>	Considere el archivo de texto como un PI. Vuelva a la primera interrogante.	<input type="checkbox"/>	Pase a la siguiente interrogante.
¿Hay información de salida que es un documento?	<input type="checkbox"/>	Considere el documento como un PI. Vuelva a la primera interrogante.	<input type="checkbox"/>	Pase a la siguiente interrogante.
¿Queda información de salida pendiente de clasificar como un PI?	<input type="checkbox"/>	Clasificar y agrupar la información de acuerdo al tipo(s) de información de salida y considerarla como PI's. Vuelva a la primera interrogante.	<input type="checkbox"/>	Vuelva a la primera interrogante.

Figura 4.4: Lista de Control para identificar PI's.

Al seguir la lista de control, las respuestas a las interrogantes irán identificando los PI's asociados al requisito.

En caso de contestar con un “sí” a la última pregunta, el miembro del DT que hará esta tarea, debe conocer la definición de un PI y el dominio del problema para realizar un correcto análisis de esta información y agruparla adecuadamente, si es necesario el DQM debe asesorarlo.

La lista de control provista puede ser ampliada y mejorada por el DQM de acuerdo a su experiencia en la aplicación de DeWIQ. Por último, algunas consideraciones adicionales sobre la identificación de los PI's:

- Para aquellos PI's que contengan sólo un atributo, se debe evaluar la posibilidad de fusionarlos con algún PI relacionado y que pueda absorber su funcionalidad.
- Aquellos PI's que comparten todos sus atributos deben ser fusionados en un sólo PI.
- Aquellos PI's cuyo atributos son un subconjunto de otro PI, deben quedar absorbidos en el PI que contenga el conjunto total de atributos.

Una vez definidos los PI's, se deben documentar en el artefacto PI's de la Aplicación presentado en la Figura 4.5.

PI's de la Aplicación					
Id PI	Nombre PI	Propósito	Descripción	Roles de Usuario de la Aplicación	Requisito
PI-1					
PI-2					
...					
PI-n					

Figura 4.5: Artefacto de Identificación de PI's.

El artefacto de la Figura 4.5 documentará todos los PI's de la aplicación a desarrollar y que han ido siendo identificados mediante el análisis de los requisitos. Por cada PI de la aplicación el artefacto contendrá los siguientes datos:

- Identificador: corresponde a la identificación que se le da al PI. Su formato es definido por el DT y el DQM.
- Nombre del PI: corresponde al nombre que lleva el PI, el cual debe ser descriptivo y preciso.
- Propósito: consiste en el objetivo que tiene el PI.
- Descripción: consiste en los atributos que lo componen.
- Roles de usuario de la aplicación: corresponde a los roles de usuario de la aplicación con los que quedarán relacionados.
- Requisito: asociados a los PI's de la aplicación.

Además, esta tabla puede ser complementada con información adicional que sirva de apoyo en caso de ser requerido, como por ejemplo fecha de actualización, número de versión, entre otros.

Luego se debe completar la información en el artefacto mostrado Matriz de Relación de Requisitos y PI's (ver Figura 4.6), la que indica la relación entre los requisitos de la aplicación y los PI's identificados.

Matriz de relación Requisitos y PI's				
Requisito	PI-1	PI-2	...	PI-n
1.-	X			
2.-	X	X		X
...				
n.-	X	X		

Figura 4.6: Artefacto Matriz de Requisitos de aplicación y sus PI's.

Esta matriz tiene como propósito la visualización de la interacción que tienen los requisitos del SI con los PI's definidos, de modo que puedan establecerse algunas consecuencias relevantes como:

- Qué PI's de la aplicación pueden ser impactados si cambia un requisito de la aplicación.
- Qué requisitos de la aplicación pueden ser afectados si se modifica un PI.
- Qué PI's son prioritarios de construir, en caso de una implementación por etapas.
- Qué requisitos cruzan con mayor frecuencia a los PI's. Estos deben ser clasificados como críticos dentro del SI y se les debieran incorporar características de DQ.

Cuando los PI's estén asociados a sus requisitos, se usará el Artefacto de Identificación de PI's, en donde los requisitos quedaron relacionados con los roles de usuario de la aplicación y se reclasificará esta asignación en roles de usuario DQ (ver Figura 4.7).

Interacción de PI's con Roles de Usuario de la Aplicación y DQ			
Productos de Información	Roles de Usuario de la Aplicación	Rol de usuario DQ	
		Consumidor de Datos	Productor de Datos
PI-1	Usuario I de Aplicación	X	
	Usuario II de Aplicación		X
	...		X
	Usuario n de Aplicación		X
PI-2	Usuario I de Aplicación	X	
	Usuario II de Aplicación	X	
	...		X
	Usuario n de Aplicación		
...			
PI-n	Usuario I de Aplicación	X	X
	Usuario II de Aplicación		
	...		
	Usuario n de Aplicación	X	

Figura 4.7: Artefacto de Interacción de PI's con Roles de Usuario de la Aplicación y DQ.

De acuerdo a la relación que tengan los roles de usuario de la aplicación con los PI's definidos, esta matriz muestra el comportamiento de los roles de usuario de la aplicación en función de los roles de usuario DQ por cada PI.

Este artefacto, por cada PI, permite visualizar lo siguiente:

- Qué roles de usuario de la aplicación interactúan más frecuentemente en su función de rol de usuario DQ: este rol de usuario tendrá mayor interacción con el PI, por lo tanto es necesaria su participación en la incorporación de características de DQ en los PI's.
- Qué usuarios de aplicación interactúan con un cierto PI: entre estos usuarios se debe determinar a quiénes se les consultará por las características de DQ que debe incluir el PI.
- Qué rol de usuario DQ es el más utilizado: permite analizar las características de DQ que deben incorporarse al PI. El rol de consumidor de datos usará el producto de información con DQ y el productor de datos es a quien se le imponen las exigencias para ingresar la información que posteriormente será parte del proceso de generación del PI.

Finalizando esta actividad de identificar los PI's, se registrará en un artefacto llamado matriz de relación de PI's la dependencia existente entre los PI's (ver Figura 4.9). Esta dependencia está dada porque un PI puede originar a otro, por lo tanto sus características de DQ están estrechamente relacionadas.

Matriz de Relación de PI's					
Producto de Información	PI-1	PI-2	PI-3	...	PI-n
PI-1		>	>		
PI-2			>		
PI-3					
...					...
PI-n	>				

Figura 4.9: Artefacto de Matriz de Relación de PI's.

Las flechas presentadas en el artefacto, indican que el PI de la fila izquierda origina al PI de la columna de derecha.

Este artefacto ayuda a una visualización global de los PI's, dando el apoyo al DQM y al DT para tomar decisiones relacionadas con la incorporación de las características de DQ, descubrir problemas posibles en la implementación de éstas y solucionar de manera rápida en esta etapa de desarrollo antes de la fase de construcción del SI. Concretamente, cuando un PI debe ajustarse a una determinada característica de DQ, se debe ubicar el o los PI's que lo generan porque sus atributos también deben estar involucrados con la característica de DQ. De lo contrario no se podrá asegurar la característica de DQ en el PI que se está analizando.

c) *Definición de los aspectos de DQ:* Por cada PI identificado se deberán definir sus características de DQ relevantes. Puede darse el caso de haber identificado una gran cantidad de

PI's y por consiguiente será una labor extensa el identificar sus características de DQ, pero para asegurar que los requisitos del SI estén realmente centrados en la DQ es necesario que todos los PI's sean sometidos a este proceso.

En esta etapa se deben elaborar encuestas, con ayuda del DQM, que entregan información sobre la importancia que los usuarios asignan a cada característica de DQ por cada PI analizado usando el modelo de DQ ISO/IEC 25012 ó el que se haya elegido como referencia. Esta encuesta debe tener una escala de puntuación claramente explicada y manejar al menos cinco valores para luego visualizar y manejar en qué rangos de valores se mueve la tendencia de los encuestados en la preferencia de las características de DQ.

Resulta lógico pensar que el usuario siempre desea que el SI a desarrollar tenga alta calidad de datos, por lo tanto las interrogantes las contestará con una probable puntuación máxima. Sin embargo, el usuario deberá tener presente que a mayor calidad incorporada mayor será el uso de recursos incorporados para lograr estos objetivos. Es por ello que por cada interrogante realizada, en la misma encuesta, se debe indicar a grandes rasgos que trabajo adicional implica la incorporación de la característica de calidad de datos consultada para que el usuario tenga una visión de los recursos extras que significa agregar esa característica de DQ.

En el anexo IV, se presenta un patrón de encuesta el cual debe ser adaptado de acuerdo al nivel de conocimientos que tiene el usuario, al contexto del SI, a la información relacionada con el PI consultado y al contexto de cada característica de DQ.

Los usuarios, en su función de rol de usuario DQ, deben indicar en la encuesta qué valor de importancia debe tener cada una de las características de DQ en cada PI que se le consulte. Esta asignación de valores debe estar guiada de acuerdo a interrogantes lo suficientemente claras para asegurar la comprensión del significado y ámbito de cada característica de DQ en los PI's que necesitan ser evaluados.

A continuación se debe consolidar la información contenida en las encuestas y realizar una tabla resumen entre los roles de usuario DQ y sus características de DQ. El artefacto mostrado en la Figura 4.10 presenta un resumen de las valoración promedio que dan los roles de usuario DQ a cada característica de DQ por cada PI definido, utilizando como marco de referencia el estándar ISO/IEC 25012.

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

Figura 4.10: Artefacto de Características de DQ y su importancia en Roles de Usuario DQ por PI.

Este resumen permite visualizar claramente cuáles características de dicho modelo son relevantes en cada PI y en qué grado. Cuando un rol de usuario DQ es compartido por más de un rol de usuario de la aplicación se debe indicar el puntaje promedio resultante de las valoraciones indicadas por cada uno de ellos en la característica de DQ en análisis.

Para obtener el valor de la ponderación promedio de cada característica de DQ, se debe realizar la siguiente fórmula:

$$Ponderación\ Promedio(CaracterísticaDQ) = \frac{(2CoD + PrD)}{3}$$

Donde,

- CoD: Valor promedio de puntuación, dado por el rol de usuario DQ consumidor de datos a la característica de DQ analizada.
- PrD: Valor promedio de puntuación, dado por el rol de usuario DQ productor de datos a la característica de DQ analizada.

Como se observa en la fórmula, se da una mayor ponderación al rol de usuario DQ consumidor de datos al ser el rol de mayor interacción con los PI's identificados en un SI y quien determina si los datos son útiles o no.

Si bien la calidad de los datos se define en términos del uso, esta fórmula debe ser aplicada a las características que son dependientes del sistema, dado que sobre ellas el productor de datos tiene una percepción más clara de la DQ en sus tareas de ingreso y actualización de datos. Además nivela las expectativas del consumidor de datos (suelen ser superiores). Para el resto de las características sólo se toma en cuenta la opinión del consumidor de datos, estas se encuentran

marcadas en la Figura 4.10 con un “N/A”, es decir, no aplica porque no obtienen puntaje del rol de usuario DQ proveedor de datos.

Se ocupa el valor promedio por ser el equilibrio que representa a todos los valores observados de medición dada a la característica de DQ. Además existe una compensación implícita entre la DQ que los usuarios requieren y su costo de alcance.

Estos valores permiten establecer qué características de DQ deben ser abordadas en un SI a desarrollar desde el punto de vista de los roles de usuario DQ consumidor y productor de datos.

d) *Especificación de requisitos incluyendo DQ:* Cuando ya se conocen las necesidades de DQ de los PI's, entonces se debe complementar con ellas la especificación de los requisitos del SI. Para esto 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, quedando un margen en referencia a la puntuación máxima definida por la escala de valores. A mayor margen se considerarán mayor número de PI's y características de DQ a incluir.

El puntaje de corte debe quedar definido, de acuerdo al criterio del DT (por tener la visión global de las soluciones que aborda el SI a desarrollar), del DQM (por ser el experto en la aplicación de DeWIQ) y del rol de usuario DQ consumidor de datos (por ser quien utilizará los datos), debido a que la planificación de los tiempos de desarrollo del SI no deben verse excesivamente aumentada por la incorporación de las características de DQ a los requisitos del SI.

En base al análisis de los resultados obtenidos de la encuesta, se debe incluir como parte de la especificación de requisitos un listado de acciones tendientes a lograr el alcance de las características de DQ, considerando no alterar la definición de los requisitos del SI a desarrollar.

Estas acciones son específicas de acuerdo al PI y deben quedar ordenadas en forma descendente en función del puntaje obtenido en la encuesta, este orden ayudará al momento de planificar su implementación en cada PI. Las características de DQ seleccionadas a incorporar deben ser aplicadas a los atributos que componen el PI.

Listado de acciones en PI's		
Producto de Información	Características de DQ	Acciones
PI-1	Característica 1 de DQ	Acción 1
		Acción 2
		Acción 3
PI-2	Característica 3 de DQ	Acción 1
	Característica 1 de DQ	Acción 2
PI-3	Característica 4 de DQ	Acción 1
		Acción 2
		Acción 3
	Característica 2 de DQ	Acción 1
		Acción 2

Figura 4.11: Listado de acciones en PI's.

La Figura 4.11 muestra un resumen de los PI's seleccionados de acuerdo al puntaje de corte definido dado a las características de DQ y su listado de acciones a realizar. Las características de DQ indicadas en el resumen quedaron seleccionadas de acuerdo a un puntaje de corte establecido y sus acciones relacionadas tienen por objetivo incorporar la característica de DQ a los requisitos definidos del SI a desarrollar. El valor ponderado representa al valor promedio dado a la característica de DQ, en base a los resultados obtenidos en la encuesta.

En cada caso de uso se deben utilizar las acciones más apropiadas de acuerdo a la utilización del PI en la especificación de requisitos y a sus PI's generados a partir de éste. Cuando un PI debe ajustarse a la característica 1, entonces debemos ubicar el o los PI's que lo originan (porque tienen los atributos involucrados con la característica de DQ), ya que estos PI's (los atributos que correspondan) deben reunir esa misma característica de DQ. El artefacto que sirve de apoyo en esta tarea es el mostrado en la Figura 4.9.

A partir de la experiencia en desarrollo de un grupo de expertos, se han especificado diferentes ejemplos de acciones que ayudan a lograr un nivel adecuado de DQ. En el anexo III se presenta el conjunto de acciones generado.

A continuación, se debe complementar la especificación de requisitos con las características de DQ definidas como importantes en la especificación de requisitos. En esta tesis, tal como se señaló anteriormente, se ha usado como referencia los casos de uso, en dicho documento se debe especificar en la sección que indica los PI's, qué características debe contener cada PI.

Caso de Uso: <<Nombre de caso de uso>>
ID: <<Identificador de Caso de Uso>>
Breve Descripción: <<Descripción máxima de dos líneas>>
Actores Principales: <<Actores Principales implicados en el caso de uso>>
Actores Secundarios: <<Actores Secundarios implicados en el caso de uso>>
Productos de Información (PI's): <<Nombre, Descripción, Características DQ>>
Precondiciones: <<Estado del sistema antes de que el caso de uso pueda empezar>> 1. 2.
Flujo Principal: <<Pasos del caso de uso>> 1. El Caso de Uso comienza ... 2.
PostCondiciones: <<Estado del sistema después de que el caso de uso ha sido terminado>> 1. 2.
Flujos Alternativos: <<Lista de alternativas al flujo principal>>

Figura 4.12: Casos de Uso con características de DQ.

La Figura 4.12 muestra el documento de caso de uso, utilizado como referencia para especificar requisitos, indicando, en la sección de los PI's, las características de DQ que deben estar incluidas.

Este documento debe anexar el documento especificado en la Figura 4.11, el que contiene las acciones para lograr las características deseadas de DQ por cada PI, seleccionando las pertinentes al caso de uso y al PI involucrado, para contar con una especificación de requisitos con elementos incorporados de DQ.

e) *Verificación de DQ a nivel de especificaciones:* Es la última actividad del ciclo, la que en términos generales determina la completitud de la especificación de requisitos en los siguientes aspectos relacionados con la DQ:

- Cada especificación de requisitos debe contener una referencia a cada uno de los PI's 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 PI's seleccionados del SI deben estar asociados al menos a una especificación de requisitos.

En términos particulares, se debe verificar que los artefactos incluidos en cada actividad contemplada en DeWIK hayan sido debidamente completados.

Como instrumento de evaluación se ha definido una lista de control (ver Figura 4.13), en la que el DT debe cotejar lo indicado anteriormente por cada requisito especificado.

Lista de Control de Verificación de Características DQ				
Interrogante	Confirmación		Negación	
	Si	Acción	No	Acción
¿La especificación del requisito incluye PI's?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	Tomar la siguiente especificación de requisitos y comience nuevamente esta lista desde la primera interrogante.
¿En la especificación del requisito se han definido las características de DQ asociadas al PI?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	En caso de que el PI tenga sus características de DQ definidas, complete el artefacto de "especificación del requisito" y luego pase a la siguiente interrogante. En caso contrario, deje esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. Si el requisito no tiene más PI's, finalice el seguimiento de esta lista y pase al siguiente requisito.
¿El PI se ha identificado con su información en el artefacto "Identificación de PI's"?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	Dejar esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. Si el requisito no tiene más PI's, finalice el seguimiento de esta lista y pase al siguiente requisito.
En el artefacto "Matriz de Requisitos de aplicación y sus PI's" ¿está el PI asociado a su requisito?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	Dejar esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. Si el requisito no tiene más PI's, finalice el seguimiento de esta lista y pase al siguiente requisito.
En el artefacto "Interacción de PI's con Roles de Usuario de la Aplicación y DQ" ¿está el PI asociado a su rol de usuario de aplicación y DQ?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	Dejar esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. Si el requisito no tiene más PI's, finalice el seguimiento de esta lista y pase al siguiente requisito.
¿Se aplicó la encuesta de características de DQ para el PI?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	Deje esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. Si el requisito no tiene más PI's, finalice el seguimiento de esta lista y pase al siguiente requisito.
En el artefacto de "Características de DQ y su importancia en roles de usuario DQ", en el PI, ¿se han identificado las características de DQ necesarias para su implementación?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	Dejar esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. Si el requisito no tiene más PI's, finalice el seguimiento de esta lista y pase al siguiente requisito.
¿Se han definido las acciones para el PI, en el artefacto de listado de acciones en PI's?	<input type="checkbox"/>	Si el requisito tiene más PI's, tome el siguiente y vuelva a la segunda interrogante de esta lista. Si el requisito no tiene más PI's, finalice el seguimiento de esta lista y pase al siguiente requisito.	<input type="checkbox"/>	Coloque el requisito en estado "Pendiente". Si el requisito tiene más PI's, tome el siguiente y vuelva a la segunda interrogante de esta lista. Si el requisito no tiene más PI's, finalice el seguimiento de esta lista y pase al siguiente requisito.

Figura 4.13: Lista de control de verificación de características DQ.

La Figura 4.13 presenta una lista de control que debe aplicarse a cada requisito especificado, en forma secuencial de acuerdo a la acción indicada por la respuesta señalada. La idea es tomar un requisito, revisar sus PI's asociados y completar su documentación si corresponde. Luego, tomar otro requisito y así sucesivamente, hasta finalizar los requisitos especificados del SI a desarrollar.

A continuación se debe hacer una revisión inversa, es decir, revisar los PI's seleccionados y verificar que al menos estén en una especificación de requisitos, esta revisión se puede apoyar con el artefacto "Identificación de PI's" y "Matriz de Requisitos de Aplicación y sus PI's", el primero por aportar la información de los PI's identificados y el segundo por proporcionar la información de los PI's cruzados con los requisitos del SI. Si hubiese algún PI no contenido en las especificaciones, debe incluirse en aquellas que correspondan de acuerdo a su definición y contexto.

Aquellos requisitos no marcados como "Pendiente", deberán marcarse como "Completo".

En esta última actividad del ciclo, es muy importante tener en cuenta que, y con el fin de evitar confusiones, solamente se revisarán aquellas especificaciones de requisitos que utilicen PI's que deban estar centrados en la DQ, dejando de lado aquellos que fueron descartados durante la aplicación de DeWIK. Como por ejemplo, los que describen especificaciones técnicas, de documentación del sistema y políticas organizacionales o externas.

Estas actividades las debe realizar el DT con supervisión del DQM.

Como resultado de la verificación de las especificaciones de requisitos y PI's del SI, las especificaciones serán clasificadas en los siguientes estados:

- **Completo**, 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 del SI.
- **Pendiente**, lo que significa que la especificación del requisito aún tiene definiciones incompletas de DQ, por lo cual debería volver a entrar en el ciclo de DeWIK.

4.2.3 Próximas Iteraciones

Es posible que al término de una iteración, probablemente producto del mismo análisis de DQ, resulten los siguientes escenarios: (a) requisitos con definiciones incompletas, es decir, con estado pendiente, (b) nuevos requisitos y (c) requisitos cuyos PI's no incluyeron DQ en la primera iteración y ahora si se les desea incorporar. Para esto es necesario iterar nuevamente el método, y repetir las distintas actividades del método partiendo de la documentación emitida en la última iteración y tratando de completar cada una de ellas de manera definitiva con las definiciones faltantes.

A continuación se explica cómo debe llevarse a cabo cada una de las actividades del método en el contexto de las próximas iteraciones de DeWIK.

a) *Especificación de Requisitos:* Se deben generar los documentos de especificación de requisitos necesarios para detallar los requisitos pendientes y actualizar aquellos casos en que se necesita modificar su definición, completando sus escenarios, sus roles de usuario y, en especial, la información de sus PI's asociados y sus características.

b) *Identificación de PI's involucrado:* En caso de no tener PI's identificados, por cada requisito, se deben seguir las listas de control de igual forma que en la primera iteración.

A continuación, debe completarse el artefacto de identificación de PI's con la nueva información de los PI's. Al igual que el documento anterior, este artefacto también puede ser completado con columnas que representen de mejor forma al PI, de acuerdo al contexto especificado del SI.

Con esta información, el DT está en condiciones de completar los artefactos "Requisitos de aplicación y sus PI's e "Interacción de PI's con Roles de Usuario de la Aplicación y DQ", de acuerdo a la definición dada en los requisitos.

c) *Definición de los aspectos de DQ:* De igual forma que en la primera iteración, en esta instancia se debe decidir a qué PI's se les definirán sus características, decisión que se puede tomar con el apoyo de los artefactos completados en esta iteración. Esta decisión lleva a centrar en la DQ aquellos requisitos, en las próximas actividades.

Para aquellos PI's en que no se han identificado sus características relevantes, se debe completar la encuesta relacionada con esta actividad. Y luego registrar sus resultados en el artefacto "Características de DQ y su importancia en Roles de Usuario DQ".

La encuesta que ya fue aplicada en la primera iteración, es un artefacto que permite ser completado con más interrogantes de acuerdo al contexto en que se desarrollarán los PI's. El tener una encuesta más refinada nos permite consultar con mayor claridad qué características de DQ y qué nivel se le debe proporcionar al PI en cuestión.

d) *Especificación de requisitos incluyendo DQ:* Se debe utilizar el mismo puntaje de corte que en la primera iteración, para mantener el ámbito de DQ que se le quiere dar al SI a desarrollar.

Se debe definir un listado de acciones por cada una de las características de DQ, asociadas a los PI's analizados. Al igual que en la primera iteración, estas acciones se pueden extraer de los ejemplos contenidos en el Anexo III, o bien incorporar otras de acuerdo a la experiencia del DT.

A continuación se deben completar las especificaciones de los requisitos con la información de sus PI's asociados y sus características de DQ seleccionadas, de acuerdo a los criterios previamente establecidos.

e) *Verificación de DQ a nivel de especificaciones*

En esta última actividad de la iteración, se debe seguir la lista de control de verificación de características de DQ para cada uno de los requisitos y PI's analizados.

Se recomienda realizar dos a tres iteraciones del método como máximo para obtener requisitos centrados en la DQ y no aumentar demasiado el costo del proyecto.

Las actividades realizadas en estas iteraciones posteriores las debe realizar el DT con la supervisión del DQM, de la misma forma que cómo se realizó en la primera iteración.

DeWIK, en esta primera versión, no ha considerado la existencia de PI's provenientes de desarrollos anteriores, con el objeto de simplificar la presentación del método, sin perjuicio de que quede abordado como trabajo futuro en aquellas actividades vinculadas con la identificación del PI.

Una vez que las especificaciones de requisitos y PI's del SI hayan alcanzado el nivel de completitud esperado, el DT 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, pero por ahora no serán abordados en nuestro trabajo.

4.3 Conclusiones

En este capítulo se ha definido el marco teórico del método DeWIQ, tomando en cuenta su objetivo general y específico.

A partir de la revisión de los estudios relacionados con abordar la DQ en las etapas de desarrollo de software, se define un método denominado DeWIQ que complementa el proceso de desarrollo de un SI, centrándolo en la DQ. Se seleccionan los elementos en los que se fundamenta el método, estos son el concepto de PI, la norma ISO/IEC 25012 y el enfoque TDQM.

DeWIQ guía a un equipo de desarrollo de software a incorporar la DQ como parte de un proceso tradicional de desarrollo de DQ, a través de un conjunto de actividades iterativas que producen como resultado un listado de acciones orientadas a centrar los requisitos del software a desarrollar en la DQ.

El conjunto de actividades que componen la implementación del método, proporciona una serie de artefactos que aportan a tener una mejor especificación de requisitos y que facilita la tarea de elaborar la documentación final relacionada con el sistema.

Capítulo 5

Casos de Estudio

En este capítulo se presenta la aplicación de DeWIK en diferentes casos de estudio. Estos casos tienen por objetivo obtener la retroalimentación necesaria, producto de la experiencia en el uso de DeWIK por parte de diferentes equipos de desarrollo, para realizar los ajustes que fueran convenientes.

Se utilizarán distintas alternativas de implementación, en que el rol de consumidor de datos o el DT defina las características de DQ a incluir, evaluando sus resultados.

Para un caso de estudio, el rol de DQM será asignado a una persona distinta a la creadora del método para evaluar los resultados de implementación y su correcta ejecución.

La información presentada en cada caso de estudio será incluida en forma resumida, procurando presentar los resultados relevantes de cada aplicación.

Posterior a ello, se incluye una sección que indica los ajustes realizados a DeWIK durante su implementación para mejorar su funcionamiento.

5.1 Caso de Estudio 1: Aplicación Web de Gestión de Información de egresados de la Carrera de Ingeniería Civil en Informática de la Universidad del Bío-Bío.

Este desarrollo forma parte de un proyecto de tesis para optar al título de Ingeniero Civil en Informática. Este proyecto consiste en el desarrollo de una aplicación Web, centrada en la calidad de datos, que permita registrar información de los egresados de la carrera de Ingeniería Civil en Informática de la Universidad del Bío-Bío, de modo que permita mantener un registro actualizado de todos ellos. Sus principales funcionalidades son:

- Permitir la creación de cuentas personales a través de las cuales los egresados puedan mantener su información actualizada.
- Contar con un rol de Administrador. Este rol poseerá privilegios que le permitirán modificar los datos de algún egresado, si esto fuese necesario.
- Permitir al Administrador generar informes estadísticos basados en la información de los egresados.
- Contar con una “Sección de Noticias y Avisos”, la cual permitirá a los usuarios, tanto ex egresados como el Administrador, la publicación de Noticias y/o Avisos de interés.
- El Sistema deberá permitir al Administrador el envío de correos masivos a los usuarios que pertenezcan a una determinada categoría, un ejemplo de categoría pueden ser: Alumnos egresados de la Carrera de Ingeniería Civil en Informática el Año 2010.

A continuación, se explica cómo fue aplicado DeWIK en este caso de estudio.

5.1.1 Etapa 1. Definición del entorno de aplicación de DeWIK

a) *Denominación y formación del DQM.* Se designa a la alumna memorista de esta tesis como DQM, no necesitó inducción por ser la autora de esta tesis.

b) *Definición de Roles.* (a) Integrante del DT, el alumno memorista es quien realiza esta tarea (b) Usuario de aplicación, se definieron los perfiles: “Egresado”, “Jefe de Carrera”, “Docente” y “WebMaster”.

c) *Entrenamiento del DT en aspectos relacionados con DQ y en el método DeWIK.* Se realizó un entrenamiento proporcionado por el DQM y lectura del Anexo II de esta tesis (documento de inducción al método). El modelo de DQ elegido fue el dado por la norma ISO/IEC 25012. Para la reclasificación de roles, todos los roles de usuario de la aplicación se comportan en función de los roles de usuario DQ consumidor y productor de datos, de acuerdo a su interacción en el SI.

5.1.2 Etapa 2. Definición de los requisitos de la aplicación con DQ

a) *Identificación de los requisitos de la aplicación.* Se utilizaron los procedimientos indicados en un desarrollo de un SI para interactuar con los usuarios y captar los requisitos y funcionalidades que ellos requieren respecto de la aplicación a desarrollar. El artefacto usado para la especificación de requisitos fueron los casos de uso. En la figura 5.1 se presenta la lista de especificación de requisitos del proyecto.

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
CU-11 Modificar Carrera

Figura 5.1: Lista de Especificación de Requisitos del C.E. N°1.

b) *Identificación de los PI's involucrados en los requisitos de la aplicación.* Se utilizó como estrategia de identificación de PI's las listas de control suministradas por el método, siguiendo sus instrucciones de aplicación.

A continuación se presenta el artefacto de PI's de la aplicación con la información relevante de algunos de los PI's identificados.

PI's de la Aplicación					
Id PI	Nombre PI	Propósito	Descripción	Roles de Usuario de la Aplicación	Requisito
FPERE	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 WebMaster Jefe de Carrera	CU-01, CU-02, CU-03
LISTEGR	Listado de Egresados	Listado de Egresados que cumplen con los criterios de filtrado aplicados.	Datos Personales Datos de Contacto Datos Laborales Datos Académicos Datos de Autenticación	Jefe de Carrera Docente WebMaster	CU-03
AV/NOT	Aviso/Noticia	Información de un Aviso/Noticia creado y enviado en el Sistema	Autor Destinatarios Titulo Contenido	Egresado Jefe de Carrera Docente WebMaster	CU-04, CU-05, CU-06
FPERD	Ficha de Perfil Docente	Ficha que posee los datos de perfil de un Docente	Datos Personales Datos de Contacto Datos Académicos Datos de Autenticación	Egresado Jefe de Carrera Docente WebMaster	CU-07, CU-08, CU-09
FICCAR	Ficha Carrera	Ficha que muestra los datos de una Carrera	Nombre Código Descripción Jefe de Carrera.	Egresado Jefe de Carrera Docente WebMaster	CU-10, CU-11

Figura 5.2: Artefacto de Identificación de PI's del C.E. N°1.

En la descripción de los PI's se colocó en forma general los datos que deben estar, pero en la práctica se deben colocar en forma detallada, es decir, los atributos que lo componen.

En la especificación de requisitos fue incorporada los PI's identificados. Luego se registró la información en el artefacto Matriz de relación de Requisitos y PI's.

Matriz de relación Requisitos y PI's					
Requisito	FPERE	LISTEGR	AV/NOT	FPERD	FICCAR
CU-01 Crear Cuenta Egresado	X				
CU-02 Editar Perfil Egresado	X				
CU-03 Filtrar Egresados	X	X			
CU-04 Crear Nuevo Aviso/Noticia			X		
CU-05 Modificar Aviso/Noticia			X		
CU-06 Eliminar Aviso/Noticia			X		
CU-07 Ingresar Nuevo Docente				X	
CU-08 Editar Datos Docente				X	
CU-09 Eliminar Docente				X	
CU-10 Agregar Nueva Carrera					X
CU-11 Modificar Carrera					X

Figura 5.3: Artefacto Matriz de relación Requisitos de aplicación y sus PI's del C.E. N°1.

A continuación se ingresa la información en el artefacto de Interacción de PI's con Roles de Usuario de la Aplicación y DQ.

Interacción de PI's con Roles de Usuario de la Aplicación y DQ			
Productos de Información	Roles de Usuario de la Aplicación	Rol de Usuario 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

Figura 5.4: Artefacto de Interacción de PI's con Roles de Usuario de la Aplicación y DQ del C.E. N°1.

Como última actividad de esta etapa, se registra la interrelación que existe en los PI's en el artefacto Matriz de Relación de PI's.

Matriz de Relación de PI's					
Productos de Información	FPERE	LISTEGR	AV/NOT	FPERD	FICCAR
FPERE		>			
LISTEGR					
AV/NOT					
FPERD					
FICCAR					

Figura 5.5: Artefacto de Matriz de Relación de PI's del C.E. N°1.

Este artefacto indica que solamente el PI Ficha de Perfil del Egresado da origen al PI de Listado, en los demás PI's no hay interrelación.

c) *Definición de aspectos de DQ.* Como alternativa de implementación de DQ se decide excluir aquellas características de DQ que no son aplicadas de acuerdo al SI a desarrollar, en base a los conocimientos adquiridos por el DT y el DQM del SI y sus experiencias previas. Se debe consultar al usuario por el resto de las características de DQ. Se adaptó el patrón de encuesta ubicado en el Anexo IV y se aplicó dicho instrumento a los usuarios pertenecientes a los roles de usuario DQ consumidor y productor datos. A continuación se muestran algunos ejemplos de PI's encuestados.

PI FPERE, Ficha Perfil Egresado			
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
PI AV/NOT, Aviso/Noticia			
Características de DQ	Consumidor de Datos	Productor de Datos	Ponderación Promedio
Exactitud	5	N/A	5
Complejidad	5	N/A	5
Consistencia	1	N/A	1
Credibilidad	4	N/A	4
Actualidad	5	N/A	5
Accesibilidad	5	5	5
Conformidad	1	1	1
Confidencialidad	5	5	5
Eficiencia	5	5	5
Precisión	4	4	4
Trazabilidad	5	5	5
Comprensibilidad	4	4	4
Disponibilidad	1	1	1
Portabilidad	5	5	5
Recuperabilidad	1	1	1
PI FPERD, Ficha de Perfil Docente			
Características de DQ	Consumidor de Datos	Productor de Datos	Ponderación Promedio
Exactitud	4	N/A	4
Complejidad	4	N/A	4
Consistencia	5	N/A	5
Credibilidad	1	N/A	1
Actualidad	1	N/A	1
Accesibilidad	4	4	4
Conformidad	1	1	1
Confidencialidad	1	1	1
Eficiencia	1	1	1
Precisión	3	3	3
Trazabilidad	1	1	1
Comprensibilidad	4	5	4
Disponibilidad	1	1	1
Portabilidad	5	5	5
Recuperabilidad	1	1	1

Figura 5.6: Artefacto de Características de DQ y su importancia en Roles de Usuario DQ por PI del C.E. N°1.

d) *Especificación de requisitos incluyendo DQ.* Se analizaron los valores de ponderación promedio y se eligieron las características que tuvieron puntuación mayor o igual al puntaje de

corte definido igual a 4. Se definieron acciones para las características de DQ definidas (ver Figura 5.7), tomando como referencia las acciones tipo contenidas en el Anexo III de esta tesis.

Listado de acciones en PI's		
PI	Características de DQ	Acciones
AV/NOT	Exactitud	<ol style="list-style-type: none"> 1. Para dar la opción al usuario de crear un aviso o una noticia, el Sistema debe mostrar esos dos valores (aviso, noticia) en un radio button. 2. Definir reglas de validación para el ingreso de datos (validación sintáctica). 3. Definir largos para los datos que se deben ingresar. Este largo será definido de acuerdo al tipo de dato especificado en el modelo de datos. 4. Antes de enviar el Aviso/Noticia, se proporcionará una vista previa al usuario.
	Complejidad	<ol style="list-style-type: none"> 1. Al momento de mostrar el formulario de ingreso de un Aviso/Noticia, marcar los datos que deben ser ingresados de forma obligatoria. 2. Validar el ingreso de los datos. 3. No enviar un Aviso/Noticia hasta asegurar que esté completo. 4. 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.
	Actualidad	<ol style="list-style-type: none"> 1. Al momento de ingresar un nuevo Aviso/Noticia, el Sistema debe registrar la fecha y la hora de la creación. 2. Incorporar estados a los Avisos/Noticias que se envían en el Sistema (Activo, Eliminado). 3. Hacer mantención a los registros de Avisos/Noticias, para eliminar aquellos en los cuales la fecha de caducidad (señalada por el usuario) haya vencido.
	Accesibilidad	<ol style="list-style-type: none"> 1. Mostrar los Avisos/Noticias ordenados cronológicamente. 2. Permitir al usuario buscar los Avisos/Noticias separados por: Avisos/Noticias enviados y Avisos/Noticias recibidos.
	Confidencialidad	<ol style="list-style-type: none"> 1. Solo los usuarios autorizados, entiéndase Autor y Remitente(s) del Aviso/Noticia, podrán ver estos.
	Eficiencia	<ol style="list-style-type: none"> 1. Para la salida de información, se usarán formatos livianos, tales como txt y pdf. 2. El modelo de datos debe ser normalizado.
	Trazabilidad	<ol style="list-style-type: none"> 1. El Aviso/Noticia debe tener claramente señalado quien es su autor. 2. El Aviso/Noticia debe registrar la fecha y la hora de su creación. 3. La bitácora del Sistema deberá poseer todos los ingresos y modificaciones a un Aviso/Noticia, indicando claramente el autor, la fecha y la hora.
	Portabilidad	<ol style="list-style-type: none"> 1. Dar la opción al usuario de exportar un Aviso/Noticia, los formatos en los que se permitirá exportar serán pdf y txt.
FPERE	Exactitud	<ol style="list-style-type: none"> 1. Definir reglas de validación para los datos a ingresar por el usuario (validación sintáctica). Por ejemplo: Que el teléfono sea un dato numérico, que el mail tenga el formato adecuado, al igual que el Rut, etc. 2. Definir largos para los datos que se deben ingresar. Este largo será definido de acuerdo al tipo de dato especificado en el modelo de datos 3. Antes de crear la Ficha de Perfil del Egresado, el Sistema proporcionara una vista previa de esta al usuario. 4. Los datos que posean valores definidos (Ejemplo año de egreso, carrera de egreso), pueden facilitar su ingreso presentándose en un select, o como se conoce comúnmente, un combo box. 5. Para los datos que solo posean dos valores posibles (Ejemplo, sexo), los valores deben presentarse en un radio button.
	Complejidad	<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.
	Accesibilidad	<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).
	Portabilidad	<ol style="list-style-type: none"> 1. Dar la opción de exportar el Perfil de un Egresado, en formatos pdf y txt.

Figura 5.7: Ejemplo de listado de algunas acciones en PI's del C.E. N°1.

A continuación, se complementó la especificación de requisitos con las características de DQ consideradas relevantes. A continuación se presentan algunos ejemplos.

Caso de Uso: Crear Cuenta Egresado
ID: CU-01
Breve Descripción: Este Caso de Uso permite al actor, crear un perfil en el Sistema, este perfil poseerá sus datos personales, sus datos académicos, sus datos laborales y los datos de Autenticación para la cuenta.
Actores Principales: Egresado
Actores Secundarios: No existen actores secundarios
Productos de Información: 1. <i>Ficha de Perfil Egresado:</i> Ficha que posee los datos de perfil de un Egresado (Datos Personales, Datos de Contacto, Datos Laborales, Datos Académicos y Datos de Ingreso al Sistema). <ul style="list-style-type: none"> • <u>Características de Calidad de Datos:</u> 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 esta Condiciones Y Reglas, y procede con el registro. 4. El Sistema despliega el formulario de registro en el cual el actor debe ingresar sus datos. El formulario de registro 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: <ul style="list-style-type: none"> • 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 la cual 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 todos 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, con los datos solicitados. 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 o incorrectos. 7.2. El Caso de Uso retorna al paso 4.

Figura 5.8: Caso de Uso Crear Cuenta Egresado del C.E. N°1.

Caso de Uso: Crear Nuevo Aviso/Noticia
ID: CU-04
Breve Descripción: Este Caso de Uso permite al actor, la creación de un nuevo Aviso y/o Noticia en el Sistema. Cada nueva noticia posee un título, un detalle, un autor y el (los) remitente(s).
Actores Principales: Egresado, Jefe de Carrera, Docente, Secretaria y WebMaster.
Actores Secundarios: No existen actores secundarios
Productos de Información: 1. <u>Aviso/Noticia</u> : Información de un Aviso/Noticia creado y enviado en el Sistema. La información que posee un Aviso/Noticia es: Autor, Destinatarios y Contenido. <ul style="list-style-type: none"> • <u>Características de Calidad de Datos</u>: Exactitud, Completitud, Actualidad, Accesibilidad, Confidencialidad, Eficiencia, Trazabilidad y Portabilidad.
Precondiciones: 1. El actor debe estar autenticado previamente en el Sistema. 2. Los remitentes del mensaje deben poseer un perfil creado en el Sistema.
Flujo Principal: 1. El Caso de Uso comienza cuando el actor selecciona la opción “Crear Nuevo Aviso/Noticia”. 2. El Sistema muestra un formulario con los detalles del nuevo Aviso/Noticia a crear, los datos que el Sistema solicita al actor son los siguientes: <ul style="list-style-type: none"> • Título • Contenido del Aviso/Noticia • Destinatarios 3. El actor ingresa los datos solicitados al formulario de creación de Aviso/Noticia. 4. El Sistema valida los datos ingresados al formulario 5. Si los datos son válidos, el Sistema crea el nuevo Aviso/Noticia con el estado de “Activo” y lo remite a el(los) remitente(s) seleccionados. 6. El Sistema registra la creación del nuevo Aviso/Noticia en el Log de Actividades al cual corresponde. 7. El Caso de Uso concluye, con el nuevo Aviso/Noticia creado y enviado a los remitentes
Post-Condiciones: 1. El nuevo Aviso/Noticia se crea y es enviado a él (los) remitente(s) del mismo. 2. La creación de un nuevo Aviso/Noticia es registrado en el Log de Actividades 3. El estado del Aviso/Noticia se asignará a “Activo”.
Flujos Alternativos: 5.1 Si los datos ingresados no son válidos, el Sistema avisa de esto al actor, el cual debe revisarlos y reingresarlos. 5.2 Se retorna al paso 2.

Figura 5.9: Caso de Uso Crear Nuevo Aviso/Noticia del C.E. N°1.

e) *Verificación de DQ a nivel de especificaciones.* En esta última actividad del ciclo, se realizó el seguimiento de la lista de control útil para verificar que las especificaciones contengan las características de DQ. A continuación se muestra un ejemplo de este seguimiento.

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 PI's 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, es por ello que se determina concluir esta primera fase de desarrollo del SI.

5.1.3 Análisis de la aplicación

Este caso de estudio nos deja las siguientes experiencias:

- 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 DQM 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.

5.2 Casos de Estudio 2: Sistema de Gestión Administrativa para Grupo de Robótica de la Universidad del Bío-Bío, sede Chillán, incorporando Calidad de Datos.

Este desarrollo forma parte de un proyecto de tesis para optar al título de Ingeniero Civil en Informática. Este proyecto consiste en el desarrollo de una aplicación web para el registro y control de la asistencia y el pago de cuotas de los integrantes del Grupo Robótica de la Universidad del Bío-Bío, sede Chillán, incorporando calidad de datos. Sus principales funcionalidades son:

- Llevar registro de ingresos y egresos de los miembros del grupo, particularmente de las cuotas pagadas e impagas.
- Llevar registro de los eventos, asistencia y justificaciones de los miembros. El ingreso de asistencia debe ser en un sitio web y un equipo móvil con sistema operativo Android.
- Generar listados que indiquen cuotas pagadas y/o vencidas por miembro, dineros disponibles / adeudados y asistencia / inasistencia a los eventos por parte de los miembros.
- Generar avisos mediante correo electrónico por el vencimiento de una cuota, indicando deudas anteriores si estas existiesen y cuando un miembro exceda un nivel preestablecido de deuda, así como de inasistencias.

A continuación se presenta la aplicación del método con sus actividades.

5.5 Etapa 1. Definición del entorno de aplicación de DeWIQ

a) *Denominación y formación del DQM.* Se designa a una persona del equipo de investigación como DQM, se realizó una inducción del método a cargo de la alumna memorista de esta tesis.

b) *Definición de Roles.* (a) Integrante del DT, el alumno memorista quien realiza esta tarea y (b) Usuario de aplicación, se definieron los perfiles: “Docente Responsable”, “Miembro del Grupo”, “Directiva”, “Tesorero”.

c) *Entrenamiento del DT en aspectos relacionados con DQ y en el método DeWIQ.* Se realizó un entrenamiento proporcionado por el DQM y lectura del Anexo II de esta tesis. El modelo de DQ elegido fue el dado por la norma ISO/IEC 25012. Para la reclasificación de roles, todos los roles de usuario de la aplicación se comportan en función de los roles de usuario DQ consumidor y productor de datos, de acuerdo a su interacción en el SI.

5.2.2 Etapa 2. Definición de los requisitos de la aplicación con DQ

a) *Identificación de los requisitos de la aplicación.* Se utilizaron los procedimientos indicados en un desarrollo de un SI para interactuar con los usuarios y captar los requisitos y funcionalidades que ellos requieren respecto de la aplicación a desarrollar. El artefacto usado para la especificación de requisitos fueron los casos de uso. En la figura 5.10 se presenta la lista especificación de requisitos del proyecto.

Requisito
CU-01 Agregar Miembro
CU-02 Editar Miembro
CU-03 Deshabilitar Miembro
CU-04 Habilitar Miembro
CU-05 Revisar Directiva Actual
CU-06 Agregar Evento
CU-07 Tomar Asistencia
CU-08 Revisar Estado de Cuentas
CU-09 Revisar Ingresos
CU-10 Revisar Egresos

Figura 5.10: Lista de Especificación de Requisitos del C.E. N°2.

b) *Identificación de los PI's involucrados en los requisitos de la aplicación.* Se utilizó como estrategia de identificación de PI's las listas de control suministradas por el método, siguiendo sus instrucciones de aplicación.

A continuación se presenta el artefacto de PI's de la aplicación con la información relevante de algunos de los PI's identificados.

PI's de la Aplicación					
Id. PI	Nombre PI	Propósito	Descripción	Roles de Usuario de la aplicación	Requisito
PI-1	Miembro Robótica	Individuo que desempeña el rol de miembro del grupo robótica.	Run, Nombres, Apellidos, Rol, Correo electrónico, Teléfono	Docente Responsable, Miembro del Grupo.	CU-01, CU-02, CU-03, CU-04, CU-05
PI-2	Informe Directiva Actual	Listado de nombres de las personas que ocupan los cargos del grupo, estos son; Presidente, Vicepresidente, Secretario y Tesorero.	Nombre y Run de cada cargo, Periodo designado.	Docente Responsable, Miembro del Grupo.	CU-05
PI-3	Formulario de designación de Directiva	Permite seleccionar de entre los miembros vigentes del grupo a quienes desempeñarán los cargos de Presidente, Vicepresidente, Secretario y Tesorero.	Nombre y Run de cada cargo.	Docente Responsable	
PI-4	Evento	Corresponde a un suceso agendado, relevante para el grupo y sujeto a la toma de asistencia de los miembros del grupo.	Título, Fecha, Hora Lugar, Descripción	Docente Responsable, Miembro del grupo.	CU-06, CU-07
PI-5	Informe de Estado de Miembros	Permite al Docente Responsable y a la Directiva conocer el estado de sus miembros vigentes, ver situación de pagos de cuotas y asistencias a eventos.	Nombre y Run del miembro activo, Porcentaje de asistencia del miembro activo Estado de pago de cuotas.	Docente Responsable, Directiva.	
PI-6	Informe de asistencia a eventos	Otorga una visión general de todos los eventos del año, indicando el porcentaje de asistencia a cada uno de ellos.	Título, fecha y porcentaje de asistencia de cada evento.	Docente Responsable, Directiva.	
PI-7	Informe de estado de cuenta	Permite conocer cuál es el monto de dinero que el grupo posee actualmente.	Total de dinero recibido, Total de dinero gastado, Diferencia entre recibido y gastado.	Docente Responsable, Tesorero.	CU-08
PI-8	Ingreso de dinero	Permite al tesorero registrar el pago de cuotas por parte de un miembro.	Miembro del grupo (run y nombre), Monto pagado.	Tesorero.	CU-10
PI-9	Egreso de dinero	Permite al tesorero registrar el gasto de dinero del grupo.	Monto del egreso, fecha y descripción de qué fue cancelado o adquirido.	Tesorero	CU-09

Figura 5.11: Artefacto de Identificación de PI's del C.E. N°2.

En la especificación de requisitos fueron incorporados los PI's identificados. Luego se registró la información en el artefacto Matriz de relación de Requisitos y PI's.

Matriz de relación Requisitos y PI's									
Requisitos	PI-1	PI-2	PI-3	PI-4	PI-5	PI-6	PI-7	PI-8	PI-9
CU-01 Agregar Miembro	X								
CU-02 Editar Miembro	X								
CU-03 Deshabilitar Miembro	X								
CU-04 Habilitar Miembro	X								
CU-05 Revisar Directiva Actual	X	X							
CU-06 Agregar Evento				X					
CU-07 Tomar Asistencia	X			X					
CU-08 Revisar Estado de Cuentas							X		
CU-09 Revisar Egresos									X
CU-10 Revisar Ingresos	X							X	

Figura 5.12: Artefacto Matriz de relación Requisitos de aplicación y sus PI's del C.E. N°2.

A continuación se ingresó la información en el artefacto de Interacción de PI's con Roles de Usuario de la Aplicación y DQ.

Interacción de PI's con Roles de Usuario de la Aplicación y DQ			
Productos de Información	Roles de Usuario de la Aplicación	Rol de usuario DQ	
		Consumidor de Datos	Productor de Datos
PI-1	Docente Responsable	X	X
	Directiva	X	X
	Tesorero	X	
	Miembro del Grupo	X	X
PI-2	Docente Responsable	X	
	Directiva	X	
	Tesorero	X	
	Miembro del Grupo	X	
PI-3	Docente Responsable		X
PI-4	Docente Responsable	X	X
	Directiva	X	X
	Tesorero	X	
	Miembro del Grupo	X	
PI-5	Docente Responsable	X	
	Directiva	X	
	Tesorero	X	
	Directiva	X	
	Tesorero	X	
PI-6	Docente Responsable	X	
	Directiva	X	
	Tesorero	X	
PI-7	Docente Responsable	X	
	Tesorero	X	
PI-8	Tesorero		X
PI-9	Tesorero		X

Figura 5.13: Artefacto de Interacción de PI's con Roles de Usuario de la Aplicación y DQ del C.E. N°2.

Como última actividad de esta etapa, se registró la interrelación que existe entre los PI's en el artefacto Matriz de Relación de PI's.

Matriz de Relación de PI's									
Productos de Información	PI-1	PI-2	PI-3	PI-4	PI-5	PI-6	PI-7	PI-8	PI-9
PI-1		>	>		>	>		>	
PI-2									
PI-3		>							
PI-4						>			
PI-5									
PI-6									
PI-7									
PI-8					>		>		
PI-9							>		

Figura 5.14: Artefacto de Matriz de Relación de PI's del C.E. N°2.

c) *Definición de aspectos de DQ.* Como alternativa de implementación de DQ, se decide consultar por las 15 características al usuario. Se adaptó el patrón de encuesta ubicado en el Anexo IV y se aplicó dicho instrumentos a los usuarios pertenecientes a los roles de usuario DQ consumidor y productor datos. A continuación se muestran algunos ejemplos de PI's encuestados.

PI 1, Miembro Robótica			
Característica de DQ	Consumidor de Datos	Productor de datos	Ponderación Promedio
Exactitud	4	N/A	4
Compleitud	5	N/A	5
Consistencia	5	N/A	5
Credibilidad	5	N/A	5
Actualidad	1	N/A	1
Accesibilidad	5	5	5
Conformidad	5	5	5
Confidencialidad	2	2	2
Eficiencia	5	5	5
Precisión	2	2	2
Trazabilidad	5	5	5
Comprensibilidad	4	4	4
Disponibilidad	1	1	1
Portabilidad	5	5	5
Recuperabilidad	3	5	5
PI 2, Informe Directiva Actual			
Característica de DQ	Consumidor de Datos	Productor de datos	Ponderación Promedio
Exactitud	3	N/A	3
Compleitud	5	N/A	5
Consistencia	4	N/A	4
Credibilidad	4	N/A	4
Actualidad	1	N/A	1
Accesibilidad	5	5	5
Conformidad	5	5	5
Confidencialidad	2	2	2
Eficiencia	5	5	5
Precisión	2	2	2
Trazabilidad	5	5	5
Comprensibilidad	3	3	3
Disponibilidad	1	1	1
Portabilidad	4	4	4
Recuperabilidad	4	4	4

Figura 5.15: Artefacto de Características de DQ y su importancia en Roles de Usuario DQ por PI del C.E. N°2

d) *Especificación de requisitos incluyendo DQ.* Se analizaron los valores de ponderación promedio y se eligieron las características que tuvieron puntuación mayor o igual al puntaje de corte definido igual a 3. Se definieron acciones para las características de DQ definidas, tomando como referencia las acciones tipo contenidas en el Anexo III de esta tesis. A continuación se presentan algunos ejemplos.

Listado de acciones en PI's		
PI	Característica de DQ	Acciones
PI-1, PI-5	Complejidad	1. Validar que el formulario esté completo antes de guardar. 2. Validar los datos ingresados, para tipos de datos de texto una longitud mínima de 3 caracteres, para datos numéricos validar que solo existan números, para el caso de direcciones de correo, fechas u horas utilizar expresiones regulares. 3. Establecer niveles de ingreso, obligatorios y no obligatorios.
	Consistencia	1. Se debe prohibir el ingreso de dos (o más) miembros con un mismo RUT. 2. Validar que un miembro no registre deudas posteriores al periodo de su primera incorporación.
	Accesibilidad	1.No se aplican funcionalidades especiales para beneficiar a personas con deficiencias psicomotoras. 2.Cada listado debe contar con la opción “buscar”, mediante cualquier palabra (o parte de ella).
	Precisión	1. Se debe definir Restricción de clave primaria en cada una de las tablas del modelo relacional involucradas.
	Comprensibilidad	1. Cada informe debe poseer el logo de la Universidad y del Grupo, el título, el periodo (si es el caso), la fecha en la cual fue generado y página actual y total. 2. Cada formulario de ingreso debe exponer que es lo que se está ingresando, y los campos que son obligatorios. 3. Utilizar errores que indiquen el nombre del campo y la descripción del error. 4. Entregar etiquetados de campos significativos.
	Recuperabilidad	1.Utilizar solo eliminación lógica para los miembros. 2.Al recargar un formulario de ingreso que posea errores, mantener los campos que se encuentren correctos.
	Exactitud	1. Los campos RUT y correo electrónico deben ser verificados en su forma y validados (sintaxis). 2. Se debe permitir ingresar campos del tipo fecha u hora sólo con un formato válido. 3. Los informes deben contener paginación, indicando la página actual y total.
	Confidencialidad	1. Se deben establecer cuatro perfiles de usuario, Docente coordinador, miembro de directiva, tesorero y miembro del grupo. 2. Solo el Docente coordinador puede agregar nuevos miembros al grupo. 3. Debe existir la posibilidad de cerrar sesión. 4. La sesión debe caducar transcurridos 30 minutos desde la última actividad sobre el sistema. 5. Las contraseñas de los usuarios del sistema deben ser ocultadas mediante el sistema de encriptación MD5. 6. Cada usuario del sistema puede ver solo la información propia (excepto el docente coordinador y la directiva, los cuales pueden ver la información de todos).
PI-2, PI-3	Complejidad	1. Ayudar a la selección de un sujeto para un cargo mediante la selección de este por sobre la escritura. 2. No guardar la designación sin haber escogido a un miembro para el cargo en cuestión. 3. Indicar en el formulario que el cargo no ha sido establecido (cuando corresponda)
	Conformidad	1. Se debe avisar mediante confirmaciones que la operación concluyó exitosamente.
	Credibilidad	1. Los datos incluidos en los informes deben ser los mismos que los expuestos en el sistema en línea. 2. Los datos ingresados en el formulario de ingreso deben proveer una vista previa. 3. Solo el docente puede definir los cargos.

Figura 5.16: Listado de acciones en PI's del C.E. N°2.

A continuación, se complementa la especificación de requisitos con las características de DQ consideradas relevantes. A continuación se presentan algunos ejemplos.

Caso de Uso: Agregar Miembro
ID: CU01
Breve Descripción: Permite incorporar un miembro al sistema, y habilitarlo como parte del Grupo Robótica quedando sujeto al cobro de cuotas y la toma de asistencia a los eventos.
Actores Principales: Docente Responsable.
Actores Secundarios: Miembro del Grupo.
Productos de Información (PI's): 1. <i>Miembro Robótica,</i> Es la representación abstracta de una persona que desempeña el rol de miembro del grupo robótica Características de DQ: Completitud, Consistencia, Credibilidad, Accesibilidad, Conformidad, Precisión, Comprensibilidad, Recuperabilidad, Disponibilidad, Exactitud, Confidencialidad.
Precondiciones: 1. El usuario debe haber completado el caso de uso “Autenticar Usuario” durante la presente sesión. 2. El usuario debe poseer el rol de Docente Responsable.
Flujo Principal: 1. El Caso de Uso comienza cuando el Usuario selecciona la opción de agregar un nuevo miembro. 2. El Sistema despliega un formulario donde se debe especificar obligatoriamente los siguientes datos del miembro: run, nombres, apellidos y correo electrónico. 3. El Usuario ingresa los datos solicitados. 4. El Sistema almacena los datos y envía un correo electrónico al nuevo miembro informándole de su inscripción y de su clave de acceso.
Postcondiciones: 1. El Sistema mantiene un registro del nuevo miembro.
Flujos Alternativos: 3. A. El Usuario no ingresa todos los datos solicitados. 4. A. El Sistema vuelve al punto 2, indicando que debe ingresar todos los datos.

Figura 5.17: Caso de Uso de Agregar Miembro del C.E. N°2.

Caso de Uso: Revisar Directiva Actual
ID: CU05
Breve Descripción: Permite conocer cuáles son los miembros que actualmente dirigen el grupo.
Actores Principales: Docente Responsable o Miembro del Grupo.
Actores Secundarios: No tiene.
Productos de Información (PI's): 1. <i>Miembro Robótica,</i> Es la representación abstracta de una persona que desempeña el rol de miembro del grupo robótica Características de DQ: Completitud, Consistencia, Credibilidad, Accesibilidad, Conformidad, Precisión, Comprensibilidad, Recuperabilidad, Disponibilidad, Exactitud, Confidencialidad. 2. <i>Informe de Directiva Actual,</i> Corresponde al nombre de cada uno de las personas que ocupan los cargos del grupo, estos son; Presidente, Vicepresidente, Secretario y Tesorero. Características de DQ: Completitud, Accesibilidad, Conformidad, Precisión, Comprensibilidad, Consistencia, Credibilidad, Recuperabilidad, Disponibilidad.
Precondiciones: 1. El usuario debe haber completado el caso de uso “Autenticar Usuario” durante la presente sesión. 2. El usuario debe poseer el rol de Docente Responsable o de Directiva.
Flujo Principal: 1. El Caso de Uso comienza cuando el Usuario selecciona la opción para ver la directiva del grupo. 2. El Sistema muestra los nombres de la directiva, y sus cargos respectivos.
Postcondiciones: No tiene.
Flujos Alternativos: No tiene.

Figura 5.18: Caso de Uso de Revisar Directiva Actual del C.E. N°2.

Caso de Uso: Agregar Evento
ID: CU07
Breve Descripción: Permite programar un evento en una fecha determinada.
Actores Principales: Docente Responsable o Directiva.
Actores Secundarios: No tiene.
Productos de Información (PI's): 1. <i>Evento</i> , Corresponde a un suceso agendado el cual es relevante para el grupo y que está sujeto a la toma de asistencia hacia los miembros del grupo Características de DQ: Completitud, Consistencia, Credibilidad, Accesibilidad, Conformidad, Precisión, Comprensibilidad, Disponibilidad, Exactitud, Recuperabilidad.
Precondiciones: 1. El usuario debe haber completado el caso de uso "Autenticar Usuario" durante la presente sesión. 2. El usuario debe poseer el rol de Docente Responsable o de Directiva.
Flujo Principal: 1. El Caso de Uso comienza cuando el Usuario selecciona la opción de agregar un nuevo evento. 2. El Sistema despliega un formulario donde se debe especificar obligatoriamente los siguientes datos del evento: título, fecha, hora, lugar y descripción. 3. El Usuario ingresa los datos solicitados. 4. El Sistema almacena los datos.
Postcondiciones: 1. El Sistema mantiene un registro del evento.
Flujos Alternativos: 3. A. El Usuario no ingresa todos los datos solicitados. 4. A. El Sistema vuelve al punto 2, indicando que debe ingresar todos los datos.

Figura 5.19: Caso de Uso de Agregar Evento del C.E. N°2.

e) *Verificación de DQ a nivel de especificaciones.* En esta última actividad del ciclo, se realizó el seguimiento de la lista de control útil para verificar que las especificaciones contengan las características de DQ. A continuación se muestra un ejemplo de este seguimiento.

Como resultado de esta actividad, todas las especificaciones de requisitos quedaron centradas en la DQ, es por ello que se determina cerrar esta primera fase de desarrollo del SI.

5.2.3 Análisis de la aplicación

Este caso de estudio nos deja las siguientes experiencias:

- Al ser dirigida esta implementación por un DQM distinto al creador del método, se quiso probar que su gestión debe ser suficiente con la inducción realizada, tarea correspondiente al comienzo de esta implementación, y con el material proporcionado para esta labor. El objetivo de esta prueba fue cumplido.
- La lista de control para identificar PI's fue enriquecida con nuevas interrogantes para clasificar la información de salida de los requisitos en productos de información.

5.3 Casos de Estudio 3: Gestor de Pacientes de un Nutricionista

Este desarrollo forma parte de un proyecto de tesis para optar al título de Ingeniería Civil en informática. Este proyecto consiste en construir un sistema de gestión de fichas de pacientes que permitirá tener de manera eficiente toda la información del paciente, de su diagnóstico inicial y de los controles anteriormente realizados o derivación de algún médico tratante. Las funcionalidades principales de este sistema son:

- Búsquedas de fichas de pacientes por medio de filtros relacionados con datos del paciente.
- Generación de una dieta, controlando un máximo de Hidratos de Carbono ingeridos. Para esta información, proporciona una tabla de comidas clasificadas por grupo alimenticio.
- Generación de minutas, con información clara para el paciente que incluya los alimentos que puede consumir y el aporte de los nutrientes que cada uno de ellos.

A continuación se describe la aplicación de DeWIQ en este caso de estudio

5.3.1 Etapa 1. Definición del entorno de aplicación de DeWIQ

a) *Denominación y formación del DQM.* Se designa a Alejandra Fuentes como DQM, no fue necesario realizar una inducción por ser una de las creadoras del método.

b) *Definición de Roles.* (a) Integrante del DT, la alumna tesista quien realiza esta tarea y (b) Usuario de aplicación, se definió el perfil: “Nutricionista”.

c) *Entrenamiento del DT en aspectos relacionados con DQ y en el método DeWIQ.* Se realizó un entrenamiento proporcionado por el DQM y lectura del Anexo II de esta tesis. El modelo de DQ elegido fue el dado por la norma ISO/IEC 25012. Para la reclasificación de roles, el rol de usuario de la aplicación se comportan en función de los roles de usuario DQ consumidor y productor de datos, de acuerdo a su interacción en el SI.

5.3.2 Etapa 2. Definición de los requisitos de la aplicación con DQ

a) *Identificación de los requisitos de la aplicación.* Se utilizaron los procedimientos indicados en un desarrollo de un SI para interactuar con los usuarios y captar los requisitos y funcionalidades que ellos requieren respecto de la aplicación a desarrollar. El artefacto usado para la especificación de requisitos fueron los casos de uso. En la Figura 5.22 se presenta la especificación de requisitos del proyecto.

Requisito
CU-01 Controlar Paciente Nuevo
CU-02 Buscar Paciente
CU-03 Crear Nuevo Tratamiento
CU-04 Ingresar Datos Personales
CU-05 Ingresar Historial Clínico
CU-06 Obtener Tratamiento
CU-07 Ingresar Control Paciente
CU-08 Ingresar Dieta
CU-09 Evolución del Tratamiento Nutricional
CU-10 Evaluación del Paciente por Patología

Figura 5.20: Listado de Especificación de Requisitos del C.E. N°3.

b) *Identificación de los PI's involucrados en los requisitos de la aplicación.* Se utilizó como estrategia de identificación de PI's las listas de control suministradas por el método, siguiendo sus instrucciones de aplicación.

A continuación se presenta el artefacto de PI's de la aplicación con la información relevante de algunos de los PI's identificados.

PI's de la Aplicación					
Id PI	Nombre PI	Propósito	Descripción	Roles de Usuario de la Aplicación	Requisito
FICHA	Ficha de Paciente	Contiene los datos personales del paciente.	Nombre, apellido paterno y materno, fecha de nacimiento, situación familiar, n° hijos, correo electrónico, teléfono, ciudad y sexo.	Nutricionista	CU-01, CU-02, CU-04, CU-07
HIST	Historial Clínico	Contiene los antecedentes de enfermedades y medicamentos del paciente.	Hábitos de tabaco y alcohol, actividad física, tipo de paciente, antecedentes familiares, nombre especialista, especialidad, diagnóstico inicial, medicamento, dosis y encuesta nutricional (Horario, comida y porción)	Nutricionista	CU-01, CU-02, CU-03, CU-05, CU-07
LPAT	Listado de Patologías	Listado de Patologías	Nombre de patología	Nutricionista	CU-04, CU-10
CONT	Control de Paciente	Contiene el peso, la talla, Contorno de Cintura (CC) y estado nutricional del paciente.	Fecha, peso, talla, Contorno de Cintura, IMC, diferencias de peso (primer control y control anterior), estado nutricional del paciente, observaciones y HC.	Nutricionista	CU-01, CU-03, CU-07, CU-08
DIETA	Dieta	Contiene las cantidades de H de C por comidas del día.	Cantidad total de HC y distribución de HC en una tabla por comidas y grupo de alimentos.	Nutricionista	CU-08
IEP	Informe de Evaluación del paciente por Patologías	Informe que indica las patologías de los pacientes según selección de enfermedades y sexo.	Rut, nombre, apellido paterno, peso inicial, último peso y peso ideal	Nutricionista	CU-10
IETN	Informe de Evolución del Tratamiento Nutricional	Informe que indica la evolución de los pacientes en tratamiento o dados de alta, según el periodo de tiempo de su tratamiento.	Rut, nombre, apellido paterno, peso inicial, último peso y Diferencia de peso (peso inicial – último peso).	Nutricionista	CU-09
TRATA	Tratamiento	Contiene los datos asociados a un tratamiento.	Estado, fecha inicio y de último control realizado a un paciente, controles, historial clínico.	Nutricionista	CU-01, CU-03
LTRA	Listado de Tratamiento	Listado de Tratamiento	Estado, fecha inicio y de último control realizado a un paciente	Nutricionista	CU-02, CU-06

Figura 5.21: Artefacto de Identificación de PI's del C.E. N°3.

En la especificación de requisitos fueron incorporados los PI's identificados. Luego se registró la información en el artefacto Matriz de relación de Requisitos y PI's.

Matriz de relación Requisitos y PI's									
Requisitos	FICHA	HIST	LPAT	CONT	DIETA	IEP	IETN	TRATA	LTRA
CU-01 Controlar Paciente Nuevo	X	X		X				X	
CU-02 Buscar Paciente	X	X							X
CU-03 Crear Nuevo Tratamiento		X		X				X	
CU-04 Ingresar Datos Personales	X		X						
CU-05 Ingresar Historial Clínico		X							
CU-06 Obtener Tratamiento									X
CU-07 Ingresar Control Paciente	X	X		X					
CU-08 Ingresar Dieta				X	X				
CU-09 Evolución del Tratamiento Nutricional							X		
CU-10 Evaluación del Paciente por Patologías			X			X			

Figura 5.22: Artefacto Matriz de relación Requisitos de aplicación y sus PI's del C.E. N°3.

Para la Interacción de PI's con los Roles de Usuario de la Aplicación y DQ se determina que el único rol de usuario de la aplicación se comporta como productor y consumidor de datos por ser el único individuo que interactúa con el SI's y consecuentemente con sus PI's definidos.

Como última actividad de esta etapa, se registra la interrelación que existe en los PI's en el artefacto Matriz de Relación de Requisitos y PI's.

Matriz de Relación de PI's									
Productos de Información	FICHA	HIST	LPAT	CONT	DIETA	IEP	IETN	TRATA	LTRA
FICHA						>	>		
HIST								>	
LPAT		>				>			
CONT						>	>	>	>
DIETA				>					
IEP									
IETN									>
TRATA									
LTRA									

Figura 5.23: Matriz de relación Requisitos y PI's del C.E. N°3.

c) *Definición de aspectos de DQ.* Se implementó como alternativa de inclusión de características de DQ en que el DT seleccionara aquellas que considerara necesarias de incluir, que no incorpore aquellas características que no son necesarias y que consulte al usuario aquellas

características en que no le es clara su inclusión. Se adaptó el patrón de encuesta ubicado en el Anexo IV y se aplicó dicho instrumento al usuario de este sistema. A continuación se muestran algunos ejemplos de PI's encuestados.

PI HIST, Historial Clínico			
Características de DQ	Consumidor de Datos	Productor de Datos	Ponderación Promedio
Exactitud	0	N/A	0
Compleitud	5	N/A	5
Consistencia	5	N/A	5
Credibilidad	4	N/A	4
Actualidad	0	N/A	0
Accesibilidad	5	5	5
Conformidad	0	0	0
Confidencialidad	0	0	0
Eficiencia	5	5	5
Precisión	0	0	0
Trazabilidad	0	0	0
Comprensibilidad	5	5	5
Disponibilidad	5	5	5
Portabilidad	0	0	0
Recuperabilidad	0	0	0
PI CONT, Control de Paciente			
Características de DQ	Consumidor de Datos	Productor de Datos	Ponderación Promedio
Exactitud	5	-	5
Compleitud	5	-	5
Consistencia	5	-	5
Credibilidad	5	-	5
Actualidad	5	-	5
Accesibilidad	5	5	5
Conformidad	0	0	0
Confidencialidad	5	5	5
Eficiencia	0	0	0
Precisión	0	0	0
Trazabilidad	0	0	0
Comprensibilidad	0	0	0
Disponibilidad	0	0	0
Portabilidad	0	0	0
Recuperabilidad	5	5	5

Figura 5.24. Artefacto de Características de DQ y su importancia en Roles de Usuario DQ por PI del C.E. N°3.

d) *Especificación de requisitos incluyendo DQ.* Se analizaron los valores de ponderación promedio y se eligieron las características que tuvieron puntuación mayor o igual al puntaje de corte definido igual a 4. Se definieron acciones para las características de DQ definidas, tomando como referencia las acciones tipo contenidas en el Anexo III de esta tesis. A continuación se presentan algunos ejemplos.

Listado de acciones en PI's		
PI	Características de DQ	Acciones
HIST	Completitud	1.Organizar los datos del historial clínico en un formulario de la siguiente manera: a)Condiciones: hábitos, actividad física y patologías del paciente. b) Antecedentes familiares: patologías que tienen los padres del paciente. c) Derivación: medicamentos ingeridos por el paciente y si es derivado de algún especialista. d) Encuesta nutricional: horarios y porciones de comidas del paciente.
	Credibilidad	1.Ingresar fecha de ingreso o última actualización de cualquier dato (condición, antecedentes familiares, derivación o encuesta) correspondiente al historial. 2.Dar la opción al usuario de verificar el ingreso de sus datos.
	Consistencia	1.Proporcionar textos o ayuda que ayuden a identificar cuando el dato ingresado no es consistente (si es un dígito válido en el campo que ingresa). 2.Clasificar y ordenar los datos en un formulario para que sean consistentes (ordenamiento de los datos determinado por la nutricionista). 3.Ordenar en una tabla la encuesta nutricional, colocando en las columnas los horarios, las comidas y las porciones.
	Eficiencia	1.Optimizar largo y tipo en las patologías del paciente y antecedentes familiares. 2.Usar listas, checkbox como una forma de ingreso de patologías y medicamentos. 3.Pantallas de fácil ingreso de información. Poco uso del mouse e incorporar ingresos del teclado, en datos: hábitos, actividad física, nombre de especialista, especialidad y diagnóstico inicial.
	Comprensibilidad	1.Incorporación de etiquetados de textos claros. 2.Incorporación de textos de ayuda en las abreviaciones del nombre de las patologías. 3.Para tipos de datos como fecha o que necesitan de un formato específico incorporar una máscara. 4.Indicar claridad en mensaje de error. 5.Definir abreviaturas en los nombres de las patologías (definidas por nutricionista).
	Disponibilidad	1.Definir la última fecha en que fueron actualizados los datos del Historial, y cargar la información. 2.Contar con una adecuada política de resguardo de la información.
CONT	Exactitud	1.Definir reglas de validación. 2.Establecer unidades de medida en la talla, contorno de cintura (cm) y peso (kg) del paciente. 3.Definir largos para talla, peso, contorno de cintura y observaciones. 4.Definir función para obtener IMC y cantidad de decimales a utilizar. 5.Realizar validaciones sintácticas en la talla, peso y contorno de cintura. 6.Definir calores para el estado nutricional.
	Completitud	1.Establecer niveles de ingreso. Por ejemplo: obligatorios, deseable, entre otros. 2.Organizar los datos en un Control, de tal forma que los datos obligatorios queden juntos. Esto sirve para identificar fácilmente datos obligatorios no ingresados. 3.Indicar aviso en los datos que no cumple con ciertas normas u otras restricciones de carácter legal o de tipo de procedimiento administrativo. 4.No guardar un formulario hasta asegurar que esté completo.
	Consistencia	1.Establecer reglas de consistencia para diferencias de peso con respecto al primer y último control realizado al paciente. 2.Validaciones de rango de valores para el peso, talla y contorno de cintura. 3.Proporcionar ayuda para identificar cuando el dato ingresado no es consistente. 4.Clasificar y ordenar los datos en un Control para que sean consistentes.
	Actualidad	1.Revisar generación de fecha, a través de la hora que proporciona el computador. 2.Determinar reglas del negocio que ayuden a tener disponible las diferencias de peso y peso ideal con respecto a los controles anteriormente realizados.
	Accesibilidad	1.Establecer formatos estándar para mostrar la información del estado nutricional. 2.Ordenamiento de datos por criterios previamente establecidos. 3.Incorporar criterios de búsqueda que ayuden al acceso a la información del Control.
	Confidencialidad	1.Definir la confidencialidad de la información del Control. 2.Asegurar el equipo que contendrá a la aplicación mediante contraseña.
	Recuperabilidad	1.En una validación de datos, limpiar sólo los datos erróneos para ingresarlos nuevamente. 2.Al cerrar una pantalla, almacenar datos ingresados o indicar transacciones pendientes de completar.

Figura 5.25: Listado de acciones en PI's del C.E. N°3.

A continuación, se complementó la especificación de requisitos con las características de DQ consideradas relevantes. A continuación se presentan algunos ejemplos.

Definición del Caso de Uso	
Código	1
Nombre	Ingresar Datos Personales
Actor Principal	Nutricionista
Descripción	Registrar los datos personales del paciente.
Productos de Información (PI's)	Ficha de Paciente Características de DQ: Exactitud, Completitud, Consistencia, Credibilidad, Actualidad, Accesibilidad, Comprensibilidad, Disponibilidad y Recuperabilidad
<i>Definición del Escenario Principal.</i>	
Precondiciones	
Flujo Principal	<ol style="list-style-type: none"> 1. El sistema despliega una pantalla con los campos: Nombres, apellido paterno, apellido materno, Rut, fecha de nacimiento, sexo, trabajo, situación familiar, número de hijos, correo electrónico, teléfono y ciudad. 2. La nutricionista ingresa los datos solicitados. 3. El sistema valida los datos ingresados. 4. El sistema verifica que al menos estén los campos: Nombres, apellido paterno, apellido materno, Rut, fecha de nacimiento y sexo. 5. El sistema ingresa los campos fecha de ingreso y edad. 6. El sistema almacena los datos en la base de datos y el caso de uso termina.
PostCondiciones	<ul style="list-style-type: none"> • Datos personales del paciente ingresados correctamente en la base de datos del sistema.
Flujos Alternativos	<ol style="list-style-type: none"> 1.1 El sistema llama al caso de uso "Consultar Ciudades". 2.1 Si está ingresado el nombre y el Rut, el sistema valida los datos y llama al caso de uso "Buscar Paciente". 3.1 Si los datos no son válidos el sistema muestra nuevamente la pantalla de ingreso de datos con el correspondiente mensaje de error y regresa al flujo 2. 4.1 Si los datos ingresados no son suficientes, el sistema muestra el correspondiente mensaje de error y regresa al flujo 2.

Figura 5.26: Caso de Uso de Ingresar Datos Personales del C.E. N°3.

Definición del Caso de Uso	
Código	2
Nombre	Ingresar Historial Clínico
Actor Principal	Nutricionista
Descripción	Registrar los datos médicos del paciente.
Productos de Información (PI's)	Historial Clínico Características de DQ Completitud, Consistencia, Credibilidad, Accesibilidad, Eficiencia, Comprensibilidad, Disponibilidad
<i>Definición del Escenario Principal.</i>	
Precondiciones	<ul style="list-style-type: none"> • Datos personales del paciente ya registrados.
Flujo Principal	<ol style="list-style-type: none"> 1. El sistema despliega una pantalla con los campos: Hábitos del paciente (tabaco y alcohol (cantidades)), actividad física, tipo de paciente, antecedentes familiares del padre y de la madre, médico, la especialidad, diagnóstico inicial, medicamentos y dosis. Además de la encuesta nutricional. 2. La nutricionista ingresa los datos solicitados. 3. El sistema valida los datos ingresados. 4. El sistema ingresa los campos fecha de ingreso y edad. 5. El sistema almacena los datos en la base de datos y el caso de uso termina.
PostCondiciones	<ul style="list-style-type: none"> • Datos del historial clínico del paciente ingresados correctamente en la base de datos del sistema.
Flujos Alternativos	<ol style="list-style-type: none"> 1.1 El sistema llama al caso de uso "Consultar medicamentos". 3.1 Si los datos no son válidos el sistema muestra nuevamente la pantalla de ingreso de datos con el correspondiente mensaje de error y regresa al flujo 2.

Figura 5.27. Caso de Uso de Ingresar Historial Clínico del C.E. N°3.

e) *Verificación de DQ a nivel de especificaciones.* En esta última actividad del ciclo, se realizó el seguimiento de la lista de control útil para verificar que las especificaciones contengan las características de DQ. Por razones de espacio esta lista no será incluida en esta tesis.

Como resultado de esta actividad, todas las especificaciones de requisitos quedaron centradas en la DQ, es por ello que se determina concluir esta primera fase de desarrollo del SI de especificación de requisitos en una iteración.

5.3.3 Análisis de la aplicación

Este caso de estudio nos deja las siguientes experiencias:

- Este sistema es clasificado como pequeño, dada las funcionalidades y cantidad de usuarios que tiene. Sin embargo, se demuestra que DeWIK se adapta a cualquier tipo de SI sin importar el alcance que éste debe tener.
- En este caso de estudio se evaluó la posibilidad de que el DT definiera las características de DQ que deben ir, aquellas características que no deben ir y consultar aquellas que no se sabe si incluirlas o no. Esta alternativa de implementación demuestra que la acción del DT asegura la incorporación de aquellas características que si deben ir desde su punto de vista.

5.4 Casos de Estudio 4: Sistema de Facturación a Proveedor.

Este desarrollo forma parte de un proyecto del área de sistemas de una conocida empresa de retail. Este proyecto consiste en el desarrollo de un Sistema de Información que permita registrar y controlar los procesos involucrados en las negociaciones realizadas con proveedores, su facturación y cobro. Sus principales funcionalidades son:

- Contar con información oportuna y eficiente para la toma de decisiones.
- Entregar información íntegra para realizar el proceso de facturación.
- Entregar información de calidad a aquellas áreas destinadas al control, seguimiento y gestión de los datos.

A continuación, se explica cómo fue aplicado DeWIK en este caso de estudio.

5.4.1 Etapa 1. Definición del entorno de aplicación de DeWIK

a) *Denominación y formación del DQM.* Se designa a la alumna memorista de esta tesis como DQM, no necesitó inducción por ser una de las creadoras del método.

b) *Definición de sus Roles.* (a) Integrante del DT, conformado por el jefe del proyecto y un equipo desarrollador de 5 personas y (b) Usuario de aplicación, se definieron los perfiles: “Administrador del Sistema”, “Ejecutivo de Confección”, “Ejecutivo Analista”, “Ejecutivo de Costos”, “Facturador” y “Generador de Informes”.

c) *Entrenamiento del DT en aspectos relacionados con DQ y en el método DeWIQ.* Se realizó un entrenamiento proporcionado por el DQM y lectura del Anexo II de esta tesis. El modelo de DQ elegido fue el dado por la norma ISO/IEC 25012. Para la reclasificación de roles, todos los roles de usuario de la aplicación se comportan en función de los roles de usuario DQ consumidor y productor de datos, de acuerdo a su interacción en el SI.

5.4.2 Etapa 2. Definición de los requisitos de la aplicación con DQ

a) *Identificación de los requisitos de la aplicación.* Se utilizaron los procedimientos indicados en un desarrollo de un SI para interactuar con los usuarios y captar los requisitos y funcionalidades que ellos requieren respecto de la aplicación a desarrollar. El artefacto usado para la especificación de requisitos fueron los casos de uso. En la Figura 5.28 se presenta la especificación de requisitos del proyecto.

Requisito
CU-01 Crear Formulario de Negociación
CU-02 Emitir Carta Explicativa
CU-03 Ingresar Canje
CU-04 Liberar Formulario de Negociación
CU-05 Consultar Formulario de Negociación
CU-06 Adjuntar Respaldo a una Negociación
CU-07 Replicar Formulario de Negociación
CU-08 Ingresar NC
CU-09 Unificar Cobros

Figura 5.28. Listado de especificación de Requisitos del C.E. N°4.

b) *Identificación de los PI's involucrados en los requisitos de la aplicación.* Se utilizó como estrategia de identificación de PI's las listas de control suministradas por el método, siguiendo sus instrucciones de aplicación.

A continuación se presenta el artefacto de PI's de la aplicación con la información relevante de algunos de los PI's identificados.

PI's de la Aplicación					
Id PI	Nombre PI	Propósito	Descripción	Roles de Usuario de la Aplicación	Requisito
FONE	Formulario de Negociación	Conjunto de negociaciones agrupadas por proveedor	id_formulario, fecha_ingreso, proveedor, tipo_negociacion, frecuencia_facturacion, condicion_pago, fecha_inicio, fecha_fin, observaciones, estado, firma_analista, destinatario, id_carta, concepto_facturacion, informacion_cliente	Administrador del Sistema Ejecutivo Analista Ejecutivo de Confección Ejecutivo de Costos Facturador	CU_01, CU_02, CU_03, CU_04, CU_05, CU_06, CU_07, CU_09
NEGO	Negociación	Acuerdo comercial realizado entre un proveedor y un ejecutivo para realizar descuentos y bonificaciones en la venta de artículos	id_negociacion, id_formulario, impacta_precio, envio_correccion, fecha_ingreso, fecha_inicio, fecha_termino, cobertura, estado, nombre, monto	Administrador del Sistema Ejecutivo Analista Ejecutivo de Confección Ejecutivo de Costos	CU_01, CU_06, CU_09
CAEX	Carta Explicativa	Carta resumen de la negociación efectuada con el proveedor.	id_carta, informacion	Facturador	CU_02
RENE	Respaldo de Negociación	Archivo de Respaldo que complementa el formulario	id_respaldo, fecha, nombre_archivo, id_formulario, id_negociacion	Administrador del Sistema Ejecutivo Analista Ejecutivo de Confección Ejecutivo de Costos	CU_06
NOCR	Nota de Crédito	Documento tributario que indica una anulación de cobro al proveedor	id_documento, folio, id_formulario, documento_referencia, tipo_documento_referencia, estado, fecha_emision, fecha_insercion, proveedor, condicion_pago, observacion, valor_neto, valor_impuesto, valor_adicional, valor_retencion, valor_bruto, valor_exento, cuenta, glosa, articulo, cantidad	Administrador del Sistema Facturador	CU_08
MEMO	Memo	Documento interno que indica que el cobro del proveedor será a través de una NC emitida por ellos	id_memo, proveedor, estado, fecha, referencia, valor_neto, valor_impuesto, valor_adicional	Facturador ERP Administrador del Sistema	CU_03

Figura 5.29: Artefacto de Identificación de PI's del C.E. N°4.

En la especificación de requisitos fueron incorporados los PI's identificados. Luego se registró la información en el artefacto Matriz de relación de Requisitos y PI's.

Matriz de relación Requisitos y PI's						
Requisitos	FONE	NEGO	CAEX	RENE	NOCR	MEMO
CU_04-Liberar Formulario de Negociación	X					
CU_05-Consultar Formulario de Negociación	X					
CU_06-Adjuntar Respaldo a una Negociación	X	X		X		
CU_07-Replicar Formulario de Negociación	X					
CU_01-Crear Form de Negociación	X	X				
CU_09-Unificar Cobros	X	X				
CU_02-Emitir Carta Explicativa	X		X			
CU_03-Ingresa Canje	X					X
CU_08-Ingresa NC					X	

Figura 5.30. Artefacto Matriz de relación Requisitos de aplicación y sus PI's del C.E. N°4.

A continuación se ingresó la información en el artefacto de Interacción de PI's con Roles de Usuario de la Aplicación y DQ.

Interacción de PI's con Roles de Usuario de la Aplicación y DQ			
Productos de Información	Roles de Usuario de la Aplicación	Rol de usuario DQ	
		Consumidor de Datos	Productor de Datos
FONE	Administrador del Sistema		
	Ejecutivo Analista	X	X
	Ejecutivo de Confeción	X	X
	Ejecutivo de Costos	X	X
	Facturador	X	
NEGO	Administrador del Sistema		
	Ejecutivo Analista	X	X
	Ejecutivo de Confeción	X	X
	Ejecutivo de Costos	X	X
CAEX	Facturador		X
RENE	Administrador del Sistema		
	Ejecutivo Analista		X
	Ejecutivo de Confeción		X
	Ejecutivo de Costos		X
NOCR	Administrador del Sistema		
	Facturador		X
MEMO	Facturador		X
	ERP	X	

Figura 5.31: Artefacto de Interacción de PI's con Roles de Usuario de la Aplicación y DQ del C.E. N°4.

Como última actividad de esta etapa, se registra la interrelación que existe en los PI's en el artefacto Matriz de Relación de PI's.

Matriz de Relación de PI's						
Productos de Información	FONE	NEGO	CAEX	RENE	NOCR	MEMO
FONE		>	>	>		
NEGO			>	>		
CAEX						
RENE						
NOCR						
MEMO					>	

Figura 5.32: Artefacto de Matriz de Relación de PI's del C.E. N°4.

c) *Definición de aspectos de DQ.* Para este caso de estudio se decidió considerar las 15 características de DQ por cada PI, por lo tanto no se realizó encuesta a los usuarios.

d) *Especificación de requisitos incluyendo DQ.* Se definieron acciones para las 15 características de DQ, tomando como referencia las acciones tipo contenidas en el Anexo III de esta tesis. A continuación se presentan algunos ejemplos.

Listado de acciones en PI's		
PI	Características de DQ	Acciones
FONE	Exactitud	1.Incorporación de selectores de información en dónde los datos están predispuestos para su ingreso de parte del usuario, evitando errores. Ejemplo: selectores de concepto de facturación, proveedor, destinatario, entre otros.
	Consistencia	1.Validación de formato de correo 2.Validación de formatos de fechas 3.Validación en textos de no ingresar caracteres especiales.
	Actualidad	1.Para selectores de fecha indica Calendario que marca fecha y/o periodo. 2.Actualizaciones on-line al consultar proveedor 3.Versión del sistema 4.Actualizaciones de artículos 5.Incorporación stock actualizado de acuerdo a fecha de inicio de negociación
	Accesibilidad	1.Incorporación de filtros búsqueda para artículo 2.Para selectores se facilitan las búsqueda por código y/o descripción 3.Incorporación de subconjunto de caracteres coincidentes con la descripción de un campo
	Confidencialidad	1.Acceso a usuario restringido por perfiles
	Eficiencia	1.Cargas de datos con consultas a Base de datos optimizadas en tiempo de respuesta 2.Las cargas están calendarizadas conforme a tiempo de sobrecarga, tasa de transferencia y velocidad de procesamiento de la información
	Precisión	1.Establecer una máximo de caracteres en los ingresos de textos 2.Validación de números con decimales y de enteros
	Comprensibilidad	1.Para el ingreso de valores monetarios se indica signo "\$" 2.Para el ingreso de fechas se indica máscara o formato con el que debe ser ingresado 3.Para el ingreso de valores de porcentajes se indica signo "%" 4.Para la incorporación de datos con descripción se incorporan mensajes flotantes para ayudar su ingreso
	Portabilidad	1.Salida Excel 2.Réplica del formulario
	Recuperabilidad	1.A través de estructura Xml 2.El usuario al cerrar ventana, se envía mensaje de advertencia

Figura 5.33: Listado de acciones en PI's de un SI del C.E. N°4.

A continuación, se complementa la especificación de requisitos con las características de DQ consideradas relevantes. A continuación se presentan algunos ejemplos.

Caso de Uso: Liberar Formulario de Negociación
ID: CU-04
Breve Descripción: Este Caso de Uso permite al actor, liberar un formulario de Negociación para su facturación.
Actores Principales: Administrador del Sistema, Ejecutivo Analista. Ejecutivo de Confección, Ejecutivo de Costos y Facturador
Actores Secundarios:
Productos de Información: 1. <u>Formulario de Negociación:</u> <ul style="list-style-type: none"> • <u>Características de Calidad de Datos:</u> Exactitud, Completitud, Consistencia, Credibilidad, Actualidad, Accesibilidad, Confidencialidad, Eficiencia, Precisión, Trazabilidad, Comprensibilidad, Recuperabilidad y Portabilidad.
Precondiciones: 1. El usuario debe estar validado en el sistema
Flujo Principal: <ol style="list-style-type: none"> 1. El caso de uso inicia cuando el usuario elige la opción Liberar Formulario 2. El sistema da la opción de consultar formularios (REF: CU_FACTPROV001_10). 3. En caso de consultar, el usuario completa los filtros de consulta y escoge la opción buscar. 4. En caso de consultar, el sistema valida los filtros de consulta. 5. El sistema muestra en pantalla el listado de los formularios que cumplen con las condiciones de búsqueda. 6. El usuario selecciona los formularios que desea liberar y luego escoge la opción Liberar Formulario. 7. El sistema cambia el estado del o los formularios seleccionados de Pendiente a Liberado.
Post-Condiciones:
Flujos Alternativos: 4.1 El sistema encuentra filtros vacíos. 4.2 El sistema no encuentra formularios de negociación que cumplan con las condiciones de búsqueda por lo que muestra vacío el listado de formularios para liberar

Figura 5.34: Caso de Uso Liberar Formulario de Negociación del C.E. N°4.

e) *Verificación de DQ a nivel de especificaciones.* En esta última actividad del ciclo, se realizó el seguimiento de la lista de control útil para verificar que las especificaciones contengan las características de DQ.

Como resultado de esta actividad, se determinó que todas las especificaciones de requisitos quedaron centradas en la DQ, es por ello que se determina concluir esta primera fase de desarrollo del SI.

5.4.3 Análisis de la aplicación

Este caso de estudio nos deja las siguientes experiencias:

- Este proyecto corresponde a un desarrollo en una empresa de retail reconocida en la industria. Sus sistemas son de gran envergadura y críticos en la información que proporcionan. Los problemas de DQ son frecuentes al interactuar con grandes volúmenes

de información. DeWIK fue implementado bajo este contexto, considerando su aplicación en forma exitosa y su disminución en la cantidad de errores resultantes.

- Se consiguió un involucramiento mayor de los usuarios y del DT en el conocimiento del problema y su alcance en el negocio, desde la etapa de especificación de requisitos y en todo el proceso de desarrollo del SI.
- Al poner en marcha este sistema, hubo disminución de recursos en las tareas de pruebas integrales, inducción a los usuarios y estabilización de la operación post paso a producción.
- Dado el éxito de esta implementación, se está evaluando que el DQM pase a ser un perfil establecido en el área para que tome parte activa en el monitoreo y revisión constante de la DQ en los sistemas de información presentes en la organización y en los proyectos actuales en carpeta.

5.5 Lecciones aprendidas

Como resultado de la experiencia en la aplicación de los cuatro casos de estudios, presentamos una tabla con las lecciones aprendidas más importantes de la implementación de DeWIK en relación a las alternativas de implementación de las características de DQ.

Esta tabla está construida en base a las alternativas de implementación de las características de DQ en la etapa de definición de requisitos con DQ, en cuál de las tareas de la implementación de DeWIK hubo mayor utilización de recursos (en la etapa de proceso de desarrollo de software especificación de requisitos) y en qué etapa del proceso de desarrollo de software posterior, hubo disminución en el uso de recursos.

Caso de Estudio	Alternativa de Implementación de DQ	Tarea de DeWIK con mayor uso de recursos	Etapas del desarrollo del SI con disminución de recursos.	Principales Aportes
1	El DT descarta aquellas características de DQ que no son necesarias en el SI y consulta por el resto al usuario.	Identificación de los PI's involucrados en los requisitos de la aplicación. El usuario debe conocer el alcance que tendrán los PI's de la aplicación para tener un juicio adecuado al seleccionar las características que se incluyen o no en cada PI.	Construcción del SI, el DT filtra aquellas características que desde su punto de vista aportan DQ al SI, por lo tanto el resultado obtenido de requisitos con DQ apunta a disminuir la construcción de la solución con características de DQ que realmente aporten a la solución.	Mayor conocimiento del DT en aspectos de DQ al seleccionar las características de DQ consultadas al usuario. El usuario debe adquirir conocimientos relacionados con el alcance real del PI para seleccionar aquellas características de DQ que desea incorporar.
2	Consultar al usuario por la inclusión de las 15 características de DQ al SI.	Definición de aspectos de DQ, al preparar la encuesta por la totalidad de las características de DQ. El usuario debe contestar esta extensa encuesta.	Puesta en Marcha, en las actividades de capacitación al usuario e implementación, al adquirir un mayor conocimiento del sistema durante la etapa de especificación de requisitos, en esta etapa no se necesitó inducir acerca de la solución funcional del sistema con DQ.	El DT debe preparar una encuesta que sea descriptiva en las características que debe indicar el usuario a incluir. Mayor participación del usuario al seleccionar las características de DQ.
3	El DT descarta aquellas características de DQ que no considera necesarias en el SI, incluye las que considera necesarias y consulta por el resto al usuario.	Identificación de requisitos de la aplicación, el DT debe conocer a cabalidad los requisitos de la aplicación para tener un juicio adecuado al seleccionar las características que se incluyen o no en el SI.	Diseño del SI, al obtener requisitos con DQ después de un proceso minucioso de selección de características de DQ, las tareas de diseño y modelado se ven centradas sólo en aquellas características de DQ realmente necesarias.	El DT indica las características de DQ necesarias en el SI, por lo tanto debe tener los conocimientos necesarios para una correcta inclusión de DQ
4	El DT incluye las 15 características al SI, no interviene el usuario.	Especificación de requisitos con DQ, al implementar acciones a las 15 características de DQ	Pruebas integrales de usuario, al tener menor cantidad de errores.	El DT debe planificar la implementación de los requisitos adecuadamente considerando el extenso listado de acciones de DQ a incluir.

Figura 5.1: Lecciones aprendidas en la implementación de DeWIK.

Este análisis nos servirá para realizar una estimación del costo de implementación del método DeWIK, que por ahora no será abordado en nuestro trabajo.

5.6 Ajustes de DeWIK

La implementación de DeWIK nos llevó a realizar ajustes al método para mejorar su funcionamiento y para cumplir con sus objetivos. A continuación se presentan estos ajustes.

- Se colocaron nombres a los artefactos para identificarlos con mayor facilidad.
- Se agregó al artefacto Matriz de relación de PI's para determinar la interrelación entre los PI's, 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. La definición de este artefacto aparece en la sección 4.2.2 de esta tesis.
- Se decide agregar en el pie de la encuesta que se debe realizar a los usuarios para identificar las características de DQ, la descripción del PI encuestado. Esto permite clarificar el contexto del PI consultado.
- La participación del usuario productor de datos es considerada sólo en aquellas características que son dependientes del sistema por tener una participación más directa en ellas.
- Se decide implementar distintas alternativas de elección de DQ. En caso de que el DT elija las características de DQ a incluir en el SI, aseguraremos la DQ desde su punto de vista, alternativa útil en caso de que el DT haya tenido experiencia previa en desarrollo de software. De igual forma que para el rol de usuario consumidor de datos, esta elección es útil para agilizar esta tarea que puede llegar a ser extensa y lenta.
- En un caso de estudio, se decide que el DQM no sea necesariamente la creadora de DeWIQ, la alumna tesista, para asegurar que las tareas de inducción e implementación del método sean realizadas en forma exitosa.

Capítulo 6

Conclusiones

6.1 Análisis de los objetivos propuestos/cumplidos

Los Sistemas de Información han evolucionado considerablemente, constituyéndose un factor crítico en las organizaciones para garantizar acertadas tomas de decisiones y adecuadas estrategias de crecimiento. Este impacto positivo en las organizaciones, lleva a incorporar avances tecnológicos de la información, implementación de nuevas metodologías y modernización de técnicas para mejorar sus procesos.

Esta tesis presenta una propuesta para abordar la Calidad de los Datos desde el proceso de desarrollo de un Sistema de Información, cuando recién se comienzan a definir los Productos de Información involucrados con él y en 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 habitual de desarrollo.

En el primer capítulo de esta tesis se expusieron los objetivos parciales que se pretende cumplir para satisfacer el objetivo global es esta tesis que es:

Desarrollar un método para definir los requisitos de un Sistema de Información centrados en la calidad de datos.

A continuación se presenta el análisis del cumplimiento de los diferentes objetivos parciales.

Objetivo 1, 2 y 3. *Estudiar propuestas existentes en la literatura que aborden la incorporación de aspectos de DQ como parte del proceso de desarrollo de software. Estudiar en detalle la metodología TDQM, con el fin de definir su aplicación en el proceso de desarrollo de software. Estudiar la norma ISO/IEC 25012 con el propósito de identificar las características de calidad de datos que pueden ser relevantes para el proceso de desarrollo de software.*

Se ha realizado una revisión sistemática de la literatura que seleccionó aquellos estudios que abordan la calidad de los datos en el proceso de desarrollo de software.

Se estudiaron en detalle aquellos aportes útiles para comprender el contexto de implementación de la calidad de información en los SI. Se fueron extrayendo los elementos bases para nuestro trabajo y se realizó un estado del arte (presentado en el capítulo 3) de acuerdo a cuatro enfoques. El primero, enfocado a los conceptos relacionados con la calidad de datos y cuáles son los problemas de DQ presentes en los sistemas de información. El segundo, se orientó al estudio de propuestas relacionadas con el desarrollo de sistemas de información y de cómo es abordada la calidad de los

datos en las etapas de desarrollo que componen un modelo genérico de desarrollo de un SI. El tercer enfoque correspondió a estudiar la norma ISO/IEC 25012 en donde se exponen las principales situaciones anómalas de calidad de datos, establece un modelo genérico que define quince características de calidad de datos para resolver estos problemas. El cuarto enfoque está orientado a estudiar la metodología TDQM y aquellas investigaciones que la implementan, indicando sus principales aportes en su aplicación.

Objetivo 4. *Definir un método, basado en los principios de TDQM y la norma ISO/IEC 25012, que permita abordar la DQ en aquellas etapas del proceso de desarrollo de software que estén relacionadas con la definición de requerimientos de un SI.*

En el capítulo 4 de esta tesis, se ha presentado la definición del método DeWIK (Development With Information Quality). Esta definición es establecida en base a: a) el concepto de producto de información (incluido en el estado del arte), usado en forma beneficiosa para indicar la utilidad de la información a los usuarios o consumidores de datos, b) la norma ISO/IEC 25012 que establece un modelo genérico de calidad de datos y c) el enfoque de TDQM que nos inspira la necesidad constante de mejoras en los PI's en un contexto organizacional.

Objetivo 5. *Evaluar el método propuesto mediante la implementación en casos de estudio representativos.*

En el capítulo 5 de esta tesis se presenta la aplicación de DeWIK en cuatro distintos casos de estudio. Estos casos de estudio consisten en el desarrollo de sistemas de información de diversa índole. Los tres primeros, son proyectos de títulos de alumnos de la carrera de ingeniería civil en informática y el último caso corresponde a un proyecto perteneciente al área de sistemas de una empresa de retail conocida en la industria.

Como resultado de la implementación de este método, se realizó una evaluación ajustando sus definiciones teóricas de acuerdo a la experiencia dada por la aplicación en cada caso.

6.2 Principales aportes

Como principales aportes de esta tesis podemos mencionar lo siguiente:

- Desarrollo de un método que permite incorporar la calidad de los datos en un proceso de desarrollo de software. Está dirigido a obtener la identificación de los requisitos de calidad de los datos durante la etapa de especificación de requisitos. Sus componentes básicos son:

el uso del concepto de Producto de Información, la norma de calidad de datos ISO/IEC 25012 y el enfoque Total Data Quality Methodology.

- Contribuye a la especificación y aseguramiento de las características de calidad de los datos con base en un estándar internacional de calidad.
- DeWIQ 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.
- DeWIQ incorpora los puntos de vista de diferentes roles dentro del proceso de desarrollo.
- La aplicación del método permite que tanto usuarios como desarrolladores tengan de forma temprana consciencia sobre la calidad de datos de los PI's que luego les proveerá un SI.
- El conjunto de actividades que componen la implementación del método, proporciona una serie de artefactos que aportan a tener una mejor especificación de requisitos y que facilita la tarea de elaborar la documentación final relacionada con el sistema.
- A largo plazo, el método posibilita que una organización tenga identificados sus PI's relevantes y de ese modo tenga una base sólida para emprender un proceso de evaluación y mejora continua de los mismos.

6.3 Contraste de resultados

La publicación aceptada para una conferencia de carácter nacional, se detalla a continuación:

1. Caro, A., Fuentes, A., Soto, M., *Desarrollando Sistemas de Información Centrados en la Calidad de Datos*, in *II CONGRESO INTERNACIONAL DE COMPUTACION E INFORMATICA DEL NORTE DE CHILE, Infonor-Chile 2011*. 2011: Copiapó, Chile.

6.4 Trabajos futuros

Las principales líneas futuras abiertas como resultado de esta investigación son las siguientes:

- Realizar estimaciones relativas a la incidencia del uso del método en los recursos requeridos por un proyecto de desarrollo. Esto indicará el costo real de implementación del método versus el beneficio de incorporar calidad de datos en forma temprana.
- Extender el método hacia las otras etapas de desarrollo de un SI. En dichas fases los requisitos de DQ, inmersos en los requisitos de la aplicación, deberán ser modelados, implementados, verificados y validados.

- Incorporar la participación de todos los roles de usuario DQ en la aplicación del método en la medida de lo necesario.
- Considerar la existencia de PI's provenientes de desarrollos anteriores, esto puede implicar el ajuste de alguna(s) actividad(es), como la obtención los PI's para verificar los considerados en este contexto.
- Incorporar como una actividad del método, la planificación de tiempos de implementación del método
- Desarrollar una herramienta que permita realizar los artefactos en forma automática, con el fin de disminuir el tiempo y uso de recursos en la implementación del método.

Anexos

En este capítulo se presentará los anexos vinculados con el trabajo de tesis, cuya finalidad es proporcionar información que complemente la investigación realizada. El Anexo I contiene información relacionada de cómo se aplicó la revisión sistemática de la literatura.

I Anexo I. Resultados de Revisión Sistemática de la Literatura

Una Revisión Sistemática de la Literatura (RSL) permite buscar en forma planificada y ordenada, los estudios relacionados con el objetivo propuesto, y con esto determinar si la propuesta es pertinente en el área de investigación (Kitchenham, 2004). En concreto, el propósito de una RSL es proporcionar un marco de trabajo y/o los antecedentes necesarios para posicionar una nueva actividad de investigación.

En particular, en el contexto de esta tesis, el objetivo es resumir la evidencia existente respecto de cómo se ha incorporado la Calidad de Datos en etapas tempranas de desarrollo de software. A continuación se exponen los resultados obtenidos de la RSL realizada en esta tesis, de acuerdo a las tres etapas principales de una RSL definidas en (Kitchenham, 2004).

I.1 Identificación de Necesidad de Revisión

La primera actividad de esta etapa es la identificación de la necesidad de llevar a cabo la revisión. La necesidad de investigar se justifica por el interés de contribuir, en este caso en el marco de un proyecto de investigación, a la mejora de la calidad de los datos en general y en particular a incorporar la DQ en etapas tempranas en desarrollo de software. Por tanto el objetivo de esta RSL es: “resumir la evidencia existente respecto de cómo se ha incorporado la DQ en etapas tempranas de desarrollo de software”.

A partir de este objetivo se definieron las interrogantes de investigación que ayudarán a la búsqueda de la literatura.

- ¿Qué experiencia existe respecto de incluir la DQ en etapas tempranas de desarrollo de software?
- ¿Es posible identificar ventajas y desventajas con la inclusión temprana de la DQ en comparación con metodologías tradicionales?
- ¿Qué otras soluciones han incluido la DQ como un factor relevante antes de contar con un software “físico”?

Se consideraron los siguientes recursos: Scirus , ScienceDirect, IEEE,ACM, Xplore, Google Scholar y conferencias especializadas en calidad, por ejemplo ICIQ (International Conference on Information Quality) de los últimos años.

La segunda actividad de esta etapa corresponde a la definición del protocolo de búsqueda. Para esto se definieron los siguientes términos de búsqueda: “data quality”, “information quality”, “software development”, “software development phases” y sus equivalentes en español. Además, se definieron criterios de búsqueda como por ejemplo: identificar sinónimos para los términos de búsqueda, usar operadores lógicos OR y AND para combinar los términos, considerar la obtención de nuevos términos que ayuden a encontrar trabajos de aporte, revisar opciones de búsqueda avanzada, revisar las referencias de los artículos ya seleccionados e identificar a los autores del área de Calidad de Datos y buscar sus publicaciones en índices como el DBLP o páginas personales.

La tercera actividad consistió en la definición del protocolo de revisión. Este protocolo considera la definición de las normas de revisión de los trabajos, los criterios de inclusión y exclusión que se emplearán, la estrategia de extracción de datos y, finalmente, la estrategia de síntesis.

Como normas de revisión se definieron las siguientes:

- Leer el resumen e introducción y aplicar criterios de inclusión y exclusión para considerar o rechazar el artículo.
- En caso de inclusión, extraer los datos relevantes de acuerdo con la estrategia de extracción definida.
- Llevar un registro acerca de los estudios primarios aceptados y rechazados con su correspondiente motivo de exclusión.

Los criterios de inclusión, considerarán los estudios primarios que aborden el tema de DQ en el contexto de la inclusión de este atributo en fases de desarrollo de software, de acuerdo a los siguientes puntos:

- Propuestas de metodologías, modelos de datos o técnicas que incluyan la DQ en antes de la construcción “física” de software.
- Experimentos de sistemas construidos con estas nuevas metodologías.
- Propuestas de normas o de otras soluciones que consideran la calidad como una necesidad importante al momento de diseñar e implementar software.

Los criterios de exclusión establecen que se descartarán aquellos estudios que:

- Correspondan a información proveniente de sitios sin la aprobación de la comunidad científica.
- Su contenido no corresponda con ningún criterio de inclusión.
- Correspondan a información de sitios personales relatando vivencias o experiencias, sin un análisis aprobado por la comunidad científica de computación.

Se estableció una norma de registro de los resultados de la búsqueda, de modo que fuera posible clasificar los artículos, mantener un registro histórico de búsquedas en formato electrónico, gestionar los resultados ya obtenidos (eliminando duplicados) y crear Tablas sobre resultados de las búsquedas. Se usó un formulario para la extracción de datos con el formato mostrado en la Figura I.1 y se estableció que adicionalmente, sobre la copia del estudio (impresa o digital), se adjuntarán comentarios o se destacarán los puntos más relevantes del trabajo.

Formulario de Extracción de Datos	
Título	Modelado de Requisitos de Calidad de Datos en el Proceso de Desarrollo de Portales Web
Autores	César Guerra-García, Ismael Caballero, Ignacio García-Rodríguez de Guzmán, Mario Piattini
Año	2009
Aporte a este estudio	Desde el punto de vista de portales web hay propuesta de soluciones para incorporar la calidad como un requisito en la etapa de modelado de datos.
Observaciones	Presenta un perfil UML para la gestión de requisitos de Calidad de Datos y contiene un ejemplo que permite explicar algunos conceptos fundamentales de DQ.

Figura I.1. Formulario de extracción de datos

En cuanto a la estrategia de síntesis, se definió en que los datos se resumirán los datos a modo de respuesta a las interrogantes de investigación.

I.2 Desarrollo de la Revisión

En la primera actividad de la búsqueda de estudios primarios, se realizó una búsqueda cuyos resultados parciales se muestran en la Tabla I.1. En la primera columna se identifica la búsqueda, la segunda columna indica la fuente utilizada, la tercera columna las palabras claves usadas en la búsqueda y finalmente en la cuarta columna se presentan el número de coincidencias obtenidas para los términos buscados.

Nueva Búsqueda	Fuente	Clave de Búsqueda	Resultados
1	Google Scholar	“data quality”	4.280.000
2	Google Scholar	“software development”	2.790.000
3	Scirus	“information quality”	3.910.000
4	Scirus	“software development phases”	687.728

Tabla I.1 Búsquedas primarias con términos de interrogación

Por ejemplo, la búsqueda número 1 en Google Scholar con clave “data quality” arrojó del orden de los cuatro millones de coincidencias lo que, evidentemente, es un número muy grande e impide realizar una revisión adecuada de todos los documentos. A medida que mezclamos las búsquedas reducimos considerablemente la cantidad de resultados.

Búsqueda Primaria	Operador	Clave de Búsqueda	Resultados
1,2	AND	data quality and software development	1.070.000
1,4	AND	“data quality and software development phases”	366.960
3,4	AND	“information quality software development phases”	382.285

Tabla I.2. Nuevas búsquedas sobre las búsquedas primarias

En la Tabla I.2 se presenta un ejemplo de nuevas búsquedas que se hicieron sobre los resultados de las búsquedas primarias, las cuales permitieron reducir el número de resultados. En dicha Tabla la columna uno hace referencia al número de búsqueda presentada en la Tabla 1, en la segunda columna se presenta el operador, en la tercera columna el nuevo criterio de búsqueda y finalmente, en la cuarta columna el número de coincidencias.

La segunda actividad del desarrollo de la revisión, selección de estudios primarios, se desarrolló en base a la aplicación del protocolo de revisión definido en la etapa anterior.

Para la realización de la estrategia de extracción y gestión de datos, tercera actividad de esta segunda etapa, se aplicó el protocolo definido en la etapa anterior. Es decir, los datos fueron registrados en el formulario de extracción de datos (ver Figura 1) y se adjuntaron comentarios directamente sobre los trabajos, cada vez que se consideró necesario.

Respecto de la gestión de los datos, se registraron los resultados de las búsquedas y selección de los trabajos.

Finalmente, la síntesis de datos, fue realizada considerando el protocolo de revisión, en que se estableció resumir los resultados como respuesta a cada una de las interrogantes de investigación.

El siguiente anexo tiene por objetivo entregar conocimientos necesarios para realizar una correcta aplicación del método DeWIK (por las siglas en inglés de Development With Information Quality).

II.1 Anexo II. Inducción del método DEWIK

A continuación, se dará a conocer los aspectos más importantes relacionados con la DQ (por las siglas en inglés de Data Quality), el modelo de DQ usado de referencia y las distintas etapas de aplicación por el método DeWIK.

II.1 Conceptos relacionados con DQ.

II.1.1 Calidad de Datos

Entre las múltiples definiciones de DQ o IQ (por las siglas en inglés Information Quality)⁴, una de la más utilizadas es “fitness for use” (Orman et al., 1996; Wang y Strong, 1996; Tayi y Ballou, 1998; Cappiello et al., 2004), en donde se destaca su adecuación al uso y su naturaleza subjetiva. Este término se utiliza con referencia a un conjunto de datos que debe poseer ciertas características, las cuales deben ser suficientes para cumplir con el objetivo definido para su uso (Batini y Sacannapieca, 2006).

Desde el punto de vista del usuario, se puede decir que la DQ “el conjunto de características que hacen que la información tenga más valor para él” (Lesca y Lesca, 1995). Este conjunto de características puede ser denominado dimensiones de DQ⁵, referido como las propiedades o características de calidad que debe poseer la información (Lee et al., 2006).

En efecto, una de las estrategias más comunes para entender este concepto consiste en descomponer la percepción global de la DQ de un conjunto de datos en una serie de “subcalidades” denominadas dimensiones o características de DQ y establecer determinados criterios de aceptación para cada una de ellas. Dichos criterios de aceptación deben estar definidos en función de los posibles rangos de los resultados de medición. Así, si tras medir el nivel de calidad de un conjunto de datos para cada una de las características de DQ, se obtienen resultados fuera de los rangos de aceptación preestablecidos, entonces se considerará como “defectuoso” y, por tanto, se puede decidir no usarlo, o al menos saber que se existe un riesgo al usarlo.

⁴ La literatura trata la DQ o IQ como términos equivalentes, por lo tanto en este trabajo se opta por usar ambos términos como sinónimos.

⁵ La literatura se refiere a los términos característica de calidad de datos, dimensión de calidad de datos, atributo de calidad de datos y aspectos de calidad de datos como términos equivalentes, por lo tanto en este trabajo se opta por usar estos términos como sinónimos.

Para cada contexto (entendiéndose como tal el ámbito de una tarea) es necesario identificar aquellas características que mejor representen los requisitos de DQ de los usuarios. Además hay investigaciones que afirman que la calidad no puede evaluarse independiente de los consumidores del producto (Deming, 1986).

II.1.2 Definición de Productos de Información

Wang se refiere al PI (por las siglas de Producto de Información) como el resultado que produce un sistema de información. Este concepto es incorporado para destacar el hecho de la información de salida que produce un sistema es transferible al consumidor (Wang et al., 2001).

De acuerdo a las definiciones dadas por Wang, se pueden señalar algunos ejemplos comunes de productos de información en una organización.

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, Libro de Ventas Contable.
Sistema de administración de empleados	Liquidación de Sueldo, Comprobante de Día Feriado, Informe de Cotizaciones en AFP, Informe de Comisión de Ventas.
Sistema de devoluciones de artículos.	Archivo de Stock, Notas de Crédito, Informe de Reposición de Artículos a Bodega, Informe de Motivos de Devolución, Listado de Productos Mermados.

Tabla II.1. Ejemplos de Productos de Información.

Como se puede observar en la Tabla II.1, los PI's pueden ser muy variados, con distintos atributos dependientes del sistema de información al que pertenece.

II.1.3 Roles de usuario de DQ

Existe una clasificación de los usuarios, de acuerdo a las funciones que desempeñan relacionadas con la DQ. (a) Consumidor de Datos: rol que utiliza los PI's para satisfacer sus necesidades de información, (b) Productor de Datos: usuario encargado de crear o recoger información para los PI's y (c) Custodio de datos: rol responsable de la correcta gestión del proceso de producción de los PI's.

II.1.4 Modelo de calidad de datos.

Para la aplicación de DeWIQ se debe usar como referencia un modelo de DQ para incorporar la DQ en el SI. Se recomienda usar la norma ISO/IEC 25012 (ISO/IEC-25012, 2008) por ser un

modelo genérico, establecido hace pocos años y porque presenta de manera formal un conjunto de definiciones de conceptos relacionados con la DQ.

Según lo que indica la norma, la gestión y mejora de la DQ es importante para solucionar problemas como: (a) La adquisición de datos de organizaciones de las cuales la DQ del proceso de producción es desconocido o débil, (b) La existencia de datos defectuosos que contribuyen a una información insuficiente, provoca resultados inutilizables y clientes insatisfechos, (c) La dispersión de estos datos entre los sistemas dueños de la información y los usuarios, a menudo carecen de una visión coherente e integrada, necesaria para garantizar la interoperabilidad y la cooperación, entre otros.

La detección de errores o ineficiencias de los datos, da lugar a la mejora e intervenciones correctivas a los datos producidos (análisis, limpieza, enriquecimiento, etc.), a componentes del sistema en el que residen los datos (software y hardware) y los procesos de negocio (formación de usuarios, mejora de los procesos, etc.).

Teniendo en cuenta que el ciclo de vida de los datos es a menudo más largo que el ciclo de vida del software, el modelo de DQ definido en esta Norma Internacional pretende responder a estas necesidades contribuyendo a:

- Definir y evaluar los datos de los requisitos de calidad de la adquisición, producción y procesos de integración de los datos.
- Identificar los criterios de garantía de DQ, además de ser útil para la re-ingeniería, evaluación y mejora de los datos.
- Evaluar la conformidad de los datos con la legislación y / o requisitos.

La norma presenta un modelo genérico de calidad de datos, categorizando la DQ en 15 características o dimensiones consideradas desde dos puntos de vista:

- **Inherente:** Referido al grado en el cual las características de DQ tienen el potencial intrínseco para satisfacer las necesidades, usado bajo condiciones específicas. Desde el punto de vista inherente, calidad de los datos se refiere a los datos en sí, en particular, a los valores de datos del dominio y sus posibles restricciones.
- **Dependiente del sistema:** Referido al grado en el cual la DQ es enriquecida y preservada dentro de un sistema cuando es usado bajo condiciones específicas. Desde este punto de vista, la DQ depende del dominio tecnológico en el que los datos se utilizan, es alcanzado por las capacidades de los componentes de los SI, tales como: dispositivos de hardware (por ejemplo, para que los datos estén disponibles o para obtener la precisión requerida),

software de sistema (por ejemplo, software de copia de seguridad para lograr valorización) y otros.

La Tabla II.2 muestra las características de la DQ. El símbolo X indica la relevancia de las características en el modelo de calidad de los datos requeridos o a evaluarse a partir de los puntos de vista de inherente y/o sistema dependiente, sin dejar de mencionar que algunas características comparten su relevancia en los dos puntos de vista.

Características	CALIDAD DE DATOS	
	Inherente	Sistema Dependiente
Exactitud	X	
Compleitud	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

Tabla II.2. Características de un Modelo de Calidad presentado en la Norma ISO/IEC 25012.

La siguiente tabla muestra las definiciones dadas por la norma a cada una de las características observables para obtener DQ.

Dimensión	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. Ejemplo 1: Cuando hay un bajo grado de exactitud sintáctica, la palabra María se almacena como Marj. Ejemplo 2: Cuando hay un bajo grado de exactitud semántica, el nombre de Juan se almacena como George. Ambos nombres son sintácticamente correctos, por el dominio de referencia en el que residen, pero George es un nombre diferente relacionado con otra persona.
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. Ejemplo: Para una base de datos de empleados, disminuye la completitud, si los registros de algunos empleados no contienen el dato de número de teléfono en el caso de llamar por una emergencia.
Consistencia	El grado en el cual el dato tiene los atributos libres de contradicción y son coherentes con otros datos en un contexto específico de uso. Ejemplo: Fecha de nacimiento de un empleado no puede ser posterior a su fecha de contrato.
Credibilidad	El grado en el cual el dato tiene atributos considerados como verdaderos y creíbles por usuarios en un contexto específico de uso. Ejemplo: Los datos de un certificado de una organización independiente y de confianza deben ser creíbles.
Actualidad	El grado en el cual el dato tiene los atributos que son del período correcto en un contexto específico de uso. Ejemplo: La programación de una estación de ferrocarril se debe actualizar con la frecuencia necesaria para permitir a los pasajeros tomar un tren, aunque haya cambios en la hora prevista o cambios de plataforma.
Accesibilidad.	El grado en el cual se puede acceder al dato en un contexto específico de uso. Ejemplo 1: La gente que necesita soporte de tecnología o una configuración especial debido a alguna inhabilidad (incapacidad). Ejemplo 2: Los datos que son manejados en un monitor de pantalla no deben ser almacenados como una imagen para evitar problemas con su acceso posterior.
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. Ejemplo: Un crédito bancario debe cumplir con las leyes y normas específicas.
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. Ejemplo: Datos relacionados con la información personal o confidencial como los de salud, deben ser accedidos sólo por usuarios autorizados, o deben ser registrados en forma codificada (cifrada).
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. Ejemplo: El uso de más espacio que el necesario para almacenar datos puede causar pérdida de almacenamiento, memoria y tiempo.
Precisión	El grado en el cual el dato tiene atributos que son exactos o que proporcionan la discriminación en un contexto específico de uso. Ejemplo: Una precisión de 5 decimales permite diferentes funcionalidades en lugar de una precisión de dos decimales.
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. Ejemplo: Las administraciones públicas deben mantener una bitácora con información de acceso de usuarios y de transacciones importantes como lo es el ingreso de datos confidenciales.
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. Ejemplo: Para representar a un estado (dentro de un país), un acrónimo estándar es más comprensible que un código numérico.
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. Ejemplo: Los datos deben estar disponibles tanto para la operación, dentro de una organización, como en una copia de seguridad.
Portabilidad	El grado en el cual el dato tiene los atributos que le permiten ser instalado, substituido o movido de un sistema a otro conservando la calidad existente, en un contexto específico de uso. Ejemplo: disponer de funcionalidades de importación y exportación de la información que opera en un sistema.
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. Ejemplo: Cuando un dispositivo de almacenamiento tiene una falla, los datos almacenados en el dispositivo pueden ser recuperados.

Tabla II.3. Descripción de Dimensiones de Calidad según Norma ISO/IEC 25012.

La Norma ISO/IEC 25012 define cada una de las características presentadas en la Tabla II.3.

II.1.5 Conjugando Elementos

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

II.2 Método DeWIQ

DeWIQ es un método iterativo que tiene por objetivo guiar el desarrollo de un SI con requisitos centrados en la DQ.

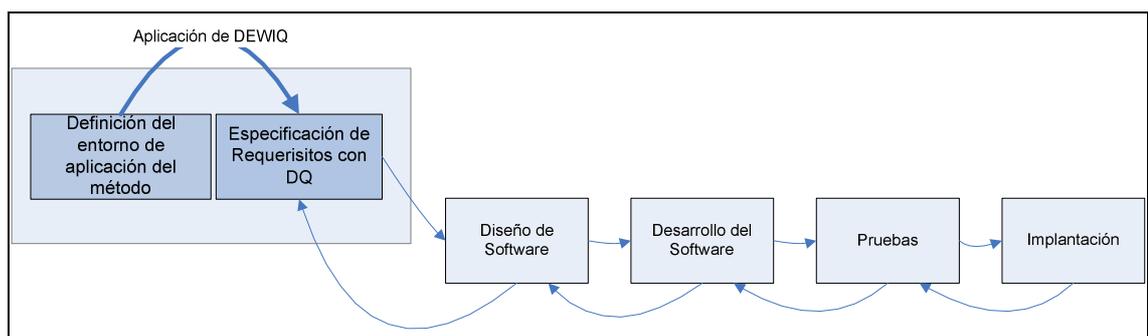


Figura II.1. DeWIQ y el proceso genérico de desarrollo de software.

La Figura II.1 muestra, de forma gráfica, una visión del proceso genérico de desarrollo de un SI (Pressman, 2005) destacando dónde es aplicable DeWIQ en esta primera versión. La idea del método es que puede ser aplicado junto con cualquier modelo o proceso de desarrollo de un SI.

A continuación se describe DeWIQ que, como lo muestra la Figura II.2, está orientado a la etapa de especificación de requisitos. DeWIQ es iterativo y se recomienda aplicarlo mientras existan requisitos por definir y sea necesario determinar las necesidades de DQ en los PI's involucrados en el producto del SI a desarrollar.

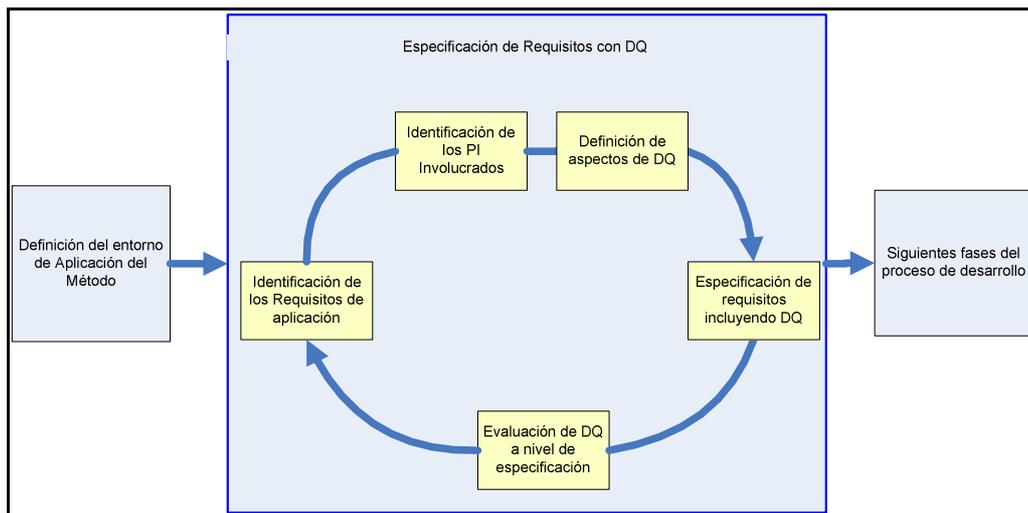


Figura II.2. Aplicación de DeWIK en la etapa de Especificación de Requisitos.

A continuación se detallan cada una de las etapas del método y las actividades involucradas en ellas:

II.2.1 Etapa 1: Definición del entorno de aplicación de DeWIK

Esta etapa, que no es parte del proceso iterativo, consiste en definir todos los elementos necesarios para la aplicación del método. El objetivo principal de ella es determinar quién dirigirá o será responsable de la aplicación de DeWIK y definir las actividades de capacitación sobre DQ necesarias para los involucrados en la especificación de los requisitos de la aplicación. Contempla 3 actividades: 1) Denominación y Formación del Gestor de Calidad de Datos, 2) Definición de Roles, y 3) Entrenamiento del DT en aspectos relacionados con DQ y el método DeWIK. A continuación se describe cada una de ellas:

a) *Denominación y Formación del Gestor de Calidad de Datos:* El Gestor de Calidad de Datos (desde ahora en adelante DQM) es el responsable de gestionar la aplicación de DeWIK a lo largo de cada una de las etapas del proceso de desarrollo de software empleado, siguiendo las tareas y directrices impuestas por el método, o bien adaptándolas a la realidad concreta del proyecto en desarrollo.

Se debe designar a la persona que será la encargada de gestionar la DQ durante el desarrollo del SI. Para esto, se pueden dar las siguientes modalidades:

- El DQM, puede ser un integrante del DT o no. En ambos casos, deberá trabajar estrechamente con el resto de los integrantes del DT.

- Si el DQM ha sido designado por primera vez para desarrollar este rol, deberá instruirse en aspectos relacionados con DQ y en el método DeWIQ. El presente Anexo está hecho para este propósito.
 - Si el DQM ya ha desarrollado este rol en proyectos previos, probablemente ya no requerirá mayor capacitación y al contrario, mejorará la aplicación de DeWIQ en base a su experiencia previa en el desarrollo de otros SI centrados en la DQ.
- b) *Definición de Roles:* Esta actividad consiste en determinar las personas que estarán involucradas, directa o indirectamente, en las actividades relacionadas con la DQ durante todo el desarrollo de la aplicación y el rol que éstas tendrán. Los roles relevantes son:
- *Integrante del DT:* Encargado de tareas propias del desarrollo como puede serlo el ingeniero de requisitos, analista, arquitecto de software, programador, tester, etc. En las diversas fases de aplicación de DeWIQ, cada uno de los integrantes del DT deberá desarrollar tareas relacionadas con DQ.
 - *Usuario de la Aplicación:* Individuo que tiene un perfil específico dentro de la organización y en la aplicación a desarrollar, ya sea proporcionando información, operando directamente u ocupando los servicios de información que el SI entregará.
- c) *Entrenamiento del DT en aspectos relacionados con DQ y el método DeWIQ:* Consiste en dar a conocer al DT temas relacionados con DQ y una inducción de DeWIQ para que sea aplicado de manera adecuada. Las actividades de esta etapa las realiza el DQM e incluye lo siguiente:
- Definición de los conceptos básicos que el equipo debe dominar respecto de DQ, por ejemplo: producto de información, calidad de datos, característica de calidad de datos y roles asociados a la calidad de datos.
 - Definición del Modelo de DQ a aplicar: Se sugiere que se use la norma ISO/IEC 25012, la cual puede ser complementada con otro modelo que se ajuste de mejor forma al tipo de SI que se desarrollará. El objetivo de lo anterior es que se identifiquen las características de DQ que serán de interés considerar en el SI a desarrollar.
 - Asignación de roles de usuario DQ: Consiste en reclasificar los roles de usuario de la aplicación a los roles clásicos de DQ como son consumidor, custodio y productor de datos. Para la aplicación del método en sus actividades posteriores, se considerará el rol de usuario DQ consumidor de datos por ser quien usa la información y el de productor de datos quien usará las reglas incorporadas a la producción de la información con calidad y que además requiere, en ocasiones, recuperar un PI para alimentar como información de

entrada a otro PI, entonces el PI consumido debe considerar ciertas características de DQ específicas de acuerdo al contexto. El rol de usuario DQ custodio de datos no se considerará, en esta etapa de especificación de requisitos, por no tener influencia directa en la DQ. En una etapa posterior a la especificación de requisitos se re-evaluarán la participación de los roles de usuario DQ.

- Inducción del método DeWIQ: Esta actividad tiene como propósito proporcionar conocimientos relacionados con el funcionamiento y modo de implementación y aplicación del método durante el proceso de desarrollo del SI. Este Anexo tiene dicho objetivo.

II.2.2 Etapa 2: Definición de los requisitos de la aplicación con DQ

Esta etapa tiene por objetivo definir los requisitos de del SI a desarrollar, incorporando aspectos de DQ a considerar en el desarrollo. Esta etapa y las subetapas que la componen se desarrollan de forma iterativa, hasta conseguir la especificación completa de los requisitos de la aplicación con DQ. Esta fase se compone de las siguientes subetapas:

a) *Identificación de los requisitos de la aplicación:* Aquí el DT debe usar sus procedimientos habituales para interactuar con los usuarios y captar los requisitos, objetivos y funcionalidades que ellos requieren respecto del SI a desarrollar. Sólo se exige que el artefacto usado para esta actividad, permitir incorporar luego, a información de sus PI's involucrados.

b) *Identificación de los PI's involucrados con los requisitos de la aplicación:* Una vez que se han determinado los requisitos, o al menos alguna parte de ellos, corresponde identificar los PI's relacionados con cada uno de ellos. Para esto, y usando el entrenamiento previo, el DT, guiado por el DQM, deberá analizar cada requisito en busca de PI's. Como estrategia de identificación de PI's, DeWIQ provee una lista de control (ver Figura II.3), donde por cada uno de los requisitos se responden interrogantes que ayudan a identificar los PI's asociados.

Lista de Control para identificar PI's				
Interrogante	Confirmación		Negación	
	Si	Acción	No	Acción
¿La implementación del requisito implica la generación de información de salida no analizada aún en esta lista de control?	<input type="checkbox"/>	Pase a la siguiente interrogante.	<input type="checkbox"/>	Finalice la determinación de PI's en el requisito analizado, pase al siguiente requisito y aplíquelo esta lista.
¿Hay información de salida es un informe?	<input type="checkbox"/>	Considere el informe como un PI. Vuelva a la primera interrogante.	<input type="checkbox"/>	Pase a la siguiente interrogante.
¿Hay información de salida es un formulario?	<input type="checkbox"/>	Considere el formulario como un PI. Vuelva a la primera interrogante.	<input type="checkbox"/>	Pase a la siguiente interrogante.
¿Hay información de salida es un archivo de texto?	<input type="checkbox"/>	Considere el archivo de texto como un PI. Vuelva a la primera interrogante.	<input type="checkbox"/>	Pase a la siguiente interrogante.
¿Hay información de salida es un documento?	<input type="checkbox"/>	Considere el documento como un PI. Vuelva a la primera interrogante.	<input type="checkbox"/>	Pase a la siguiente interrogante.
¿Queda información de salida pendiente de clasificar como un PI?	<input type="checkbox"/>	Clasificar y agrupar la información de acuerdo al tipo(s) de información de salida y considerarla como PI's. Vuelva a la primera interrogante.	<input type="checkbox"/>	Vuelva a la primera interrogante.

Figura II.3. Lista de control para identificar PI's.

Al seguir la lista de control, las respuestas a las interrogantes irán identificando los PI's asociados al requisito. Cuando se llega a la última pregunta, el encargado de clasificar la información debe conocer la definición de un PI y el dominio del problema para realizar un correcto análisis de esta información y agruparla adecuadamente, si es necesario el DQM debe asesorarlo.

La lista de control provista puede ser ampliada y mejorada por el DQM de acuerdo a su experiencia en la aplicación de DeWIQ. Por último, algunas consideraciones adicionales sobre la identificación de los PI's:

- Aquellos PI's definidos y que contienen sólo un atributo, evaluar la posibilidad de incorporarlos en otro PI con el cual esté relacionado, siempre y cuando su funcionalidad pueda ser absorbida.
- Aquellos PI's que comparten todos sus atributos deben ser fusionados en un sólo PI.
- Aquellos PI's cuyo atributos son un subconjunto de otro PI, deben quedar absorbidos en el PI que contenga el conjunto total de atributos.

Una vez definidos los PI's, se deben documentar en el artefacto PI's de la Aplicación presentado en la Figura II.4.

PI's de la Aplicación					
Id PI	Nombre PI	Propósito	Descripción	Roles de Usuario de la Aplicación	Requisito
PI-1					
PI-2					
...					
PI-n					

Figura II.4. Artefacto de Identificación de PI's.

El artefacto de la Figura II.4 documentará todos los PI's de la aplicación a desarrollar y que han ido siendo identificados mediante el análisis de los requisitos. Por cada PI de la aplicación el artefacto contendrá los siguientes datos:

- Identificador: corresponde a la identificación que se le da al PI. Su formato es definido por el DT y el DQM.
- Nombre del PI: corresponde al nombre que lleva el PI, el cual debe ser descriptivo y preciso.
- Propósito: consiste en el objetivo que tiene el PI.
- Descripción: consiste en los atributos que lo componen.
- Roles de usuario de la aplicación: corresponde a los roles de usuario de la aplicación con los que quedarán relacionados.
- Requisito: asociados a los PI's de la aplicación.

Además, esta tabla puede ser complementada con información adicional que sirva de apoyo en caso de ser requerido, como por ejemplo fecha de actualización, número de versión, entre otros.

Luego se debe completar información en el artefacto mostrado Matriz de relación de Requisitos y PI's (ver Figura II.5), la que establece la relación entre los requisitos de la aplicación y los PI's identificados.

Matriz de relación Requisitos y PI's				
Requisito	PI-1	PI-2	...	PI-n
1.-	X			
2.-	X	X		X
...				
n.-	X	X		

Figura II.5. Artefacto Matriz de Requisitos de aplicación y sus PI's.

Esta matriz tiene como propósito la visualización de la interacción que tienen los requisitos del SI con los PI's definidos, de modo que puedan establecerse algunas consecuencias relevantes como:

- Qué PI's de la aplicación pueden ser impactados si cambia un requisito de la aplicación.
- Qué requisitos de la aplicación pueden ser afectados si se modifica un PI.
- Qué PI's son prioritarios de construir, en caso de una implementación por etapas.
- Qué requisitos cruzan con mayor frecuencia a los PI's. Estos deben ser clasificados como críticos dentro del SI y se les debe dar interés para incorporar características de DQ.

Cuando los PI's estén asociados a sus requisitos, se usará el Artefacto de Identificación de PI's, en donde los requisitos quedaron relacionados con los roles de usuario de la aplicación y se reclasificará esta asignación en roles de usuario DQ (Ver Figura II.6).

Interacción de PI's con Roles de Usuario de la Aplicación y DQ			
Productos de Información	Roles de Usuario de la Aplicación	Rol de usuario DQ	
		Consumidor de Datos	Productor de Datos
PI-1	Usuario I de Aplicación	X	
	Usuario II de Aplicación		X
	...		X
	Usuario n de Aplicación		X
PI-2	Usuario I de Aplicación	X	
	Usuario II de Aplicación	X	
	...		X
	Usuario n de Aplicación		
PI-n	Usuario I de Aplicación	X	X
	Usuario II de Aplicación		
	...		
	Usuario n de Aplicación	X	

Figura II.6. Artefacto de Interacción de PI's con Roles de Usuario de la Aplicación y DQ.

De acuerdo a la relación que tengan los roles de usuario de la aplicación con los PI's definidos, esta matriz muestra el comportamiento de los roles de usuario de la aplicación en función de los roles de usuario DQ por cada PI. En la Figura 6 se presentan como ejemplo tres PI's, su relación con tres roles de usuario genéricos de la aplicación y su reclasificación en los roles de usuario DQ de acuerdo al modo de interacción que tienen con el PI.

Esta tabla, por cada PI, permite visualizar lo siguiente:

- Qué roles de usuario de la aplicación interactúan más frecuentemente en su función de rol de usuario DQ: este rol de usuario tendrá mayor interacción con el PI, por lo tanto su participación en el proceso de desarrollo e implementación del SI es relevante.
- Qué rol de usuario DQ es el más utilizado: permite analizar qué características de DQ deben incorporarse al PI para que este rol sea ejecutado de manera adecuada.
- Qué usuarios de aplicación interactúan con un cierto PI: entre estos usuarios se debe determinar a quienes se les consultará por las características que debe incluir el PI. El rol de consumidor de datos usará el producto de información con DQ y el productor de datos es a quien se le imponen las exigencias para ingresar la información que posteriormente será parte del proceso de generación del PI.
- Qué PI tiene mayor interacción con el usuario en función de su rol de usuario DQ: estos PI's deben ser clasificados como críticos, es decir, jugarán un papel esencial en el SI que se desarrollará.

Finalizando esta actividad de identificar los PI's, se registrará en un artefacto llamado matriz de relación de PI's la dependencia existente entre los PI's (Ver Figura II.7). Esta dependencia está dada porque un PI puede originar a otro, por lo tanto sus características de DQ también están estrechamente relacionadas.

Matriz de Relación de PI's					
Productos de Información	PI-1	PI-2	PI-3	...	PI-n
PI-1		>	>		
PI-2			>		
PI-3					
...					...
PI-n	>				

Figura II.7. Artefacto de Matriz de Relación de PI's.

Las flechas presentadas en el artefacto, indican que el PI de la fila izquierda origina al PI de la columna de derecha.

Este artefacto ayuda a una visualización global de los PI's, dando el apoyo al DQM y al equipo de desarrollo para tomar decisiones, descubrir problemas y solucionar de manera rápida en esta etapa de desarrollo antes de la fase de construcción. SI y como afectan los cambios a los elementos de algunos de ellos. Concretamente, cuando un PI debe ajustarse a una determinada característica, se debe ubicar el o los PI's que lo generan porque tienen los atributos involucrados con la característica de DQ, ya que estos PI's (los atributos que correspondan) deben reunir esa misma

característica. De lo contrario no se podrá asegurar la característica de DQ en el PI generado por el SI.

c) *Definición de los aspectos de DQ:* Por cada PI identificado se deberán definir sus características de DQ relevantes. Puede darse el caso de haber identificado una gran cantidad de PI's y por consiguiente una labor extensa en identificar sus características de DQ, pero para asegurar que los requisitos del SI estén realmente centrados en la DQ es necesario que todos los PI's sean sometidos a este proceso.

En esta etapa se deben elaborar encuestas que entregan información sobre la importancia que los usuarios asignan a cada característica de DQ por cada PI analizado usando el modelo de DQ ISO/IEC 25012 ó el que se haya elegido como referencia. Esta encuesta debe tener una escala de puntuación claramente explicada y manejar no más de cinco valores para luego visualizar y manejar en qué rangos de valores se mueve la tendencia de los encuestados en la preferencia de las características de DQ.

Resulta lógico pensar que el usuario siempre desea que el SI a desarrollar debe tener calidad de datos, por lo tanto las interrogantes las contestará con una probable puntuación máxima. Sin embargo, el usuario deberá tener presente que a mayor calidad incorporada mayor será el uso de recursos incorporados para lograr estos objetivos. Es por ello que por cada interrogante realizada, en la misma encuesta, se debe indicar a grandes rasgos que trabajo adicional implica la incorporación de la característica de calidad de datos consultada para que el usuario tenga una visión de los recursos extras que significa agregar esa característica de DQ.

Los usuarios, en su función de rol de usuario DQ, deben indicar en la encuesta qué valor de importancia debe tener cada una de las características de DQ en cada PI que se le consulte. Esta asignación de valores debe estar guiada de acuerdo a interrogantes lo suficientemente claras para asegurar la comprensión del significado y ámbito de cada característica de DQ en los PI's que necesitan ser evaluados.

A continuación se debe consolidar la información contenida en las encuestas y realizar una tabla resumen (Ver Figura II.8) entre los roles de usuario DQ y sus características de DQ.

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

Figura II.8. Artefacto de Características de DQ y su importancia en Roles de Usuario DQ por PI.

Este resumen permite visualizar claramente cuáles características de dicho modelo son relevantes en cada PI y en qué grado. Cuando un rol de usuario DQ es compartido por más de un rol de usuario de la aplicación se debe indicar el puntaje promedio resultante de las valoraciones indicadas por cada uno de ellos en la característica de DQ en análisis.

Para obtener el valor de la ponderación promedio de cada característica de DQ, se debe realizar la siguiente fórmula:

$$Ponderación\ Promedio(CaracterísticaDQ) = \frac{(2CoD + PrD)}{3}$$

Donde,

- CoD: Valor promedio de puntuación, dado por el rol de usuario DQ consumidor de datos a la característica de DQ analizada.
- PrD: Valor promedio de puntuación, dado por el rol de usuario DQ productor de datos a la característica de DQ analizada.

Como se observa en la fórmula, se da una mayor ponderación al rol de usuario DQ consumidor de datos al ser el rol de mayor interacción con los PI's identificados en un SI y quien determina si los datos son útiles o no.

Si bien la calidad de los datos se define en términos del uso, esta fórmula debe ser aplicada a las características que son dependientes del sistema, dado que sobre ellas el productor de datos tiene una percepción más clara de la DQ en sus tareas de ingreso y actualización de datos. Además nivela las expectativas del consumidor de datos (suelen ser superiores). Para el resto de las características sólo se toma en cuenta la opinión del consumidor de datos, estas se encuentran

marcadas en la Figura II.8 con un “N/A”, es decir, no aplica porque no obtienen puntaje del rol de usuario DQ proveedor de datos.

Se ocupa el valor promedio por ser el equilibrio que representa a todos los valores observados de medición dada a la característica de DQ. Además existe una compensación implícita entre la DQ que los usuarios requieren y su costo de alcance.

Estos valores permiten establecer qué características de DQ deben ser abordadas en un SI a desarrollar desde el punto de vista de los roles de usuario DQ consumidor y productor de datos.

d) *Especificación de requisitos incluyendo DQ.* Cuando ya se conocen las necesidades de DQ de los PI's, entonces se debe complementar con ellas la especificación de los requisitos del SI. Para esto 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, quedando un margen en referencia a la puntuación máxima definida por la escala de valores. A mayor margen se considerarán mayor número de PI's y características de DQ a incluir.

El puntaje de corte debe quedar definido, de acuerdo al criterio del DT (por tener la visión global de las soluciones que aborda el SI a desarrollar) y del DQM (por ser el experto en la aplicación de DeWIQ), debido a que la planificación de los tiempos de desarrollo del SI no deben verse excesivamente aumentada por la incorporación de las características de DQ a los requisitos del SI. El usuario no participa en esta definición de puntaje corte por ser un rol compartido por más de un individuo y sus percepciones de DQ probablemente difieren entre uno y otro, de acuerdo al uso que le darán a la información.

En base al análisis de los resultados obtenidos de la encuesta, se debe incluir como parte de la especificación de requisitos un listado de acciones tendientes a lograr el alcance de las características de DQ (ver Figura 9), considerando no alterar la definición de los requisitos del SI a desarrollar.

Estas acciones son específicas de acuerdo al PI que pertenecen y deben quedar ordenadas de mayor a menor, de acuerdo al puntaje obtenido por cada característica de DQ seleccionada. Las características de DQ seleccionadas a incorporar deben ser aplicadas a los atributos que componen el PI. Esta aplicación de DQ quedará en función de las acciones que fueron definidas para su mejora.

Listado de acciones en PI's		
Producto de Información	Características de DQ	Acciones
PI-1	Característica 1 de DQ	Acción 1 Acción 2 Acción 3
PI-1	Característica 2 de DQ	Acción 1
PI-2	Característica 3 de DQ	Acción 1 Acción 2
PI-2	Característica 1 de DQ	Acción 1
PI-3	Característica 4 de DQ	Acción 1 Acción 2 Acción 3 Acción 4
PI-3	Característica 1 de DQ	Acción 1
PI-3	Característica 2 de DQ	Acción 1 Acción 2

Figura II.9. Listado de acciones en PI's.

La Figura II.9 muestra un resumen de los PI's seleccionados de acuerdo al puntaje de corte definido y su listado de acciones a realizar. Se tomó como referencia una encuesta con una puntuación máxima de 5 puntos y se decidió un puntaje de corte de 3 puntos. Las características de DQ indicadas en la tabla quedaron seleccionadas de acuerdo a la referencia descrita y sus acciones relacionadas tienen por objetivo incorporar la característica de DQ a los requisitos definidos del SI a desarrollar. El valor ponderado representa al valor promedio dado a la característica de DQ, en base a los resultados obtenidos en la encuesta.

En cada caso de uso se deben utilizar las acciones más apropiadas o más afines de acuerdo a la utilización del PI en la especificación de requisitos y a los PI's que se generan a partir del PI en que se ejecutará esta acción. Cuando un PI debe ajustarse a la característica 1, entonces debemos ubicar el o los PI's que lo originan (porque tienen los atributos involucrados con la característica, los mismos o las fuentes para generar aquellos), ya que estos PI (los atributos que correspondan) deben reunir esa misma característica. De lo contrario no se podrá asegurar la característica en el PI generado por el SI.

A continuación, se debe complementar la especificación de requisitos con las características de DQ definidas como importantes en la especificación de requisitos. Por ejemplo, al utilizar los casos de uso, en dicho documento se debe especificar en la sección que indica los PI's, qué características debe contener cada PI.

Caso de Uso: <<Nombre de caso de uso>>
ID: <<Identificador de Caso de Uso>>
Breve Descripción: <<Descripción máxima de dos líneas>>
Actores Principales: <<Actores Principales implicados en el caso de uso>>
Actores Secundarios: <<Actores Secundarios implicados en el caso de uso>>
Productos de Información (PI's): <<Nombre, Descripción, Características DQ>>
Precondiciones: <<Estado del sistema antes de que el caso de uso pueda empezar>> 1. 2.
Flujo Principal: <<Pasos del caso de uso>> 1. El Caso de Uso comienza ... 2.
PostCondiciones: <<Estado del sistema después de que el caso de uso ha sido terminado>> 1. 2.
Flujos Alternativos: <<Lista de alternativas al flujo principal>>

Figura II.10. Casos de Uso con características de DQ.

La Figura II.10 muestra el documento de caso de uso, utilizado como referencia para especificar requisitos, indicando, en la sección de los PI's, las características de DQ que deben estar incluidas.

Este documento debe anexar el documento especificado en la Figura II.9, el que contiene las acciones para lograr las características deseadas de DQ por PI, seleccionando las pertinentes al caso de uso, para contar con una especificación de requisitos con elementos incorporados de DQ.

e) *Verificación de DQ a nivel de especificaciones:* Es la última actividad del ciclo, la que en términos generales determina la completitud de la especificación de requisitos en los siguientes aspectos relacionados con la DQ:

- Cada especificación de requisitos debe contener una referencia a cada uno de los PI's 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 PI's seleccionados del SI deben estar asociados al menos a una especificación de requisitos.

En términos particulares, se debe verificar que los artefactos incluidos en cada actividad contemplada en DeWIK hayan sido debidamente completados.

Como instrumento de evaluación se ha definido una lista de control (ver Figura 11), en la que el DT debe cotejar lo indicado anteriormente.

Lista de Control de Verificación de Características DQ				
Interrogante	Confirmación		Negación	
	Si	Acción	No	Acción
¿La especificación del requisito incluye PI's?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	Tomar la siguiente especificación de requisitos y comience nuevamente esta lista desde la primera interrogante.
¿En la especificación del requisito se han definido las características de DQ asociadas al PI?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	En caso de que el PI tenga sus características de DQ definidas, complete el artefacto de "especificación del requisito" con las características asociadas a su PI y luego pase a la siguiente interrogante. En caso contrario, deje esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. Si el requisito no tiene más PI's , finalice el seguimiento de esta lista y pase al siguiente requisito.
¿El PI se ha identificado con su información en el artefacto "Identificación de PI's"?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	Dejar esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. . Si el requisito no tiene más PI's , finalice el seguimiento de esta lista y pase al siguiente requisito.
En el artefacto "Matriz de Requisitos de aplicación y sus PI's" ¿está el PI asociado a su requisito?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	Dejar esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. Si el requisito no tiene más PI's, finalice el seguimiento de esta lista y pase al siguiente requisito.
En el artefacto "Interacción de PI's con Roles de Usuario de la Aplicación y DQ" ¿está el PI asociado a su rol de usuario de aplicación y DQ?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	Dejar esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. Si el requisito no tiene más PI's, finalice el seguimiento de esta lista y pase al siguiente requisito.
¿Se aplicó la encuesta de características de DQ para el PI?	<input type="checkbox"/>	Pase a la siguiente Interrogante.		Deje esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. En caso contrario, finalice el seguimiento de esta lista y pase al siguiente requisito.
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?	<input type="checkbox"/>	Pase a la siguiente Interrogante.	<input type="checkbox"/>	Dejar esta especificación en estado "Pendiente", tome el siguiente PI de la especificación de requisitos y comience nuevamente esta lista desde la segunda interrogante. Si el requisito no tiene más PI's, finalice el seguimiento de esta lista y pase al siguiente requisito.
¿Se han definido las acciones para el PI en el artefacto de listado de acciones en PI's?	<input type="checkbox"/>	Si el requisito tiene más PI's, tome el siguiente y vuelva a la segunda pregunta de esta lista. En caso contrario, finalice este seguimiento y pase al siguiente requisito.	<input type="checkbox"/>	Coloque el requisito en estado "Pendiente". Si el requisito tiene más PI's, tome el siguiente y vuelva a la segunda interrogante de esta lista. Si el requisito no tiene más PI's , finalice el seguimiento de esta lista y pase al siguiente requisito.

Figura II.11. Lista de control de verificación de características DQ.

La Figura II.11 presenta una lista de control que debe aplicarse a cada requisito especificado, en forma secuencial de acuerdo a la acción indicada por respuesta señalada. La idea es tomar un requisito, revisar sus PI's asociados y completar su documentación en caso de que corresponda. Luego, tomar otro requisito y así sucesivamente, hasta finalizar los requisitos especificados del SI a desarrollar.

A continuación se debe hacer una revisión inversa, es decir, revisar los PI's seleccionados y que al menos estén en una especificación de requisitos, para esta revisión se puede apoyar con el artefacto de "Identificación de PI's" y en el de "Matriz de Requisitos de Aplicación y sus PI's", el primero por aportar la información de los PI's identificados y el segundo por proporcionar la información de los PI's cruzados con los requisitos del SI. En caso de que haya algún PI que no esté contenido en las especificaciones, debe ser incluido en las especificaciones de requisitos que corresponda de acuerdo a su definición y contexto.

Para aquellos requisitos en que no fueron marcados con estado "Pendiente", marcarlo con estado "Completado".

En esta última actividad del ciclo, es muy importante tener en cuenta, y con el fin de evitar confusiones, solamente se revisarán aquellas especificaciones en que se necesita que sus requisitos estén centrados en la DQ, dejando de lado aquellos que fueron descartados durante la aplicación de DeWIQ.

Como resultado de la verificación de las especificaciones de requisitos y PI's del SI, las especificaciones serán clasificadas en los siguientes estados:

- **Completo**, 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 del SI.
- **Pendiente**, lo que significa que la especificación del requisito aún tiene definiciones incompletas de DQ, por lo cual debería volver a entrar en el ciclo de DeWIQ.

II.2.3 Próximas Iteraciones

Es probable que al término de una iteración, probablemente, producto del mismo análisis de DQ, resultaron los siguientes escenarios: (a) requisitos con definiciones pendientes de completar, es decir, con estado pendiente, (b) nuevos requisitos y (c) requisitos cuyos PI's no fueron considerados para incorporar DQ en la primera iteración y ahora si se desea incorporarle. Para esto es necesario iterar nuevamente el método, y repetir las distintas actividades del método partiendo

de la documentación emitida en la última iteración y tratando de completar cada una de ellas de manera definitiva con las definiciones faltantes.

A continuación se explica, como debe llevarse a cabo cada una de las actividades del método en el contexto de las próximas iteraciones de DeWIQ.

a) *Especificación de Requisitos:* Se deben emitir los casos de uso necesarios para especificar los requisitos pendientes (o en su defecto el documento usado para la especificación de requisitos) y completar aquellos casos en que se necesita actualizar su definición, completando sus escenarios, sus roles de usuario y en especial la información de sus PI's asociados y sus características.

b) *Identificación de PI's involucrados:* En caso de no tener PI's identificados, se deben seguir las listas de control de igual forma que en la primera iteración. Tener presente que este artefacto puede ser completado en cada iteración de acuerdo a la interacción que tenga el DT con los requisitos de un SI y los PI's que pueden ser identificados.

A continuación, debe completarse el artefacto de identificación de PI's con la nueva información de los PI's. Al igual que el documento anterior, este artefacto también puede ser completado con columnas que representen de mejor forma al PI, de acuerdo al contexto especificado del SI.

Con esta información, el DT está en condiciones para completar los artefactos de "Requisitos de aplicación y sus PI's y el de "Interacción de PI's con Roles de Usuario de la Aplicación y DQ", de acuerdo a la definición dada en los requisitos.

c) *Definición de los aspectos de DQ:* De igual forma que en la primera iteración, en esta instancia se debe decidir a qué PI's se les definirán sus características, decisión que se puede tomar con el apoyo de los artefactos completados en esta iteración. Esta decisión lleva a centrar en la DQ aquellos requisitos asociados a los PI's seleccionados en la DQ, en las actividades que vienen a continuación.

Para aquellos PI's en que no se han identificado sus características relevantes, se debe completar la encuesta relacionada con esta actividad. Y luego registrar sus resultados en el artefacto de "Características de DQ y su importancia en Roles de Usuario DQ".

La encuesta que ya fue aplicada en la primera iteración, es un artefacto que permite ser completado con más interrogantes de acuerdo al contexto en que se desenvolverán los PI's. Al tener una encuesta más refinada nos permite consultar con mayor claridad que características de DQ y que nivel se les debe proporcionar al PI en cuestión.

d) *Especificación de requisitos incluyendo DQ:* Consecuentemente, se debe definir un puntaje de corte en la encuesta realizada, para seleccionar aquellas características en que su resultado sea superior o igual al puntaje de corte definido. Se recomienda utilizar el mismo puntaje de corte que en la primera iteración, para mantener uniforme el nivel de calidad que se le quiere dar al SI a desarrollar, a menos que se quiera aplicar un criterio distinto como por ejemplo: “al aplicar una nueva iteración aumenta la calidad de los PI’s, pero también aumenta el costo de implementación. Por lo tanto se desea un puntaje de corte bajo para incorporar más características y otorgar realmente una calidad de datos esperada”.

Se deben definir un listado de acciones por cada una de las características de DQ, asociadas a los PI’s analizados. Al igual que en la primera iteración, estas acciones se pueden extraer de los ejemplos contenidos en el Anexo III, o bien incorporar de acuerdo a las experiencias del DT. Deben quedar ingresados en el mismo orden que en la primera iteración.

A continuación se deben completar las especificaciones de los requisitos con la información de sus PI’s asociados y sus características de DQ seleccionadas, de acuerdo a los criterios previamente establecidos.

e) *Verificación de DQ a nivel de especificaciones:* En esta última actividad de la iteración, se debe seguir la lista de control de verificación de características de DQ para cada uno de los PI’s analizados.

Se recomienda realizar dos a tres iteraciones del método como máximo para obtener requisitos centrados en la DQ y no aumentar demasiado el costo del proyecto.

Una vez que las especificaciones de requisitos y PI’s del SI haya alcanzado el nivel de completitud esperado, el DT se encuentra en condiciones de avanzar a las siguientes fases del desarrollo del SI. En dichas fases los requisitos de DQ, inmersos en los requisitos del SI, deberán ser modelados, implementados, verificados y validados, pero por ahora no serán abordados.

A partir de la experiencia en desarrollo de un grupo de expertos, se han definido diferentes acciones que ayudan a lograr un nivel adecuado de DQ por cada característica de DQ a considerar en el desarrollo de un SI, presentado en el siguiente anexo.

III Anexo III. Acciones que ayudan a lograr un nivel adecuado de DQ

a) Exactitud

- Definir reglas de validación
- Considerar estándar IEEE de requisitos para incorporar validaciones de este tipo.
- Acotar dominio de los datos de acuerdo a su tipo.
- Utilizar máscaras para ingreso de datos.
- Establecer unidades de medida de acuerdo al tipo de datos.
- Definir largos para tipo de datos.
- Realizar validaciones de tipo sintáctica.
- Controlar caracteres especiales.
- Para tipos de datos que los valores son definidos dar opción de sólo indicar sus valores posibles.
- Normalizar datos, subdividir en más datos para asegurar la exactitud. Ejemplo: Para el dato dirección dividir en calle, número, villa.
- Vista Previa de Información de ingreso.
- En informes realizar paginaciones.

b) Completitud

- Establecer niveles de ingreso. Por ejemplo: obligatorios, deseable, entre otros.
- Organizar datos en un formulario, de tal forma que los datos obligatorios queden juntos. Esto sirve para identificar fácilmente cuando hay datos incompletos de ingresar.
- Indicar aviso de datos que no cumple con ciertas normas u otras restricciones de carácter legal o de tipo de procedimiento administrativo.
- Revisar datos de salida y en base a resultados esperados, validar los datos obligatorios de entrada. Se puede establecer en una especie de estándar de requerimientos.
- No guardar un formulario hasta asegurar que esté completo.
- Incorporar estrategias de ingreso. Especie de resumen para informes o envío de información. Ejemplo: Artículos para revistas de investigación.
- Ayuda de ejemplo para el ingreso de información.

- c) Consistencia
- Establecer reglas de consistencia propias del negocio o de información histórica.
 - Validaciones de rango de acuerdo a reglas del negocio.
 - Realizar cuestionario para extraer reglas.
 - Proporcionar textos o ayuda que ayuden a identificar cuando el dato ingresado no es consistente.
 - Clasificar y ordenar los datos en un formulario o informe para que los datos sean consistentes.
- d) Credibilidad
- Ingresar responsable o autor del dato.
 - Ingresar fecha de ingreso o última actualización del dato.
 - Indicar fuente de origen de los datos.
 - Indicar estado del dato, a nivel de proceso.
 - Incorporar historial de transacciones.
 - Dar la opción al usuario de verificar el ingreso de sus datos. Por ejemplo a través de una vista previa.
 - Para datos que tienen un alto flujo de cambios, incorporar herramientas que aseguren que el dato es actual.
- e) Actualidad
- Revisar generación de fecha, a través de un servidor confiable.
 - Incorporación de estados del dato.
 - Asignar atributos de estados para PI. Validaciones de datos que sean “activos” o “vigentes”. Ejemplo: ventas de clientes “vigentes”.
 - Determinar por la frecuencia de transacciones, estados en los datos. Esto puede quedar en una regla de negocio.
 - Considerar caducidad en el dato.
 - Incorporación de métodos de almacenamiento de datos en repositorios históricos y de proceso actual, de acuerdo a intervalos de tiempo parametrizados.
- f) Accesibilidad
- 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”.
- Establecer formatos estándar para interpretar el dato en forma correcta.
- Ordenamiento de datos por criterios previamente establecidos.
- Incorporar Criterios de búsqueda que ayuden a tener un mejor acceso a la información.

g) Conformidad

- Incorporar recomendaciones a nivel de diseño, que serán traducidas como trigger o procedimiento almacenado a fin de facilitar el seguimiento, validación y actualización de los datos.
- Parametrizar reglas de validación.
- Seguir o definir estándares o convenciones del negocio, regulaciones o normativas existentes.
- Incorporar ventanas para confirmar la transacción con el usuario.

h) Confidencialidad

- Definir a nivel de negocio la confidencialidad de la información.
- Establecer perfiles y grupos de usuarios.
- Cierre de sesión de usuario
- Definir reglas de validación
- Encriptación de datos.
- Ocultar o mostrar información, dependiendo del perfil de usuario.
- Incorporación de estándares de seguridad y marcos de trabajo que ayudan a su cumplimiento.
- Construcción de un modelo de datos que permita categorizar el dato en confidencial.
- Niveles de confidencialidad a nivel de proyección (campos o columnas) y a nivel de selección (fila o registro).
- Incorporar estrategias de ingreso. Especie de resumen para informes o envío de información. Ejemplo: Artículos para revistas de investigación.
- Ayuda de ejemplo para el ingreso de información.

i) Eficiencia

- Envío de información a histórica dejando la base de proceso más liviana para transacciones relacionadas con la operación normal del sistema.
- Optimizar largo y tipo de datos.
- Utilizar compresión de datos.

- Correcta definición de Índices para accesos más rápido.
- Generar estadísticas de acceso a tablas.
- Reducir fragmentación externa e interna dentro de la base de datos.
- Usar listas, checkbox como una forma de ingreso de datos.
- Cargar datos predeterminados. Por ejemplo: fechas.
- Pantallas de fácil de ingreso de información. Poco uso del mouse e incorporar ingresos del teclado.
- Ayudar con dispositivos de accesos. Lector de código de barra.
- Para salida de información, usar formatos livianos. Por ejemplo: Para informes.
- A nivel de construcción de programas, eficiencia en escribir códigos, conexiones a BD, entre otros.
- Normalización del modelo de datos.
- Tener consultas de datos optimizadas para evitar demoras en tiempos de respuesta.

j) Precisión

- Establecer la precisión de acuerdo al negocio, al PI requerido y al tipo de usuario.
- Para datos de salida determinar de acuerdo al perfil de usuario o tipo de operación.
- Establecer unidades de medidas de acuerdo al tipo de datos.
- Para tipos de datos numéricos, determinar números de decimales
- Para tipos de datos de texto, determinar y validar un máximo de caracteres.
- Para tipos de datos de fecha/hora, incorporar máscaras para precisar su formato.

k) Trazabilidad

- Incorporar fuente de origen.
- Indicar usuario, dueño y responsable en la intervención del dato en distintos tiempos para un PI.
- Indicar cambio de un estado a otro.
- Incorporar opción de almacenamiento con fechas para tener un registro de transacciones de los datos.
- Incorporar una bitácora de transacciones de modo de hacerle un seguimiento a alguna información en particular.

l) Comprensibilidad

- Incorporar un patrón de diseño con estilos definidos y uniformes para todos los formularios e informes de la aplicación.

- Incorporación de etiquetados de textos claros.
- Incorporación de textos de ayuda.
- Para tipos de datos como fecha o que necesitan de un formato de específico incorporar una máscara.
- Indicar claridad en mensaje de error. Ejemplo “campo x no está del rango 999”.
- Indicar campo con error en un formulario de ingreso.
- Incorporar glosas descriptivas (más de una) para identificar claramente el campo, dependiendo del tipo de usuario para ingreso y/o consulta de la información. Ejemplo: Descripción corta o descripción larga. Nombre científico de especie o descripción de especie.
- Definir unidades de medida para mostrar el dato de acuerdo al contexto. Por Ejemplo: Largo de decimales, datos abreviados, entre otros.
- Reportes gráficos con identificación de colores e indicar una leyenda que indique sus significados.
- Uso explícito de los datos de acuerdo a perfiles o roles de usuario.

m) Disponibilidad

- Establecer Niveles de disponibilidad para la información y acciones para cada uno de los niveles.
- Relacionar los niveles definidos con los PI's.
- Definir rango de fechas parametrizado por la organización para carga de información a histórico.
- Definir carga y actualización de información a bases de datos y repositorios. Esto depende del negocio, puede ser muy frecuente (reposición en supermercado) o de baja frecuencia (información personal de clientes).
- Dar la opción de trabajar en modo on-line y off-line, este último cuando problemas de conectividad en red de datos.
- Tener un control adecuado de versiones del software y que quede de fácil acceso para su actualización por parte de los usuarios o área de soporte.
- Contar con una adecuada política de resguardo de la información, antivirus actualizados y soporte en línea.
- Definir de una arquitectura sencilla, modular y extensible, que permita configuraciones avanzadas de compartición, balanceo de carga e integración de la información.
- Desarrollar esquemas de replicación que se adapten a los requisitos impuestos en este escenario.

- Proponer optimizaciones a la alta disponibilidad basadas en técnicas de detección y diagnóstico de errores, y aprendizaje supervisado.

n) Portabilidad

- Definir estándares de portabilidad y ocupar medios que apoyen esta tarea. Por ejemplo: servicios web, vistas, etc.
- Dar la opción de importar y exportar datos en formatos tradicionales de utilizar. Por ejemplo, planilla electrónica, archivo plano de texto, archivos de datos, entre otros.
- Considerar los formatos que necesitan otros sistemas al hacer uso de los PI's.
- Para los PI's usar grillas con opción de exportación.
- Para reportes usar opción de exportación.
- Establecer qué conjunto de datos son portables.
- Usar pre-formatos para planillas electrónicas. Exporta, ingresa y luego importa planilla para ingreso de información en el sistema.
- Para carga masiva de información. Por ejemplo: uso de batch-input.

o) Recuperabilidad

- Para una validación de datos, limpiar sólo los datos erróneos para ingresarlos nuevamente.
- Al cerrar una pantalla, almacenar datos ingresados o indicar transacciones pendientes de completar.
- Incorporar tipo de auto-guardado.
- Incorporar estados de continuación para ingreso de información. Por ejemplo: Estado Borrador.
- Establecer políticas de seguridad. Ejemplo: Fallas de discos, fuente poder, respaldos, réplicas, entre otros.
- Para datos que se necesitan eliminar incorporar la funcionalidad de "eliminación lógica", es decir, que el dato quede en estado eliminado y dar la opción de recuperarlo en caso de que sea necesario.
- Dar la opción de confirmación o anulación de transacción (Commit y Rollback)

Estas acciones son específicas de acuerdo al tipo de PI y a la aplicación de software a desarrollar, cuyo propósito es mejorar los PI's de acuerdo a las características de DQ que se desean incorporar. Las acciones deben quedar establecidas en las especificaciones de los requisitos incluyendo DQ y su implementación debe ser realizada en la fase de construcción del sistema.

El siguiente anexo tiene presenta un patrón de encuesta que será aplicado a los usuarios de DQ para determinar el nivel de importancia que le dan a cada una de las características de DQ en un PI determinado.

IV Anexo IV. Encuesta de DQ

Encuesta de DQ en PI XXX ⁶

Esta encuesta tiene por objetivo obtener los requisitos de calidad de datos que debe tener la información contenida en el PI XXX.

A continuación se indica un listado de interrogantes, de acuerdo al contexto especificado de la característica de calidad de datos indicado en la columna izquierda, señalando además el costo adicional, en el desarrollo e implementación del PI, que implica incorporar la característica de calidad de datos analizada.

Responder por favor, indicando valores pertenecientes al rango de 1 a 5, de acuerdo a lo siguiente:

- 1: No es importante
- 2: Poco importante
- 3: Indiferente
- 4: Importante
- 5: Muy Importante

DQ	Pregunta	Costo Implementación	Respuesta
E X A C T I T	Para el ingreso de información, ¿deben existir validaciones y formatos especiales de ingreso de información para asegurar que los datos contenidos en XXX sean exactos?	Estas implementaciones aumentan el tiempo de ingreso de los datos y obliga el uso del Mouse de acuerdo a los formatos que quedarán preestablecidos, dejando más lenta la interacción del usuario con la pantalla de ingreso de la información para XXX. Además, se debe definir qué datos deben contener este formato.	
U D	¿Se debe contar con utilidades de visualización de ingreso de la información para XXX. Por ejemplo vista previa, reportes de consulta, entre otros?	Esto significa un desarrollo adicional en construir estas utilidades y la toma de requerimientos adicional con el usuario para su realización.	
C O M P L	En un ingreso de información para XXX, ¿se debe indicar los datos que son obligatorios y validar esta condición?	Con estos ingresos aumenta el tiempo de ingreso de los datos y su tiempo de procesamiento y validación.	

⁶ Descripción del PI XXX

E T I T U D	Quando la información obtenida para XXX se origina desde múltiples fuentes, ¿se necesita consolidar esta información para asegurar que la información entregada esté completa?	Se debe implementar un proceso de análisis de los datos que sea capaz de recolectar la información desde las múltiples fuentes, eliminar las redundancias, corregir errores y darle consistencia. Este proceso debe realizarse con el cuidado de no usarlo mientras esta información esté en interacción con el usuario y con una periodicidad a definir.	
C O N S I S T E N C I A	Los datos contenidos en XXX, ¿no deben tener duplicidades de información, deben estar libre de redundancias y con coherencia?	Deben existir procesos de limpieza de esta información para asegurar que estos atributos estén presentes. Deben existir validaciones al momento de ingresar la información para asegurar estas características deseadas.	
C R E D I B I L I D A D	¿Se debe dar a conocer la procedencia o fuente de origen de los datos proporcionados para XXX?	El usuario conocerá el origen de los datos y se deben adoptar las medidas necesarias para administrar esta información.	
L I D A D	Los datos contenidos en XXX, ¿deben tener un alto grado de objetividad, deben estar libres de interpretaciones incorrectas y sin errores?	Al momento de la implementación de XXX debe haber una marcada participación del usuario, sobre todo en el ciclo de pruebas, para asegurar que la información contenida en el PI contenga estos atributos indicados y de esta forma conseguir su credibilidad.	
A C T U A L I D A D	Los datos proporcionados en XXX, ¿deben tener actualizaciones en línea de modo que el usuario pueda visualizar las modificaciones en forma inmediata?	Considerar tiempos adicionales en la ejecución de las aplicaciones que contengan la información de XXX, al estar constantemente mostrando los tiempos de actualizaciones de la información.	
L I D A D	¿Es necesario dar a conocer a los usuarios información que complementa a XXX para saber la actualidad de sus datos. Por ejemplo fechas y hora?	Considerar un tiempo adicional en incorporar en el diseño e implementación de las pantallas que generan la información de XXX, estos datos que indiquen actualidad de la información.	
A C C E S O	La información contenida en XXX, ¿debe ser de fácil acceso, evitando tiempos excesivos de búsqueda de la información solicitada?	Se debe considerar un tiempo adicional en el diseño de pantallas que otorguen un fácil acceso de acuerdo a las necesidades del usuario.	

S I B I L	Las búsquedas de información, ¿deben proporcionar datos de búsqueda obligatorios de modo de acceder a la información contenida en XXX en un tiempo adecuado de consulta?	El usuario debe conocer los datos obligatorios al momento de hacer la consulta.	
I D A D	¿Es útil contar con accesos a información histórica de XXX para consultar de acuerdo a las necesidades del usuario?	Considerar tiempos prolongados de procesamiento al trabajar con grandes volúmenes de información.	
C O N F O R M I D A D	Los datos producidos para XXX, ¿deben ser definidos de manera estandarizada, para que todos los usuarios los entiendan, evitando dudas o diferentes interpretaciones?	Se deben definir plantillas antes de implementar XXX para asegurar la conformidad del producto.	
E F I C I E N C I A	La información contenida en XXX, ¿debe tener constantes procesos de mantención, para asegurar un uso eficiente de recursos como velocidad de respuesta, uso de memoria y almacenamiento?	Se deben considerar costos de implementación adicionales relacionados con esta tarea.	
	¿Se debe contar con procesos encargados de limpiar información innecesaria para XXX?	Considerar un tiempo adicional en desarrollar procesos automáticos de limpieza de ejecución con calendarios de ejecución adecuados para no alterar la ejecución normal de interacción con el PI.	
P R E C I S I O N	¿Es necesario definir formatos de acuerdo al tipo de datos para la información contenida en XXX?	Estos formatos deben contener tipo de letra, largo máximo especificado de caracteres, número especificado de decimales, entre otras definiciones. Considerar posibles lentitudes de ingreso de información al validar la precisión de sus datos.	

T R A Z A B I L I D A D	¿Es necesario almacenar la información de cambios realizados por los usuarios a los datos para XXX?	Se debe incorporar una bitácora para almacenar las transacciones que realiza el usuario con los datos en el PI y luego traer al sistema la información de estas transacciones en un lenguaje comprensible para que el usuario use esta información de acuerdo a lo que necesita.	
C O M P R E N S I B I L I D A D	Los datos contenidos en XXX, ¿deben ser claros y fáciles de comprender?	Realizar pantallas amigables para el ingreso de la información, con simplicidad de colores, textos y ayuda para facilitar su ingreso. Definir como será dada esta ayuda, por ejemplo con una información que aparezca cada vez que el usuario ingresa a la pantalla u otra forma fácil de identificar que es una ayuda a la cual puede acceder el usuario.	
D	Los datos contenidos en XXX, ¿deben ser exactos y estrictamente relevantes?	Hay que considerar un tiempo adicional al desarrollo en la incorporación de validaciones y habrá un mayor tiempo en el ingreso de información por parte del usuario al usar las validaciones para asegurar que la información contenida en XXX sea comprensible.	
C O N F I D E N C I A L I D A D	¿Se requiere que el acceso del PI XXX sea restringido, es decir, sólo para usuarios autorizados?	Se debe contar con información de quienes están autorizados a acceder la información contenido en el PI y definir el nivel de restricción que tendrá esta información.	
	¿Es necesario guardar la información de XXX en un código ilegible para evitar accesos de lectura indeseados?	Esto significa un tiempo extra al desarrollo e implementación de XXX para incorporar lógicas de guardado en clave y de desciframiento cuando un usuario autorizado lo quiera leer.	
	¿Es necesario contar con políticas de seguridad para resguardar la información presente en XXX?	Esto lleva a restringir el uso de la información del PI, definir el nivel de impacto que tendrán estas políticas en el resto de los PI's y establecer su ámbito de acción.	
P O R	¿Es necesario contar con funcionalidades que permitan importar o exportar la información	Esto indica la necesidad de implementar lógicas necesarias para armar una estructura que contenga el formato de la información útil	

T A	contenida en XXX?	para importar y exportar el PI en un formato estándar.	
B I L I D A D	¿Es necesario que XXX pueda ser transportable de un sistema a otro sin alterar su esencia y calidad?	Esto implica un desarrollo adicional para preparar el PI en formatos exportables, de acuerdo al sistema que lo necesite.	
R E C U P E R A	¿XXX puede ser recuperado por los usuarios autorizados cada vez que ellos lo requieran?	Esto requiere una lógica de seguridad para autorizar a ciertos usuarios para recuperar información y tener la disponibilidad de la información que se necesite recuperar. Considerar esfuerzos adicionales en asegurar la consistencia e integridad de los datos recuperados.	
B I L I D A D	¿Es necesario contar con políticas de recuperabilidad de los datos que conforman a XXX?	Considerar el uso de recursos en diseñar e implementar esta política de recuperabilidad una vez recolectados los antecedentes necesarios de este requerimiento	
D I S P O	¿Se necesita que la información de XXX esté siempre disponible?	Esto significa un costo adicional en implementar una lógica de guardar información histórica y traerla al sistema cuando se requiera. Considerar lentitud en el procesamiento de grandes volúmenes de información.	
N I B I L I D A D	¿Se necesita que la información de XXX esté disponible aunque el sistema esté fuera de línea?	Esto significa un costo adicional en implementar lógica para trabajar con datos fuera de línea y reestablecer la información al momento de recuperar las conexiones de red.	

Referencias

- Adelakun, O. (1997). Quality - What Does it Mean for Strategic Information Systems. ICIQ. D. Strong, Kahn, B. Cambridge, Massachusetts, USA: 132-147.
- Ballou, D., Tayi, G. (1989). "Methodology for allocating resources for data quality enhancement." Communications of the ACM 32(3): 329.
- Ballou, D., Wang, R., Pazer, H., Kumar, G. (1998). "Modeling information manufacturing systems to determine information product quality." Management Science 44(4): 462-484.
- Baškarada, S., Koronios, A., Gao, J. (2006). Towards a Capability Maturity Model for Information Quality Management: A TDQM Approach: 10-12.
- Baskerville, R. (1999). "Investigating information systems with action research." Communications of the AIS 2(3es): 4.
- Batini, C., Scannapieca, M. (2006). Data quality: Concepts, methodologies and techniques, Springer-Verlag New York Inc.
- Boelens, R. (2006). "Product-attribute approach to total data quality management." Product Data Quality University of Twente.
- Cappiello, C., Francalanci, C., Pernici, B. (2004). Data quality assessment from the user's perspective.
- De Miguel, A., Piattini, M. (1993). "Concepción y diseño de bases de datos: Del Modelo E/R al Modelo Relacional." Madrid, Ra-Ma.
- Dedeke, A., Kahn, B. (2002). Model-Based quality evaluation: a comparison of Internet classified operated by newspapers and non-newspaper firms. ICIQ. C. Fisher, Davidson, B. Cambridge, Massachusetts, USA: 142-154.
- Deming, W. (1986). "Out of the crisis, Massachusetts Institute of Technology." Center for Advanced Engineering Study, Cambridge, MA.
- Doggett, A. (2006). "Root cause analysis: A framework for tool selection." Quality Control and Applied Statistics 51(3): 279.
- Firth, C., Wang, R. (1996). "Data quality systems: Evaluation and implementation." Cambridge Market Intelligence Ltd., London.
- French, W., Bell, C. (1996). Organizational Development: Behavioral Science Interventions for Organization Improvement. London.
- Gackowski, Z. (2007). Pervasiveness Of Materiality Of Factors In Operations And Their Changes In Decision Situations. ICIQ. M. Robbert. Cambridge, Massachusetts, USA: 118-130.
- Guerra-García, C., Caballero, I., De Guzmán, I., Piattini, M. (2009). Modelado de Requisitos de Calidad de Datos en el Proceso de Desarrollo de Portales Web. Jornadas de Ing. del Software y BBDD. 3: 124-133.
- Guimaraes, T., Staples, D., Mckeen, J. (2004). "Empirically testing some main user-related factors for systems development quality." Quality Control and Applied Statistics 49: 333-336.
- Heravizadeh, M., Mendling, J., Rosemann, M. (2008). "Root Cause Analysis in Business Processes." QUT Digital Repository.
- Hess, K., Talburt, J. (2004). Applying Name Knowledge to Name Quality Assessment. ICIQ. S. Chengalur-Smith. Cambridge, Massachusetts, USA: 211-217.
- ISO/IEC-25012 (2008). "ISO/IEC 25012: Software Engineering - Software Quality Requirements and Evaluation (SQuARE) - Data Quality Model."

- Jiang, L. (2007). **Data Quality By Design: Agoal-ORIENTED APPROACH.** ICIQ. M. Robbert. Cambridge, Massachusetts, USA: 249-263.
- Kahn, B., Strong, D., Wang, R. (2002). "Information quality benchmarks: product and service performance." Communications of the ACM 45(4): 192.
- Kaomea, P. (2005). **Strategic Data Quality: From Real World Sources To Real World Uses.** ICIQ. F. Naumann, Gertz M., Madnick S. Cambridge, Massachusetts, USA: 2-11.
- Kerr, K., Norris, T., Stockdale, R. (2004). **Data Quality Information and Decision Making: A Healthcare Case Study.** ICIQ. S. Chengalur-Smith. Cambridge, Massachusetts, USA: 218-230.
- Kitchenham, B. (2004). **Procedures for Performing Systematic Reviews.** Technical Report 0400011T.1.
- Lee, Y., Pipino, L., Funk, J., Wang, R. (2006). "Journey to Data Quality." Cambridge, MA, USA: Massachusetts Institute of Technology.
- Lesca, H., Lesca, E. (1995). Gestion de l'information:(qualité de l'information et performances de l'entreprise).
- Lewin, K. (1947). "Frontiers in group dynamics." Human relations 1(2): 143.
- Liu, L., Chi, L. (2002). **Evolutional data quality: A theory-specific view.** ICIQ. C. Fisher, Davidson, B. Cambridge, Massachusetts, USA: 292-304.
- Madnick, S., Wang, R., Yang, L., Hongwei, Z. (2009). "Overview and Framework for Data and Information Quality Research." Data and Information Quality 1(1): 1-22.
- McTaggart, R. (1991). "Principles for participatory action research." Adult Education Quarterly 41(3): 168.
- Mielke, M. (2005). **IQ Principles in Software Development.** ICIQ. F. Naumann, Gertz, M., Madnick, S. Cambridge, Massachusetts, USA: 358-369.
- Orman, L., Storey, V., Wang, R. (1996). **Systems Approaches to Improving Data Quality.** ICIQ. R. Wang. Cambridge, Massachusetts, USA: 117-126.
- Pierce, E. (2004). **Developing, implementing and monitoring an information product quality strategy.** ICIQ. S. Chengalur-Smith. Cambridge, Massachusetts, USA: 13-26.
- Pipino, L., Lee, Y., Wang, R. (2002). "Data quality assessment." Communications of the ACM 45(4): 211-218.
- Pressman, R. (2005). Software engineering: a practitioner's approach, McGraw-Hill New York.
- Scannapieca, M., Mirabella, V., Mecella, M., Batini, C. (2002). "Data Quality in e-Business Applications." Web Services, E-Business, and the Semantic Web: 121-138.
- Schmidt, A., Otto, B. (2008). **A Method for the Identification and Definition of Information Objects.** ICIQ. P. Neely. Cambridge, Massachusetts, USA: 214-228.
- Sessions, V. (2007). **Employing the TDQM Methodology: An Assessment of the SC SOR.** ICIQ. M. Robbert. Cambridge, Massachusetts, USA: 519-537.
- Storey, V., Wang, R. (1998). **Modeling quality requirements in conceptual database design.** ICIQ. C. Chengalur-Smith, Pipino, L. Cambridge, Massachusetts, USA: 64-84.
- Strong, D., Lee, Y., Wang, R. (1997). "Data Quality in Context." Communications of the ACM Vol. 40, N° 5: 103 -110.
- Tayi, G., Ballou, D. (1998). "Examining data quality." Communications of the ACM 41(2): 57.
- Wadsworth, Y. (1998). "What is Participatory Action Research? ." Action Research International, Paper 2.
- Wang, R. (1998). "A Product Perspective on Total Data Quality Management." Communications of the ACM 41(2): 58-65.
- Wang, R., Henry, K., Madnick, S. (1992). **Data Quality Requirements Analysis and Modelling.** Conference of Data Engineering. Vienna, Austria.
- Wang, R., Lee, Y. (1998). Integrity Analyzer: A Software Tool for Total Data Quality Management. Cambridge Research Group, Cambridge, MA.
- Wang, R., Strong, D. (1996). "Beyond accuracy: What data quality means to data consumers." Journal of management information systems 12(4): 33.
- Wang, R., Ziad, M., Lee, Y. (2001). Data Quality. Massachusetts, Kluwer Academic Pub.

- Wende, K., Otto B. (2007). A Contingency Approach To Data Governance. ICIQ. M. Robbert. Cambridge, Massachusetts, USA: 163-176.**
- Wright, J., Vesonder, G., (2004). Stewardship And Staging Of Information For Enterprise Mining. ICIQ. C. Chengalur-Smith. Cambridge, Massachusetts, USA: 248-259.**