

Calidad y Medición de Sistemas de Información

TRABAJO:

Introducción de Dimensiones de Calidad de Datos (DQ) como requisitos añadidos en el proceso de desarrollo de un producto software

Presentado por:

César A. Guerra García.

9 Enero de 2009.

Índice

| | |
|--|----|
| 1. <i>Introducción</i> | 3 |
| 2. <i>Estado del Arte</i> | 5 |
| 2.1. <i>Funcionalidades de un Portal Web</i> | 5 |
| 2.2. <i>Medición de Calidad de Datos</i> | 6 |
| 2.3. <i>Técnica de Obtención de Requisitos</i> | 8 |
| 3. <i>DQ-VORD: Una metodología para la identificación / definición de Requisitos de Calidad de Datos para portales Web</i> | 11 |
| 3.1. <i>Análisis acerca de cuáles dimensiones de DQ afectan a las distintas funcionalidades de un portal</i> | 11 |
| 3.2. <i>Metodología DQ-VORD</i> | 12 |
| 4. <i>Conclusiones y Tendencias Futuras</i> | 16 |
| 5. <i>Referencias</i> | 16 |

1. Introducción

Los problemas en la calidad de datos (DQ) se han venido incrementando y son realmente evidentes, particularmente en bases de datos organizacionales. En registros criminales computerizados de EE.UU., entre el 50% y el 80% de los datos fueron encontrados inexactos, incompletos o ambiguos. El impacto económico y social por una pobre calidad de datos ha tenido un coste económico importante para las organizaciones [6] [10]. Por otra parte, Levis [12] menciona algunos ejemplos de escenarios en los que datos con niveles de calidad inadecuados originan problemas que afectan negativamente a los Sistemas de Información, y por lo tanto, al rendimiento organizacional. Los problemas producidos por esta falta de calidad pueden ser clasificados a diferentes niveles dependiendo de su naturaleza [1]: *técnicos* (por ejemplo, problemas en la implementación de almacenes de datos o en el diseño de Wrappers Web para la integración de aplicaciones), *organizacional* (tales como, pérdida de clientes, pérdidas financieras, etc.) y *legales* (la violación del artículo cuarto del Apartado II de la LOPD de 1999).

La calidad de datos es tratada como un concepto específico, independientemente del contexto en el cual el dato es producido, almacenado o utilizado, es decir; en un enfoque más genérico, los problemas de DQ sólo se visualizan como errores en los datos almacenados. Sin embargo, este enfoque fracasa al resolver problemas organizacionales complejos [16], porque el concepto de calidad no debe ser entendido sólo como “cero defectos” en los datos, sino como una adecuación al uso de los mismos, entre la que se encuentra esta exigencia. Además, de cara a determinar el grado de adecuación de los datos para su uso, debe tenerse en cuenta que cada usuario desempeñando un rol tendrá una percepción de la calidad de los datos para cada una de las tareas que realice utilizando ciertas herramientas software.

El enfoque de nuestra investigación se centra precisamente en esas herramientas software que los usuarios utilizan para desempeñar esas tareas. Así asumimos que, normalmente los datos son tratados utilizando determinados productos software construidos para satisfacer los requisitos del negocio. Se podría pensar que si dichos productos software estuvieran también diseñados para dar soporte a la gestión necesaria de la calidad de datos (siempre desde el punto de vista de adecuación al uso, y no simplemente de “cero defectos” en los datos), entonces sería posible garantizarle al usuario una estimación del error que puede estar cometiendo al usar ciertas colecciones de datos, de modo que pueda tomar decisiones sobre si utilizar o no los datos.

Este enfoque de partida nos lleva a formular la hipótesis de que si la gestión de la calidad de datos se pudiese tratar como unos requisitos más del negocio a la hora de construir las herramientas software, entonces, podrían ser implementados mecanismos para su aseguramiento

desde fases tempranas del desarrollo de los productos software, garantizando así la satisfacción de los distintos usuarios de los datos.

Desafortunadamente no existen en la bibliografía relacionada descripciones detalladas de tales mecanismos que permitan gestionar específicamente los requisitos de calidad de datos a nivel de desarrollo software, tan sólo algunas propuestas a nivel de modelo de datos, como la propuesta en [17] para el modelo relacional y a nivel de tecnologías semánticas como la propuesta en [2].

Un primer paso para afrontar el objetivo propuesto en este trabajo de introducir como requisitos software los aspectos de calidad de datos que deberían tener las aplicaciones, podría ser buscar en la bibliografía aquellas situaciones descritas por diferentes autores que han identificado niveles inadecuados de DQ en diferentes contextos. Los orígenes más comunes de esos niveles inadecuados son algunos baches¹ que pueden ser encontrados durante el ciclo de vida de los datos, tal como los describe Strong y colegas [15]. Entre estos baches podrían citarse los siguientes: Múltiples fuentes de la misma información producen diferentes valores, la información es producida usando juicios subjetivos conduciendo a prejuicios, errores sistemáticos en la producción de la información generan pérdidas de datos, fácil acceso a la información que puede entrar en conflicto con requisitos de seguridad, privacidad y confidencialidad, entre otros.

Determinar si el nivel de DQ es el adecuado para la realización de las tareas organizacionales implica necesariamente la definición de medidas de DQ sobre los datos, y el establecimiento de algunos rangos de aceptación válidos para los valores obtenidos, siempre desde el punto de vista de los usuarios que están ejecutando esas tareas. En este sentido, sería interesante aplicar técnicas y herramientas clásicas de gestión de calidad, para localizar los datos con niveles inadecuados de DQ y sus causas de una manera más eficiente [1].

Como se ha constatado que el problema de abordar la introducción de mecanismos de aseguramiento de calidad de datos en el desarrollo de productos software, se ha decidido hacer una primera aproximación limitándose a un único tipo de productos software debido a su gran presencia hoy en día y su continua y creciente demanda en un futuro cercano: los portales Web. Para ser coherente con la idea de estudiar la DQ para las distintas tareas que un usuario puede realizar con un sistema, se ha planteado como una estrategia interesante para este trabajo estudiar la DQ desde el punto de vista de las funcionalidades de un portal Web, enumeradas por Collins en [5] y tomando como base el trabajo de Caro en su tesis doctoral [4]. Para ello, se analiza para cada funcionalidad qué dimensión o dimensiones de calidad de datos deben ser tenidas en cuenta (estas relaciones se han representado mediante una matriz en la sección 3.1). Con los resultados de este análisis se pretende formular una serie de requisitos relacionados con

¹ Traducido de potholes.

la calidad de datos para ser introducidos en el desarrollo de un producto software tipo “Portal Web”. Debido a que la definición de dichos requisitos puede ser una tarea compleja, el trabajo se va a apoyar en el método VORD (Definición de Requerimientos Orientados a Puntos de Vista) descrito en la sección 2.3, lo que permitirá tener en cuenta el enfoque de los distintos usuarios.

Así el resto del trabajo está organizado como sigue: la sección dos (2) describe el estado del arte de los aspectos relacionados con las funcionalidades principales de un portal Web, aspectos relacionados con la medición de calidad de datos, junto con una breve descripción del modelo de calidad de datos propuesto en el estándar ISO-25012 [8] recientemente hecho público, y una descripción de la técnica de obtención de requisitos orientados a puntos de vista; la sección tres (3), muestra por un lado el análisis realizado para la identificación de las dimensiones de DQ que son consideradas necesarias/importantes para las funcionalidades de un portal Web, y además, por otro, una propuesta para la obtención de requisitos con conciencia de calidad de datos (*DQ-VORD*); por último, en la sección cuatro (4) se mencionan las conclusiones y el trabajo futuro.

2. Estado del Arte

2.1. *Funcionalidades de un Portal Web*

Como ya se mencionó anteriormente, uno de nuestros objetivos del trabajo es proponer una matriz de relación, entre las funcionalidades de un portal Web y las Dimensiones de DQ, por lo que es fundamental enumerar y repasar dichas funcionalidades, las cuales son mostradas en [5]. A continuación se muestran ordenadas usando como criterio la probabilidad de utilización en un portal Web y por tanto de ser susceptibles de encontrar niveles inadecuados de DQ que supongan que el usuario del portal pueda no estar satisfecho con el mismo. Además se añade un pequeño comentario sobre las perspectivas de uso de datos para dicha funcionalidad. Es necesario indicar que el orden que se propone está determinado en base a mi experiencia y conocimientos en el área de DQ.

1. **Gestión de Contenido.** La publicación y distribución de información debe soportar la creación de contenido, autorización e inclusión (o exclusión) de colecciones de contenido. Las organizaciones tienen volúmenes de información y material de referencia que se dejan disponible desde el portal.
2. **Procesos y Acciones.** Habilitan a los usuarios del portal a iniciar y participar en procesos de negocios de la organización.

3. **Capacidades de búsqueda.** Se deben proveer servicios de búsqueda de información para los usuarios, que soporten búsqueda a lo largo de la empresa, en el World Wide Web, y en motores de búsqueda de catálogos e índices.
4. **Administración.** Esta funcionalidad provee dos servicios. El primero, un servicio correspondiente a actividades de mantenimiento y utilización de las tareas asociadas con el Sistema del Portal Corporativo. El segundo, un servicio que puede ser configurado únicamente por el administrador con el Sistema del Portal Corporativo, así como por cada usuario a través de la personalización.
5. **Seguridad.** Provee una descripción de los niveles de acceso de cada usuario o grupo de usuarios, para cada aplicación y función de software incluidas en el portal.
6. **Integración y puntos de datos.** Capacidad para acceder a la información desde un amplio rango de fuentes de información internas y externas y mostrar la información resultante como un simple punto de acceso en el escritorio.
7. **Comunicación y Colaboración.** Estos elementos facilitan la discusión, la identificación de ideas innovadoras y soluciones ingeniosas o emprendedoras. Estos elementos dan la posibilidad a los empleados de trabajar juntos de forma más cualitativa mediante la creación de un entorno virtual compartido (colaboración), soportando mensajería electrónica (comunicación) y agregando características de comunicación y colaboración al proceso de negocios (coordinación).
8. **Presentación.** Provee la experiencia visual para el usuario, que encapsula todas las funcionalidades del portal.
9. **Taxonomía.** La taxonomía provee información de contexto incluyendo categorías específicas de la organización que soportan y reflejan los negocios de ésta. Provee terminología común usada en la organización de reconocimiento rápido y mejora la semántica para los usuarios del portal (empleados, socios, clientes, etc.).
10. **Personalización.** Este es un componente crítico para crear un ambiente de trabajo organizado y configurado especialmente para cada empleado en la organización. La clave está en lograr un balance entre la información y los atributos que se necesitan para ser consistentes y constantes en la personalización del portal y las características únicas requeridas por los empleados individuales.
11. **Características de Ayuda.** Asistencia en el uso del portal. La ayuda debe estar referida tanto a características del portal como a características específicas de la organización.

2.2. *Medición de Calidad de Datos*

Una de las estrategias más interesantes para abordar el estudio de calidad de datos para un contexto en específico, es dividirla en partes más pequeñas llamadas “calidades menores”,

conocidas como dimensiones de calidad de datos [11]. El conjunto de dimensiones de calidad de datos usables en un contexto es conocido como “Modelo de calidad de datos”.

La calidad de datos es un concepto multidimensional, por lo cual, para medirla es necesario descomponerla en características observables llamadas dimensiones de calidad de datos, en base a las cuales sea posible definirla, identificarla y medirla. El modelo propuesto por el estándar ISO/IEC 25012 [8] categoriza los atributos de calidad de datos en 15 características o dimensiones considerados desde dos puntos de vista:

- *Inherente*: La calidad de datos *inherente* se refiere al grado en el cual las características de calidad del dato tienen el potencial intrínseco para satisfacer las necesidades implicadas cuando el dato es usado bajo condiciones específicas.
- *y, dependiente del sistema*. la calidad de datos *dependiente del sistema* se refiere al grado en el cual la calidad del dato es enriquecida y preservada dentro de un sistema de cómputo cuando el dato es usado bajo condiciones específicas.

A continuación se describen cada una de las dimensiones de calidad de datos (véase Tabla 1. Dimensiones de Calidad de Datos según ISO/IEC 25012):

| Inherentes | |
|--|--|
| Dimensión | Descripción |
| Exactitud (Accuracy) | El grado en el cual el dato tiene atributos que correctamente representan el valor correcto del atributo intencionado de un concepto o evento en un contexto específico de empleo. |
| Complejidad (Completeness) | El grado al cual el dato del sujeto asociado con una entidad tiene valores para todos los atributos esperados e instancias de entidad relacionadas en un contexto específico de uso. |
| Consistencia (Consistency) | El grado en el cual el dato tiene los atributos que son libres de contradicción y son coherente con otros datos en un contexto específico de uso. |
| Credibilidad (Credibility) | El grado en el cual el dato tiene atributos que son considerados como verdaderos y creíbles por usuarios en un contexto específico de uso. |
| Actualidad (Currentness) | El grado en el cual el dato tiene los atributos que son del período correcto en un contexto específico de uso. |
| Inherentes y Dependientes del sistema | |
| Accesibilidad (Accessibility) | El grado en el cual el dato puede ser accedido en un contexto específico de uso, en particular por la gente que necesita el soporte de tecnología o una configuración especial debido a alguna inhabilidad (incapacidad). |
| Conformidad (Compliance) | El grado en el cual el dato tiene atributos que se adhieren a normas, convenciones o regulaciones vigentes y reglas similares relacionadas con la calidad de datos en un contexto específico de uso. |
| Confidencialidad (Confidentiality) | El grado en el cual el dato tiene los atributos que aseguran que éste es sólo accesible e interpretable por usuarios autorizados en un contexto específico de uso. |
| Eficiencia (Efficiency) | El grado en el cual el dato tiene los atributos que pueden ser procesados y proporciona los niveles esperados de funcionamiento (desempeño) usando las cantidades y los tipos de recursos apropiados en un contexto específico de uso. |
| Precisión (Precision) | El grado en el cual el dato tiene atributos que son exactos o que proporcionan la discriminación en un contexto específico de uso. |
| Trazabilidad (Traceability) | 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. |
| Entendibilidad (Understandability) | 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. |
| Dependientes del Sistema | |
| Disponibilidad (Availability) | El grado en el cual el dato tiene atributos que le permiten ser recuperados por usuarios autorizados y/o aplicaciones en un contexto específico de uso. |
| Portabilidad (Portability) | 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. |
| Recuperabilidad (Recoverability) | 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. |

Tabla 1. Dimensiones de Calidad de Datos según ISO/IEC 25012

2.3. *Técnica de Obtención de Requisitos.*

La Ingeniería de Requisitos es el proceso sistemático de desarrollar requerimientos a través de un proceso iterativo y cooperativo de análisis del problema, documentando las observaciones resultantes en una variedad de formatos de representación, y verificando la precisión del entendimiento ganado [13].

Los enfoques orientados a puntos de vista para la ingeniería de requisitos toman en consideración los diferentes puntos de vista y los utilizan para estructurar y organizar tanto el proceso de obtención como los requisitos mismos. El punto clave del análisis orientado a puntos de vista es que toma en cuenta la existencia de varias perspectivas y provee un marco de trabajo para descubrir conflictos en los requisitos propuestos por diferentes stakeholders² [7].

Los puntos de vista se pueden utilizar como una forma de clasificar los stakeholders y otras fuentes de requerimientos. Existen tres tipos genéricos de puntos de vista [14]:

1. *Puntos de vista de los interactuadores*: representan a las personas u otros sistemas que interactúan directamente con el sistema.
2. *Puntos de vista indirectos*: representan a los stakeholders que no utilizan el sistema ellos mismos pero que influyen en los requisitos de algún modo.
3. *Puntos de vista del dominio*: representan las características y restricciones del dominio que influyen en los requisitos del sistema.

Por lo general, estos puntos de vista proporcionan diferentes tipos de requisitos:

- Los *puntos de vista de los interactuadores* proporcionan requisitos detallados del sistema que cubren las características, funcionalidades e interfases del sistema.
- Los *puntos de vista indirectos* es más probable que proporcionen requisitos y restricciones organizacionales de alto nivel.
- Los *puntos de vista del dominio* proporcionan restricciones del dominio que se aplican al sistema.

Como ilustración se puede considerar la jerarquía de puntos de vista mostrada en la Figura 1, es un diagrama relativamente sencillo de los puntos de vista que se pueden consultar en la obtención de requisitos para un sistema Web de una biblioteca.

² *Personas involucradas de alguna manera en el proceso de desarrollo de software.*

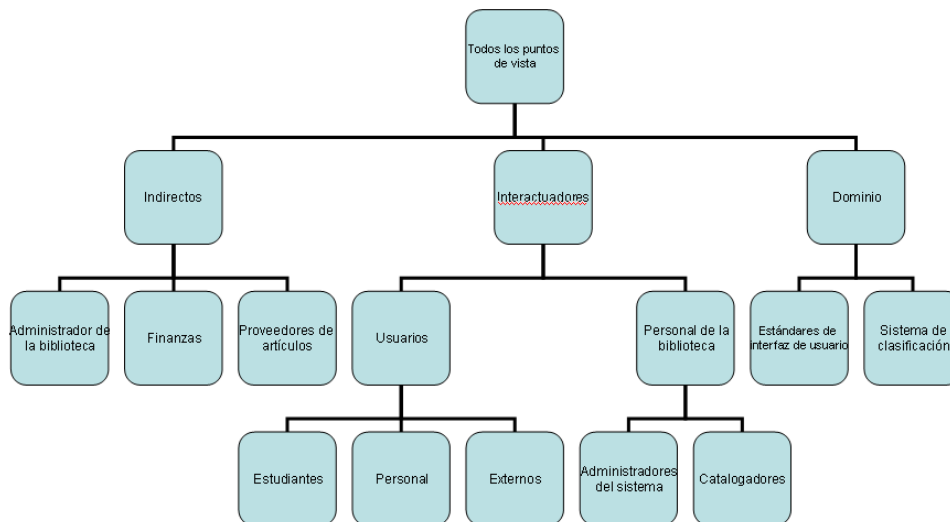


Figura 1. Ejemplo de puntos de vista para un portal Web de una Biblioteca

La identificación inicial de los puntos de vista que son relevantes a un sistema a veces puede llegar a ser difícil. Para ayudar a este proceso, se debería intentar identificar tipos de puntos de vista más específicos [14]:

1. Los proveedores de servicios al sistema y los receptores de los servicios del sistema.
2. Los sistemas que deben interactuar directamente con el sistema a especificar.
3. Las regulaciones y estándares que se aplican al sistema.
4. Las fuentes de los requisitos no funcionales y de negocio del sistema.
5. Los puntos de vista de la ingeniería que reflejan los requisitos de las personas que tienen que desarrollar, administrar y mantener el sistema.
6. Los puntos de vista de la mercadotecnia y otros que generan requisitos sobre las características del producto esperadas por los clientes y cómo el sistema debería reflejar la imagen externa de la organización.

Los puntos de vista de la ingeniería pueden ser importantes por dos razones. En primer lugar, los ingenieros que desarrollan el sistema pueden tener experiencia con sistemas similares y sugerir requerimientos a partir de esa experiencia. En segundo lugar, el personal técnico que tiene que administrar y mantener el sistema puede tener requerimientos que ayuden a simplificar el soporte del sistema. Los requerimientos de administración del sistema son cada vez más importantes debido a que los costes de administración son una proporción creciente de los costes totales de un sistema. Finalmente, los puntos de vista que proporcionan requerimientos

pueden venir de los departamentos de mercadotecnia y asuntos externos en una organización. Esto es especialmente cierto para sistemas basados en Web, particularmente sistemas de comercio electrónico y productos software empaquetados. Los sistemas basados en Web deben presentar una imagen favorable de la organización además de entregar funcionalidad al usuario [14].

Para cualquier sistema no trivial, existe un enorme número de posibles puntos de vista, y es prácticamente imposible obtener requisitos de todos ellos. Por lo tanto, es importante organizar y estructurar los puntos de vista en una jerarquía, es probable que los puntos de vista en la misma rama compartan requerimientos comunes.

El método VORD (Definición de Requisitos Orientado a Puntos de Vista) [9] se ha diseñado como un marco de trabajo orientado a servicios para la obtención y análisis de requisitos. Las etapas principales de este método son:

1. *E1. Identificación de puntos de vista*, que implica descubrir los que reciben los servicios del sistema e identificar los servicios específicos que se suministran a cada punto de vista.
2. *E2. Estructuración de puntos de vista*, que comprende agrupar los relacionados en una jerarquía. Los servicios comunes se ubican en los niveles altos de la jerarquía y se heredan los puntos de vista de bajo nivel.
3. *E3. Documentación de puntos de vista*, que comprende refinar la descripción de éstos y los servicios identificados.
4. *E4. Trazado del punto de vista del sistema*, que comprende identificar los objetos en un diseño orientado a objetos utilizando la información del servicio encapsulado en los puntos de vista.

3. DQ-VORD: Una metodología para la identificación / definición de Requisitos de Calidad de Datos para portales Web

3.1. *Análisis acerca de cuáles dimensiones de DQ afectan a las distintas funcionalidades de un portal.*

Una vez identificadas tanto las funciones básicas de un portal Web como las distintas dimensiones de calidad de datos, el siguiente paso es el llevar a cabo una clasificación de cada una de las dimensiones de DQ en relación a las funcionalidades de un portal, es decir, determinar qué atributos de calidad tienen algún tipo de relación al momento de implementar alguna de las distintas funcionalidades deseables en un portal Web (véase Tabla 2). Con esto se pretende determinar un camino que dirija la investigación sobre las características de DQ que deberíamos empezar a tener en cuenta primero, para su posterior implementación durante el proceso de desarrollo de portales Web.

Tomando como base el análisis realizado por Caro y colegas [3], en donde se realiza una clasificación de los atributos de DQ para portales Web de acuerdo al modelo de calidad de datos definido por [16], el signo “X” en la matriz muestra estas relaciones; mientras que el signo “O” representa la relación entre las funciones y las dimensiones de calidad de datos propuestas en el estándar internacional [8]. En la Tabla 3 se muestran cada una de las relaciones marcadas con el signo “O” (función, dimensión).

| DQ Dimensiones | Exactitud | Completitud | Consistencia | Credibilidad | Actualidad | Accesibilidad | Confidencialidad | Confidencialidad | Eficiencia | Flexibilidad | Integrabilidad | Disponibilidad | Portabilidad | Recuperabilidad |
|-------------------------------|-----------|-------------|--------------|--------------|------------|---------------|------------------|------------------|------------|--------------|----------------|----------------|--------------|-----------------|
| Integración y puntos de datos | X | X | X | X | X | | | | X | | | X | O | |
| Taxonomía | O | | X | X | X | | | | X | | | X | | |
| Capacidades de búsqueda | X | X | X | X | X | | | | X | | O | X | O | |
| Características de Ayuda | | | X | | | | X | | | | | X | | |
| Gestión de Contenido | X | X | X | X | X | X | X | O | X | | O | X | | O |
| Procesos y Acciones | X | X | X | X | X | X | X | | X | | | X | | |
| Comunicación y Colaboración | O | | X | O | O | X | | | | | | X | O | |
| Personalización | X | X | | | X | X | | | | | | | O | |
| Presentación | X | X | X | X | X | | | | X | | | X | | |
| Administración | | | X | | | X | O | | O | O | | X | | O |
| Seguridad | X | X | X | | X | X | | O | | | O | X | | |

Tabla 2. Matriz de relación de Dimensiones de DQ [8] y Funcionalidades de un Portal Web [5]

| Funciones de un portal Web | Dimensión de DQ | Descripción |
|--------------------------------|------------------|--|
| Integración y puntos de datos. | Disponibilidad | Los datos deberían de estar siempre al alcance de los usuarios en el momento que sean requeridos. |
| Taxonomía | Exactitud | Los datos intrínsecamente deberían de ser sintáctica y semánticamente correctos, mejorando la terminología común utilizada en las organizaciones. |
| Capacidades de búsqueda | Disponibilidad | Los datos deberían poder ser recuperados al momento de ejecutar algún tipo de búsqueda específica. Ello implica que el dato esté accesible, que los caminos de acceso estén disponibles y que el usuario tenga los permisos adecuados para su lectura. |
| | Trazabilidad | El sistema debería proporcionar información acerca de la cantidad de veces que ha sido accedido y/o modificado un determinado dato o conjunto de datos. |
| Gestión de Contenido | Confidencialidad | Los datos en sí deberían sólo ser accesibles e interpretables por usuarios autorizados. |
| | Trazabilidad | Los datos deberán proporcionar información de cuándo y quién publicó o distribuyó éstos y quién los podrá acceder. |
| | Portabilidad | Los datos podrán ser instalados o movidos dentro de cualquier sistema en la organización. |
| Comunicación y Colaboración | Exactitud | Los datos deberán ser utilizados y comprendidos en el mismo contexto por todos los usuarios (empleados), mejorando con esto la comunicación y el desempeño de ellos. |
| | Credibilidad | Los datos deberán ser considerados como verídicos por todos los usuarios, mejorando la colaboración entre cada uno de ellos. |
| | Actualidad | Los datos deberán estar actualizados, ayudando con ello a realizar soluciones ingeniosas e innovadoras en los procesos de negocio. |
| | Disponibilidad | Los datos deberán estar disponibles para los usuarios autorizados, ayudándolos a discutir y solucionar problemas, así como la toma de decisiones. |
| Personalización | Disponibilidad | Los datos deberán estar disponibles para los usuarios autorizados, creando un ambiente de trabajo configurado especialmente para cada empleado. |
| Administración | Conformidad | Los datos deberán adherirse a normas o convenciones definidas por la organización. |
| | Eficiencia | Los datos puede ser accedidos con un nivel aceptable de funcionamiento. |
| | Precisión | Los datos deberán estar definidos acorde a la precisión requerida. |
| | Portabilidad | Los datos podrán ser instalados o substituidos por el Administrador del portal durante actividades de mantenimiento. |
| | Recuperabilidad | Los datos podrán ser mantenidos y recuperados por el Administrador del portal durante actividades de mantenimiento, o en caso de algún tipo de falla del sistema. |
| Seguridad | Confidencialidad | Los datos serán sólo accesibles por usuarios autorizados. |
| | Trazabilidad | Los datos proporcionan información de cuándo y quién lo utilizó. |

Tabla 3. Descripción de relaciones de las Dimensiones de DQ y las Funciones de un Portal

3.2. Metodología DQ-VORD

Como se mencionó al inicio de este trabajo, uno de los objetivos es introducir como requisitos software los aspectos de calidad de datos que deberían tener las aplicaciones Web, para lo cual se propone la incorporación de los mecanismos adecuados relacionados directamente con las diferentes dimensiones de calidad, como parte del proceso de desarrollo de software, particularmente en la fase de obtención de requisitos.

En este sentido, se tomó como referencia uno de los métodos principales de obtención de requerimientos “Orientados a Puntos de Vista” [14], con el objetivo de incorporar los aspectos de gestión de DQ durante el proceso, como si de requisitos normales se tratara.

La descripción de cada una de las etapas del método DQ-VORD, así como sus productos de entrada y salida relacionados se muestran a continuación:

1. *DQ-E1. Identificación de puntos de vista de un portal Web (IPV)*, que implica descubrir los que recibirán los servicios del portal Web (véase Tabla 4).

| | |
|-----------------------------------|---|
| Producto de Entrada | - Documento de Factibilidad del sistema. |
| Producto de Salida | - Lista de Puntos de Vista identificados en el sistema. |
| Técnica y/o herramienta propuesta | - Entrevistas con usuarios. - Observación. - Sesiones de trabajo grupales. - Estudio de documentación. |

Tabla 4. Artefactos para IPV.

2. *DQ-E2. Identificación de las funciones de un portal Web (IFP)*, implica identificar las funciones específicas que se suministran a cada punto de vista (véase Tabla 5).

| | |
|-----------------------------------|---|
| Producto de Entrada | - Lista de Puntos de Vista identificados. |
| Producto de Salida | - Listado de las funciones relacionadas a cada punto de vista. |
| Técnica y/o herramienta propuesta | - Entrevistas. - Estudio de documentación. - Cuestionarios. - Tormenta de ideas. |

Tabla 5. Artefactos para IFP.

3. *DQ-E3. Identificación de dimensiones de DQ (IDDQ)*, implica identificar las distintas dimensiones de calidad de datos relacionadas a cada una de las funciones descritas para cada punto de vista (véase Tabla 6).

| | |
|-----------------------------------|---|
| Producto de Entrada | - Lista de Puntos de Vista identificados. - Lista de funciones identificadas relacionadas a los puntos de vista. |
| Producto de Salida | - Listado de las dimensiones de calidad de datos asociadas a las distintas funciones. |
| Técnica y/o herramienta propuesta | - Entrevistas. - Sesiones de trabajo. - Lluvia de ideas. |

Tabla 6. Artefactos para IDDQ.

4. *DQ-E4. Estructuración de puntos de vista (EPV)*, comprende agrupar los puntos de vista relacionados en una jerarquía. Las funciones comunes se ubican en los niveles altos de la jerarquía y se heredan los puntos de vista de bajo nivel (véase Tabla 7).

| | |
|-----------------------------------|--|
| Producto de Entrada | - Lista de Puntos de Vista identificados. - Lista de funciones identificadas. |
| Producto de Salida | - Lista jerarquizada de los puntos de vista y funciones relacionadas. |
| Técnica y/o herramienta propuesta | - Sesiones de trabajo. - Juicio de expertos. |

Tabla 7. Artefactos para EPV.

5. *DQ-E5. Clasificación de las dimensiones de DQ (CDDQ)*, consiste en clasificar las dimensiones de calidad de datos, en base al nivel de prioridad que tienen las funciones del portal (véase Tabla 8).

| | |
|-----------------------------------|--|
| Producto de Entrada | - Lista jerarquizada de los puntos de vista y funciones relacionadas. - Listado de las dimensiones de calidad de datos asociadas a las distintas funciones. |
| Producto de Salida | - Lista de clasificación de las dimensiones de calidad de datos. |
| Técnica y/o herramienta propuesta | - Sesiones de trabajo - Juicio de expertos. |

Tabla 8. Artefactos para CDDQ.

6. *DQ-E6. Documentación de puntos de vista (DPV)*, que comprende refinar la descripción de los puntos de vista y las funciones identificadas (véase Tabla 9).

| | |
|-----------------------------------|---|
| Producto de Entrada | - Lista jerarquizada de los puntos de vista y funciones relacionadas. |
| Producto de Salida | - Documento de especificación de requisitos |
| Técnica y/o herramienta propuesta | - Juicio de expertos - Sesiones de trabajo - Herramientas como procesadores de texto. |

Tabla 9. Artefactos para DPV.

7. *DQ-E7. Documentación de las dimensiones de DQ (DDDQ)*, consiste en documentar y/o modelar de ser posible, las dimensiones identificadas (por ejemplo, mediante casos de uso). Véase Tabla 10.

| | |
|-----------------------------------|--|
| Producto de Entrada | - Lista de clasificación de las dimensiones de calidad de datos. - Documento de especificación de requisitos. |
| Producto de Salida | - Documento de especificación de requisitos con conciencia de calidad de datos. |
| Técnica y/o herramienta propuesta | - Sesiones de trabajo. - Juicio de expertos. - Herramientas como procesadores de texto - Herramientas de modelado de UML. |

Tabla 10. Artefactos para DDDQ.

8. *DQ-E8. Trazado del punto de vista del sistema (TPV)*, que comprende identificar los objetos en un diseño orientado a objetos utilizando la información del servicio encapsulado en los puntos de vista (véase Tabla 11).

| | |
|-----------------------------------|--|
| Producto de Entrada | - Documento de especificación de requisitos con conciencia de calidad de datos. |
| Producto de Salida | - Documento de diseño de alto nivel. |
| Técnica y/o herramienta propuesta | - Herramientas de modelado orientado a objetos (Rational Rose, Visual Paradigm, Poseidon, ArgoUML, etc.) |

Tabla 11. Artefactos para TPV.

9. *DQ-E9. Modelado de las dimensiones de DQ (MDQ)*, consiste en modelar las distintas dimensiones de calidad de datos en un modelo de datos y posteriormente en un modelo de procesos (véase Tabla 12).

| | |
|-----------------------------------|--|
| Producto de Entrada | - Documento de especificación de requisitos con conciencia de calidad de datos. - Documento de diseño de alto nivel |
| Producto de Salida | - Documento de diseño de alto nivel con conciencia de calidad de datos (Modelo de datos y de Procesos) |
| Técnica y/o herramienta propuesta | - Herramientas de modelado orientado a objetos (Rational Rose, Visual Paradigm, Poseidon, ArgoUML, etc.) |

Tabla 12. Artefactos para MDQ.

10. *DQ-E10. Validación del modelo (VM)*, consiste en validar con los stakeholders el modelo completo (véase Tabla 13).

| | |
|-----------------------------------|---|
| Producto de Entrada | - Documento de especificación de requisitos con conciencia de calidad de datos. - Documento de diseño de alto nivel con conciencia de calidad de datos |
| Producto de Salida | - Documento final aprobado de requisitos del sistema Web. |
| Técnica y/o herramienta propuesta | - Sesiones de trabajo. - Técnicas de negociación interpersonal. |

Tabla 13. Artefactos para VM.

En la Figura 2 se muestra mediante un diagrama de actividades las distintas etapas de la metodología DQ-VORD.

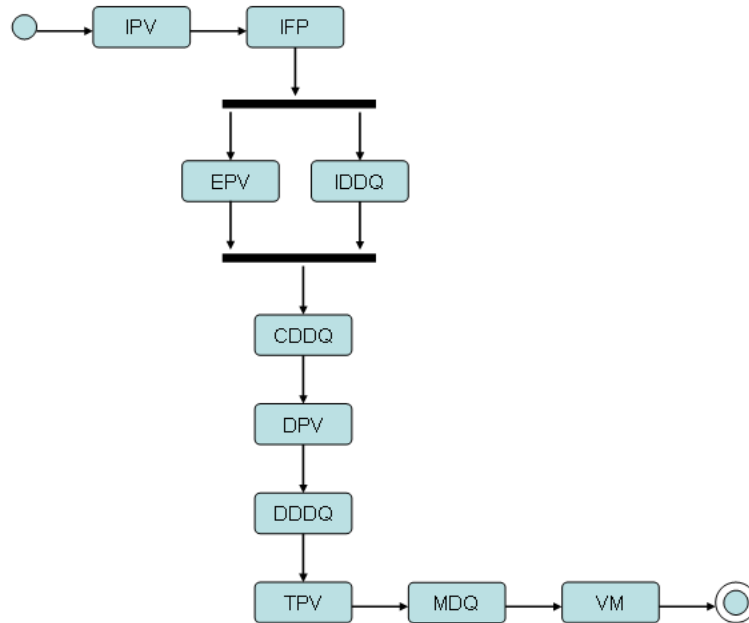


Figura 2. Etapas de la metodología DQ-VORD

Las siguientes plantillas pueden ser llenadas, a manera de resumir y complementar los principales aspectos del método, la primera de ellas está relacionada a los puntos de vista (véase Tabla 14), y la segunda plantilla orientada a los servicios (véase Tabla 15).

| | |
|-------------|---|
| Referencia | Nombre del punto de vista |
| Atributos | Descripción de las características propias del punto de vista. |
| Eventos | Referencia a un conjunto de escenarios describiendo como reacciona el sistema a eventos específicos del punto de vista. |
| Servicios | Referencia a un conjunto de descripciones de los servicios. |
| Excepciones | Descripción de las excepciones que habrá de manejar el sistema. |

Tabla 14. Plantilla de Punto de Vista.

| | |
|--------------------------------|---|
| Referencia | Nombre del servicio |
| Razón | Justificación y/o descripción del servicio. |
| Especificación | Referencia a una lista de especificaciones de servicio (en caso que exista) pudieran estar en distintas notaciones. |
| Puntos de Vista | Lista de Puntos de Vista que reciben el servicio. |
| Requisitos No Funcionales | Descripción de los requisitos no funcionales relacionados con este servicio. |
| Requisitos de Calidad de Datos | Descripción de los requisitos de calidad de datos relacionados con este servicio. |

Tabla 15. Plantilla de Servicios.

4. Conclusiones y Tendencias Futuras

De forma ideal uno desearía imaginar a un analista o diseñador que, usando una simple herramienta de desarrollo de software, pudiera ser capaz de desarrollar un sistema completo, gestionando desde las etapas iniciales todos los requerimientos (funcionales, no funcionales, de calidad de datos), diseño (alto y bajo nivel), pruebas, hasta la generación automática de código.

Un primer acercamiento a esta solución es la mostrada en este trabajo, en el cual se describe cuáles dimensiones de calidad de datos estarían presuntamente implicadas directamente con las distintas funciones de un portal Web, así como la propuesta de un mecanismo para la obtención y gestión de requisitos de calidad de datos, lo que en un momento dado, facilitaría a los desarrolladores a tener conciencia del nivel de calidad de datos que necesita implementarse para cada una de las funciones durante todo el proceso de desarrollo de software.

Como parte del trabajo futuro se pretende realizar un estudio para la incorporación del enfoque de MDA (Model Driven Architecture), para el modelado de las dimensiones de DQ y su incorporación al proceso de desarrollo de software.

El enfoque dirigido por modelos en general y la propuesta de MDA en particular, han dado un giro para ser una gran ventaja en el desarrollo de software. La separación en niveles de abstracción (CIM, PIM y PSM), así como la definición de reglas de transformación entre los modelos definidos dentro de esos niveles y entre niveles, son los fundamentos de la propuesta MDA. Este enfoque también tiene la ventaja de ser acatador de las tecnologías de UML2 y OMG (Grupo de Gestión de Objetos), asegurando una integración mucho más fácil con otras herramientas y metodologías que obedezcan estándares similares.

5. Referencias

1. Caballero, J., et al. MMPRO: A Methodology Based on ISO/IEC 15939 to Draw Up Data Quality Measurement Processes. in *ICIQ*. 2008.
2. Caballero, J., et al. DQRDFS: Towards a Semantic Web Enhanced with Data Quality. in *Web Information Systems and Technologies*. 2008. Funchal, Madeira, Portugal.
3. Garc, A., et al., A proposal for a set of attributes relevant for Web Portal Data Quality. *Software Quality Journal*, 2008.
4. Garc, A.S., Modelo de Calidad de Datos para Portales Web, in *Tecnologías y Sistemas de Información*. 2007, Universidad de Castilla-La Mancha: Ciudad Real, España.
5. Collins, H., Corporate Portal Definitions and Features. 2001, New York, NY, U.S.A.: Amacom Books.
6. Eppser, M. and M. Helffert. A Classification and Analysis of Data Quality Costs. in *International Conference on Information Quality*. 2004. MIT, Cambridge, MA, U.S.A.

7. Guerra García, G., Obtención de Requerimientos. Técnicas y Estrategias. *Software Guru*, 2007: p. 28-31.
8. ISO-25012, ISO/IEC 25012: Software Engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data Quality Model. 2008.
9. Kotonya, G. and J. Sommerville, Requirements engineering with viewpoints. *Software Engineering Journal*, 1996. 11 (1): p. 5-18.
10. Laudon, K. G., Data Quality and Due Process in Large Interorganizational Record System. *Communications of the ACM*, 1986. 29 (1): p. 4-11.
11. Lee, Y. W., et al., Journey to Data Quality. 2006, Cambridge, MA, USA: Massachusetts Institute of Technology.
12. Levis, M., M. Helfert, and M. Brady. Information Quality Management: Review of an Evolving Research Area. in ICIQ'07. 2007. MIT, Cambridge, MA, USA.
13. Loucopoulos, P. and P. Karakostas, System Requirements Engineering. 1995. New York NY USA: McGraw-Hill.
14. Sommerville, J., Software Engineering. 6th Edition. 2000.
15. Strong, D., Y. Lee, and R. Wang, Ten Potholes in the Road to Information Quality. *IEEE Computer*, 1997: p. 38-46.
16. Strong, D. M., Y. W. Lee, and R. Y. Wang, Data Quality in Context. *Communications of the ACM*, 1997. 40 (5): p. 103-110.
17. Wang, R. Y. and S. Madnick. Data Quality Requirements: Analysis and Modelling. in Ninth International Conference on Data Engineering (ICDE'93). 1993. Vienna, Austria: IEEE Computer Society.