

**Universidad Latinoamericana de Ciencia y Tecnología (ULACIT)**

**Facultad de Ingeniería**

**Escuela de Ingeniería Informática**

**Trabajo final para optar al grado de licenciado en Ingeniería  
Informática con énfasis en Desarrollo de Software**

**Calidad de Datos**

**Autor: Minor Zúñiga Sedó**

**Cédula: 1-1013-0964**

**Profesor tutor: Miguel Pérez Montero**

**Agosto del 2007**

## ÍNDICE

<b>Introducción</b> .....	1
<b>¿Qué es calidad de datos?</b> .....	2
<b>¿Cómo se originan los problemas de calidad de datos en los sistemas?</b> .....	3
Aspectos de formación académica .....	4
<b>Impacto de la mala calidad de los datos</b> .....	5
Internet.....	6
<b>Atributos, dimensiones y categorías de datos</b> .....	7
<b>Herramientas para facilitar los procesos de la calidad de los datos</b> .....	10
Metadatos .....	11
<b>Pasos para aplicar procesos de calidad de datos en las organizaciones</b> .....	12
El ciclo de los datos .....	12
¿Cómo aplicar un proceso de calidad de datos en las organizaciones? .....	14
<b>Beneficios de los procesos de calidad de datos</b> .....	19
<i>E-business</i> .....	19
Determinando el grado de recuperación de inversión en proyectos de calidad de los datos .....	22
<b>Conclusiones</b> .....	23
<b>Bibliografía</b> .....	24

# CALIDAD DE DATOS

## Resumen

En la actualidad, la información se ha vuelto clave para las organizaciones independientemente del mercado en el que compitan. Adicionalmente, los sistemas informáticos en los que se almacena esta información son indispensables para que la organización pueda llevar a cabo sus procesos de negocio. Por tanto, garantizar y mantener una buena calidad en los datos es crucial para lograr cumplir metas y objetivos.

Este artículo presenta una mirada al concepto de “calidad de datos” en las organizaciones, analiza los principales problemas asociados con los datos incorrectos y muestra cómo llevar a cabo un proceso cíclico e iterativo de análisis y mejora de la información aplicando una estructura jerárquica y multidimensional.

**Palabras clave:** calidad de datos, atributos de datos, dimensiones de calidad, categorías de calidad.

## Abstract

Nowadays the information has become key for the organizations independently of the market in which they compete. Additionally, the computer systems in which this information is stored are indispensable for the organization so they can carry out his processes of business. Therefore, to guarantee and to maintain a good quality in the data become crucial to fulfill its goals and objectives.

This article presents a look to the concept of Data Quality in the organizations, it analyzes the main problems associated with the incorrect data and shows how to carry out a cyclical and iterative process of analysis and improvement of the information applying a hierarchic and multidimensional structure.

**Keywords:** data quality, data attributes, quality dimensions, quality categories.

## **Introducción**

El siglo XXI se ha definido como la “Era de la Información”, por lo que esta se ha convertido en el producto más importante para las empresas y organizaciones. Para obtener una ventaja competitiva en los negocios es necesario obtener y analizar información de todos los tipos y fuentes. La información permite conocer a fondo a los clientes y al mercado, lo que representa una ventaja para la organización.

El rol de la información para obtener ventajas competitivas es crucial; sin embargo, la información por sí sola no crea conocimiento ni ventajas, es necesario que esta sea de calidad y analizada en diferentes contextos para que se convierta en una ventaja competitiva.

Actualmente, las organizaciones no solo dependen de la tecnología de la información en un nivel de soporte para los procesos organizacionales, sino que también la utilizan en la toma de decisiones, con lo que se abarcan de esta forma también los procesos estratégicos para la consecución de los objetivos del negocio.

Conforme la tecnología se convierte en una herramienta indispensable para que las empresas realicen sus actividades cotidianas, la información que se almacena en los sistemas transaccionales se transforma en el activo más importante de la organización; de esta forma, un error o inconsistencia en los datos podría generar un impacto negativo en aspectos operativos, monetarios, funcionales, tácticos, técnicos y estratégicos. La mala calidad de datos puede costarle a la organización clientes potenciales y afectar la imagen en el mercado. Como consecuencia de esto, en los últimos años, los gerentes de las organizaciones han tenido que implementar proyectos para mejorar la calidad de los datos en los sistemas e instaurar mecanismos y controles que minimicen la entrada de datos inconsistentes en los repositorios de datos primarios.

Para lograr el éxito con una iniciativa de calidad de datos es necesario analizar la información tanto desde una perspectiva técnica (verificar el cumplimiento de requerimientos o especificaciones), como desde una perspectiva de usuario (validar si es apta para su utilización en el contexto del negocio), a fin de determinar si un dato es correcto o no. Para llevar a cabo estos análisis, existen herramientas que permiten verificar grandes cantidades de datos basados en el concepto de dimensiones de calidad. La interpretación de los resultados obtenidos en el análisis nos puede llevar a determinar las causas de los errores de datos y, por ende, a implementar las correcciones del caso, con la finalidad de evitar la propagación de inconsistencias en los sistemas.

En este documento se analizan algunas de las causas que provocan una mala calidad de datos en los sistemas, el impacto que genera esta problemática así como los pasos para aplicar una iniciativa de calidad de datos en las organizaciones, la cual derivará en beneficios inmediatos para la empresa y podrá

ayudar a implementar procesos de mejoramiento continuo y a desarrollar proyectos para la utilización de la información de la empresa bajo un enfoque estratégico como la Minería de Datos<sup>1</sup> (en inglés, *Data Mining*).

### ¿Qué es calidad de datos?

Se puede decir que no existe una definición simple para el concepto de “calidad de datos”; no obstante, aspectos como la exactitud, la consistencia y la completitud son utilizados comúnmente para representarla. Para Ballou y Tayi (1998), un dato puede considerarse de calidad si cumple con su propósito y es apto para utilizarse, es decir, el dato se analiza desde un enfoque de exactitud, consistencia y completitud, pero referenciados en un contexto de usuario más que en un contexto puramente técnico.

Entre otras definiciones del concepto de calidad de datos, podemos mencionar las siguientes:

- Para J.M. Juran (1989), los datos pueden considerarse de calidad si cumplen con su propósito desde una perspectiva operativa, de toma de decisiones y de planeación.
- Glosario GIS (2007): La calidad de datos se refiere al grado de excelencia que posean los datos en relación con el ámbito en que se encuentren definidos.
- *Government of British Columbia* (2007): La calidad de datos es el grado de completitud, consistencia, persistencia y exactitud que hace un dato apropiado para un uso específico.
- *Glossary of Quality Terms* (1997): Calidad de datos es la totalidad de características que les permiten a los datos cumplir con un propósito específico.

Por otra parte, en el estándar ISO<sup>2</sup> 8402: *Quality Management and Quality Assurance Vocabulary*, se establece una definición formal de calidad como:

---

<sup>1</sup> Según Wikipedia (2007), “bajo el nombre de minería de datos se engloban un conjunto de técnicas encaminadas a la extracción de “conocimiento” procesable implícito en las bases de datos de las empresas”.

<sup>2</sup> Según Wikipedia (2007), “la Organización Internacional para la Estandarización o International Organization for Standardization (ISO), es una organización internacional no gubernamental, compuesta por representantes de los organismos de normalización (ONs) nacionales, que produce normas internacionales industriales y comerciales. Dichas normas se conocen como normas ISO y su finalidad es la coordinación de las normas nacionales, en consonancia con el Acta Final de la Organización Mundial del Comercio, con el propósito de facilitar el comercio y el intercambio de información, y contribuir con unos estándares comunes para el desarrollo y transferencia de tecnologías”.

“Conjunto de propiedades y características de un producto o servicio, que le confieren aptitud para satisfacer necesidades explícitas e implícitas” (pág. 25).

Adicionalmente, el estándar ISO 8402 ofrece las siguientes definiciones referidas al concepto de calidad de datos:

- Control de Calidad: “Conjunto de técnicas y actividades de carácter operativo, utilizadas para verificar los requerimientos relativos a la calidad del producto o servicio” (pág. 37).
- Garantía de Calidad: “Conjunto de acciones planificadas y sistemáticas necesarias para proporcionar la confianza adecuada de que un producto o servicio satisfará los requerimientos dados sobre calidad” (pág. 37).
- Gestión de Calidad: “Aspecto de la función de gestión que determina y aplica la política de calidad de datos, los objetivos y las responsabilidades y que lo realiza con medios tales como la planificación de la calidad, el control de la calidad, la garantía de la calidad y la mejora de la calidad” (pág. 38). La gestión de la calidad es responsabilidad de todos los niveles ejecutivos de la organización, pero debe ser guiada por la alta dirección.
- Sistema de Gestión de Calidad: “Conjunto de la estructura de la organización, de responsabilidades, procedimientos, procesos y recursos que se establecen para llevar a término la gestión de calidad” (pág. 38). Debe contener los alcances suficientes que permitan conseguir los objetivos de calidad y estar previsto para satisfacer las necesidades específicas de la organización.

De esta forma, se puede decir que un dato cumple con un nivel de calidad si satisface los requerimientos establecidos para un rol en particular y si esa especificación refleja las necesidades implícitas y explícitas de los usuarios.

Por tanto, se alcanza un nivel aceptable de calidad si los datos se encuentran de acuerdo con las especificaciones (son conformes) y sirven para lo que fueron definidos (son útiles). Estos dos conceptos son concordantes con el estándar ISO 8402 y se asemejan a otras referencias que definen calidad como “satisfacción del cliente”, “acorde con su uso”, “acorde con su propósito” o “de acuerdo a los requerimientos” (Juran 1989, pág. 57., Crosby 1980, pág. 46).

### **¿Cómo se originan los problemas de calidad de datos en los sistemas?**

Los datos incorrectos se pueden generar de muchas formas, por errores en la entrada de datos, información incorrecta cargada desde formularios, discrepancias entre diferentes sistemas o datos incorrectos obtenidos por fuentes externas.

También es común encontrar datos correctos mezclados con información desactualizada, lo que dificulta distinguir los datos válidos de los inválidos.

Para Fisher y Marinos (2003), algunos de los problemas más comunes relacionados con calidad que se pueden encontrar en los sistemas son:

- Problemas de estandarización: Errores de ortografía en los nombres de clientes (personas, organizaciones, empresas, etc.) así como en el uso de abreviaturas.
- Inconsistencia en los formatos: Son datos que no cumplen con un formato definido. Se presenta especialmente en los campos tipo fecha; por ejemplo, se tienen datos almacenados con el formato dd/mm/aaaa u otros bajo aaaammdd.
- Codificación incorrecta: Datos almacenados que no corresponden a valores válidos de un catálogo.
- Datos incorrectos: Se permite incluir datos que cumplen con el formato, pero que no son válidos. Dentro de estos podemos mencionar, por ejemplo, la fecha 01/01/1900 o en nuestro país, el número de cédula 0-0000-0000.
- Datos que no cumplen con las reglas del negocio: Por ejemplo, un salario que se sale de la escala de máximos y mínimos establecido en la organización.
- Datos estáticos: Corresponde a información que cambió en el tiempo y que hoy en día está desactualizada, por lo que no genera un valor para el negocio. Un ejemplo de esto son las direcciones o números telefónicos desactualizados.
- Datos ambiguos entre sistemas que no permiten su consolidación: Este caso se presenta cuando se intenta unificar información de dos sistemas en los cuales la misma información se encuentra almacenada bajo llaves primarias diferentes o con formatos diferentes. Por ejemplo, dos catálogos maestros de clientes, uno en el cual la llave primaria fue generada a partir de una combinación de número de persona física-jurídica y nombre del cliente, y otra en la que la llave primaria es una codificación interna consecutiva del sistema.

### Aspectos de formación académica

Si se ve el tema de calidad de datos desde la perspectiva de formación de los profesionales de la computación y la informática, se determinará que el tema es prácticamente ignorado en todos los programas de las carreras universitarias.

Por lo general, el tema se aborda indirectamente a través de los cursos de bases de datos, al analizar aspectos relacionados con la integridad de datos en las bases de datos relacionales, es decir, a través del diseño del modelo relacional de las bases de datos. Este es otro de los factores que, según expertos, genera que en los sistemas de información no se plasmen suficientes reglas de negocio y validaciones que minimicen la incidencia de errores que conducen a almacenar datos inconsistentes.

### **Impacto de la mala calidad de los datos**

Los avances en la tecnología le han permitido a las organizaciones crear, almacenar y procesar grandes cantidades de información. Conforme se ha incrementado el uso de la información almacenada en los sistemas para mantener los procesos organizacionales, es evidente que los problemas relacionados con datos inconsistentes afectarán negativamente la organización en términos de eficiencia y eficacia.

A pesar del costo que representa tener información incorrecta, las organizaciones por lo general ignoran el tema de la calidad de los datos hasta que este se convierte en un riesgo palpable o un obstáculo para completar los objetivos organizacionales. La mala calidad de los datos en los sistemas de información es un tema que impacta negativamente las organizaciones día con día, ya sea en el orden público o en el privado, por lo se ha convertido en uno de los problemas críticos que enfrentan las organizaciones en la actualidad. Conforme la información almacenada en los sistemas se vuelve indispensable para alcanzar los objetivos organizacionales, la calidad de los datos se convierte en un tema importante para las organizaciones.

El costo de una mala calidad de datos en los sistemas se puede ilustrar desde varios puntos de vista, por ejemplo, una reducción en el nivel de satisfacción de nuestros clientes y, por consiguiente, un deterioro en la imagen de la organización ante el público en general o un incremento en los gastos asociado a problemas relacionados con aspectos operacionales, tácticos y estratégicos.

Es más, el problema de la mala calidad de los datos puede ser tan grave que podría extenderse más allá del nivel organizacional e impactar el mercado en sí, puesto que cada vez que los datos son transferidos entre sistemas, existe el riesgo de que información incompleta o incorrecta genere una falla generalizada. La mala calidad de los datos es persistente y costosa para la industria en general, y al referirnos a “la industria” no debemos limitarnos al campo de la informática y computación, sino al mercado en general, indistintamente del ámbito de desempeño.

Para algunos, los problemas de calidad de datos se encuentran intrínsecamente relacionados con el volumen de información almacenada en los sistemas de información así como en la manera de obtener los datos, de esta forma, los sistemas que se alimentan de múltiples fuentes de información son más susceptibles de tener inconsistencias que aquellos que utilizan solamente una o dos fuentes.

### Internet

En la actualidad, todo se mueve en torno a la Internet. Los negocios que no tienen presencia en la web se encuentran claramente en desventaja. En la web todo se resume al contenido, pero, ¿qué sucede si el contenido es incorrecto? La respuesta es obvia, posiblemente los usuarios no podrán utilizar el sitio (esto indistintamente del tipo de sitio al que nos estemos refiriendo, ya sea buscadores de sitios web, tiendas virtuales o portales con información general como sitios de noticias, por ejemplo).

Los siguientes son algunos ejemplos simples de problemas generados por la mala calidad de los datos en los sitios web:

1. Imagine que un cliente busca en una tienda virtual un disco compacto de música, una película en DVD o un libro, pero por un problema con la información del título o el autor almacenado en el sistema, el cliente no obtiene un resultado positivo de su búsqueda. Probablemente lo que este usuario haría es buscar otra tienda que “sí posea” lo que buscaba.
2. Un usuario realiza una transacción monetaria un día en cualquier parte de Centroamérica, pero la fecha que queda almacenada en los sistemas es la fecha del servidor, la cual se encuentra ubicada, por ejemplo, en Europa, al otro lado del mundo. Esta discrepancia de fechas que parece irrelevante, podría generar inconvenientes financieros relacionados con factores para realizar cálculos de intereses o el establecimiento de fechas límites para transferencias.

En cualquiera de los casos, un error en los datos produce que el sitio sea inservible para un usuario promedio. Errores como estos podrían costarle a la empresa perder clientes potenciales y generar un deterioro en su imagen ante el público en general, ya que de todos es conocido lo fácil que se puede divulgar la mala publicidad a través de Internet.

Un factor que afecta tanto la facilidad de uso como la velocidad de los sitios web es que los clientes deban digitar la información por ellos mismos en los sistemas. Factores como la falta de cuidado en el suministro de información así como los errores de tecleo en los formularios web, generan el riesgo de que información incorrecta sea almacenada en los sistemas.

Producto de esta interacción entre clientes y proveedores se han generado muchos cambios en la forma de manejar la información. La pregunta que se plantea, entonces, es si se podrá integrar la información de baja calidad capturada mediante los formularios web con la información de alta calidad que se almacena en los sistemas operacionales en la organización, conociendo de previo que la información web es susceptible de errores e inconsistencias y en donde no todos los campos son requeridos. Para lograr esta integración de datos de baja calidad a los sistemas operacionales, es necesario llevar a cabo un proceso para completar, validar y estandarizar la información antes de trasladarla. Este proceso puede llevarse a cabo mediante un centro de llamadas<sup>3</sup> (en inglés, *Call Center*) o utilizando otras fuentes de datos (catálogos de información). Una vez integrada la información, se podrá compartir en toda la organización y utilizarse en la toma de decisiones.

### **Atributos, dimensiones y categorías de datos**

Kahn, Strong y Wang (2002) mencionan que existen varios marcos teóricos que permiten entender de una mejor manera el concepto de calidad de datos, uno se enfoca desde la perspectiva del producto (cumplimiento de las especificaciones), mientras que otro se enfoca desde la perspectiva del servicio (alcanzar las expectativas del cliente).

Para lograr una medición de “conformidad” y “utilidad”, es necesario contar con un set de características o atributos de los datos. Por lo general, estas características se encuentran definidas por atributos tales como exactitud, completitud, consistencia, relevancia, persistencia, unicidad y validez.

Dependiendo de los requerimientos del sistema de información, el grupo de atributos puede variar. Sin embargo, aun y cuando se defina el grupo correcto de atributos, es claro que existirá un grado de interdependencia entre ellos, por lo cual los análisis de los atributos no deben realizarse de forma aislada.

En lugar de realizar iniciativas de calidad de datos desde un nivel de atributos, los expertos coinciden en la necesidad de organizar los atributos para poder identificar los problemas de la calidad de los datos. De esta forma, los atributos que poseen relación entre sí se agrupan creando dimensiones. Para Abate, Diegert y Allen (1998), agrupar atributos en dimensiones tiene las siguientes ventajas:

- Las dimensiones son más fáciles de entender que los atributos.

---

<sup>3</sup> Según Wikipedia (2007), “un centro de llamada (o “Call Center” en inglés) es un área donde agentes, especialmente entrenados realizan llamadas (llamadas salientes o en inglés, *outbound*) o reciben llamadas (llamadas entrantes o *inbound*) llamadas desde y/o hacia: clientes (externos o internos), socios comerciales, compañías asociadas u otros”.

- Al unir atributos interdependientes en dimensiones, los analistas de calidad pueden consolidar y organizar la información necesaria para la interpretación y comprensión de los datos de manera más integral.
- Las dimensiones ayudan a los analistas de calidad a identificar problemas sistemáticos de calidad en las aplicaciones.

Por su conceptualización, las dimensiones de la calidad de los datos han sido comparadas por algunos expertos en ingeniería del software con el concepto de “requerimientos no funcionales”, es decir, aquellas situaciones que no pueden ser plasmadas dentro de un desarrollo de un sistema, pero que son requeridas para su correcto funcionamiento; por ejemplo, la oportunidad de la información o qué tan rápido responde el sistema a una petición del usuario.

Wang (1994) discutió cómo construir dimensiones específicas de la calidad de los datos y agrupó 179 atributos en 15 dimensiones, las cuales se muestran en la siguiente tabla:

<b>Tabla 1: Dimensiones de la calidad de los datos</b>	
<b>Dimensión</b>	<b>Descripción</b>
Seguridad de acceso	El acceso a los datos debe ser restringido, mantenerse seguro.
Accesibilidad	Los datos deben estar disponibles de forma ágil y sencilla.
Exactitud	Los datos deben ser correctos, confiables, libres de errores.
Cantidad apropiada de datos	La cantidad o volumen de información disponible debe ser la apropiada.
Credibilidad	Los datos se deben considerar como verdaderos, reales y creíbles.
Complejidad	Los datos deben tener la longitud, profundidad y alcance para el fin establecido.
Representación concisa	La representación de los datos debe ser la correcta.
Facilidad de entendimiento	Los datos deben ser claros, sin ambigüedades y de fácil comprensión.
Interpretabilidad	Los datos deben estar en el lenguaje y unidades correctas; su definición debe ser la apropiada.
Objetividad	Los datos deben ser imparciales.
Relevancia	Los datos deben ser de utilidad y aplicables al ámbito de competencia.
Consistencia en la representación	La información debe presentarse siempre con el mismo formato y ser compatible con información previa.
Reputación	Los datos deben ser confiables en

	términos de su fuente o contexto.
Persistencia en el tiempo	La "edad" de los datos debe ser apropiada para su utilización actual
Valor agregado	Los datos deben ser beneficiosos y proveer ventajas para su uso.

Fuente: Abate, Diegert y Allen (1998).

Mientras los atributos muestran el nivel más bajo en que los problemas de calidad de datos pueden ser definidos y entendidos, las dimensiones muestran un nivel más alto de entendimiento.

Generalmente, las deficiencias en la calidad de los datos se detectan a nivel de los atributos; sin embargo, detenerse en este nivel y generar conclusiones puede llevar a obviar situaciones más graves de calidad que pueden visualizarse cuando se realizan análisis a nivel de dimensiones. Al analizar resultados agrupando los atributos en las dimensiones, podremos detectar el origen de las inconsistencias, lo que nos permitirá establecer soluciones acordes con cada problema.

Ahora bien, los datos pueden presentar problemas en una dimensión y ser correctos en otras. Cuando esto sucede es posible que la situación que origina el problema pueda provocar inconsistencias en varias dimensiones. Es por esto que, al igual que con los atributos, las dimensiones pueden agruparse para facilitar el análisis y poder encontrar patrones de problemas de calidad. Wang (1994), propuso la siguiente agrupación de dimensiones se acuerdo con cuatro categorías:

<b>Categoría</b>	<b>Dimensiones</b>	<b>Indicadores de deficiencias</b>
Intrínsecas	Exactitud, objetividad, credibilidad y reputación	Carencia o debilidad en los procesos para crear datos que correspondan a valores actuales o verdaderos.
Contextuales	Valor agregado, relevancia, persistencia en el tiempo, completitud y cantidad apropiada de datos	Carencia o debilidad en los procesos para generar la información pertinente para que los usuarios lleven a cabo sus tareas.
Representación	Interpretabilidad, fácil de entender, representación consistente, representación concisa	Carencia o debilidad en los procesos para suministrar datos claros.

Accesibilidad	Accesibilidad y seguridad en el acceso	Carencia o debilidad en los procesos para proporcionar datos de manera ágil y segura
---------------	--	--

Fuente: Abate, Diegert y Allen (1998).

Las categorías mostradas en la Tabla 2 representan un tercer pilar de entendimiento para los problemas de la calidad de los datos. En este tercer nivel de análisis es posible detectar las constantes y debilidades en los procesos que generan las condiciones apuntadas en el segundo nivel. El tercer nivel de entendimiento es necesario para poder proponer soluciones a las causales originarias de los problemas de calidad.

Las debilidades reveladas en cada una de las categorías corresponden a una deficiencia funcional o conceptual y, por tanto, a un problema de calidad de datos. Cuando se presentan deficiencias funcionales, es probable que sea necesario cambiar los procesos o sistemas utilizados para adquirir los datos, mientras que las deficiencias conceptuales pueden generar cambios tanto en los procesos utilizados para capturar la información como en los procesos utilizados para manipular los datos.

### **Herramientas para facilitar los procesos de la calidad de los datos**

Actualmente, el mercado provee herramientas que permiten mejorar la calidad de los datos en los sistemas y realizar análisis de algunas de las dimensiones de calidad indicadas anteriormente. Adicionalmente, otras herramientas más avanzadas permiten detectar inconsistencias antes de que los datos se almacenen en las bases de datos.

La mayoría de las herramientas de calidad de datos facilitan opciones para llevar a cabo las siguientes acciones:

- ❖ Data Profiling: Corresponde a un proceso de evaluación primaria de los datos que permite establecer un índice inicial de calidad y determinar, de forma general, posibles inconsistencias en los datos desde una perspectiva meramente técnica.
- ❖ Estandarización de los datos: Permite definir reglas de negocio para aplicar a los datos y garantizar que estos están de acuerdo con lo establecido.
- ❖ Geocodificación: Corresponde a una funcionalidad que permite corregir problemas en la información de las direcciones generadas por errores de codificación. Esta funcionalidad es muy utilizada en Europa y los Estados Unidos, puesto que en estos lugares existen catálogos de información

postal que facilitan la comparación y validación de información de localización de manera automatizada.

- ❖ Macheo: Permite comparar datos que son similares pero distintos a nivel de sistemas, con la finalidad de unificarlos en un único registro. Su función principal es determinar si existe duplicidad de información y brindar soluciones a dicha problemática. Para este caso, algunas herramientas cuentan con algoritmos sumamente complejos que permiten realizar comparaciones fonéticas o literales para determinar candidatos a ser “registros similares” en una vasta cantidad de datos.
- ❖ Monitoreo: Permite realizar un seguimiento a nivel de la calidad de los datos en el tiempo, así como modificar los estándares de calidad definidos según las necesidades del momento.
- ❖ Análisis en tiempo real y definición de procesos fuera de línea (Batch): Una vez que los datos se encuentran estandarizados y se ha realizado una limpieza de la información, algunas herramientas proporcionan opciones de validación en línea que permiten generar interfases entre las aplicaciones y la base de datos. Estos elementos intermedios facilitan, por ejemplo, capturar la información que es digitada por un usuario en un formulario web y aplicarle algoritmos de validación y estandarización a la información antes de almacenar los datos en el repositorio. Igualmente, algunas herramientas posibilitan el desarrollo de rutinas o procesos de estandarización o “macheo” de información para ser ejecutadas en momentos específicos, de esta forma, posterior a realizar un proceso de limpieza, se podría establecer la ejecución de una rutina de estandarización o *match* con una regularidad diaria o semanal, dependiendo de factores de incidencia de inconsistencias.

## Metadatos

Los metadatos son datos acerca de los datos, es información que nos permite conocer ciertas características de los datos (tipo, tamaño, valores válidos, precisión numérica, etc.). Esta información es utilizada para validar si un dato es correcto o incorrecto y si cumple con su propósito, lo que genera un beneficio obvio cuando se realizan proyectos de calidad de datos.

Los metadatos pueden proporcionar más información relevante sobre el contexto de los datos, que aplicar un cuestionario a un usuario para obtener esa información. Los metadatos pueden obtenerse desde dos perspectivas, una vista técnica y una vista de negocio. La perspectiva de negocio nos muestra características sobre qué representan los datos más allá de la representación misma de los datos, aspectos que obtenemos al analizar la información meramente desde una perspectiva técnica.

Las herramientas de calidad de datos existentes pueden obtener información a partir de los metadatos almacenados en las bases de datos relacionales; igualmente, existen algunos productos que permiten complementar esta información con ciertas reglas de negocio, lo que facilita los análisis de los datos y la detección de posibles errores.

## **Pasos para aplicar procesos de calidad de datos en las organizaciones**

Por lo regular, el área de tecnologías de la información en las empresas se encarga de administrar los recursos tecnológicos, no así los datos, puesto que estos son responsabilidad de los usuarios de las aplicaciones, es decir, de la organización en sí. Sin embargo, como hemos visto, la calidad de los datos es un tema que debe visualizarse desde la perspectiva del negocio, por lo que es indispensable la participación tanto de la parte de tecnologías como de los usuarios.

### El ciclo de los datos

Los datos deben visualizarse como activos del negocio, así como lo son el capital, los materiales, la maquinaria, la tecnología o los recursos humanos. Las organizaciones deben reconocer que los datos y la información son un activo del negocio y definir responsables y responsabilidades de administrar estos recursos. Al contrario de la mayoría de activos de las organizaciones, los datos son altamente dinámicos. Entender el ciclo de vida de los datos es importante para entender su naturaleza, lo que nos permitirá explicar de mejor forma los pasos para aplicar un proceso de calidad de datos. Redman (1996) propuso un ciclo de vida para los datos que incluye dos fases: el ciclo de adquisición de los datos y el ciclo de uso de los datos.

En el ciclo de adquisición de los datos, el objetivo es justamente almacenar los datos en el repositorio. El modelado de los datos y la obtención de la información son componentes necesarios para el ciclo de adquisición de los datos. Este ciclo se compone de cuatro pasos:

1. Definición del modelo lógico de los datos o “definición de la vista”.
2. Diseño físico de los datos e implementación o “implementación de la vista”.
3. Obtención y carga de los datos (“llenar” la base de datos).
4. Actualización de los datos (almacenamiento y mantenimiento de los datos).

Las fases de definición, diseño e implementación son encargadas, por lo general, a grupos técnicos, es decir, la parte informática de la organización; mientras que el proceso de llenado de la base de datos recae usualmente en la contraparte usuaria, es decir, en quienes van a utilizar la información.

El ciclo de utilización de los datos comienza una vez que los datos han sido almacenados. Además de utilizarse en la operación regular de la organización, los datos también son necesarios para los procesos de toma de decisiones, los cuales requieren combinar una gran cantidad de datos, trasladarlos a un contexto diferente y, posteriormente, interpretar los resultados obtenidos del análisis.

Los cuatro pasos en este ciclo son los siguientes:

1. Diseñar la consulta o “definir la subvista”.
2. Obtener la información o “procesar la consulta”.
3. Manipular la información (ordenar, analizar, formatear y agregar datos).
4. Presentación de la información en el nuevo contexto.

Al igual que en el ciclo anterior, los pasos de diseño de la consulta y procesamiento recaen sobre los informáticos, mientras que en las fases de manipulación y presentación de resultados, son los usuarios o encargados del negocio quienes los realizan.

Según recalcan los expertos, los problemas relacionados con la poca calidad de los datos en los sistemas informáticos se atribuyen a los siguientes factores:

1. Problemas relacionados con los procesos.
2. Problemas relacionados con los sistemas.
3. Problemas relacionados con las políticas o procedimientos organizacionales.
4. Problemas relacionados con el diseño de los datos.

Entender los procesos que generan, utilizan y almacenan los datos es esencial en las iniciativas de calidad de los datos. Redman (1996) recomienda el uso de un modelo denominado Funciones de Procesamiento de Información (en inglés, *Functions of Information Processing* o FIP), para ayudar a los usuarios o “dueños” a describir las cadenas de información. Los diagramas FIP modelan cómo los datos son creados, transferidos, almacenados, filtrados y asociados en una cadena de información, lo que resulta bastante útil para identificar fuentes de problemas de calidad de datos.

Según los expertos, los principales problemas de la calidad de los datos se originan por inconvenientes en los procesos.

Mejorar la calidad de los datos es una tarea complicada, puesto que por lo general, la información se encuentra distribuida en la organización, almacenada en fuentes diversas. Un programa de calidad es esencial para mejorar la calidad de los datos en una organización. Según Redman (1996), un buen programa de calidad de datos debe contener los siguientes aspectos:

1. Visión clara de las metas y objetivos del negocio.
2. Correcta asignación de los responsables de los datos, los cuales deben estar provistos de las herramientas necesarias para cumplir con sus funciones.
3. Poseer un plan operacional de mejoras que establezca cuáles métodos de mejora deben ser utilizados para cada set de datos.
4. Establecer un programa de administración de los recursos.

Para solventar las inconsistencias de los datos es necesario utilizar una metodología de calidad de datos de forma estructurada. Para algunos expertos en la materia, una iniciativa de calidad de datos no puede ser implementada de forma independiente de los procesos que generan los datos, más allá del contexto en el cual estos son utilizados; es decir, uno de los principales aspectos de la metodología de calidad de datos es que debe aplicarse a las fuentes de la información sin importar, por ejemplo, que los datos posteriormente se utilicen en un ámbito diferente al que los generó, como lo sería un proceso de minería de datos.

### ¿Cómo aplicar un proceso de calidad de datos en las organizaciones?

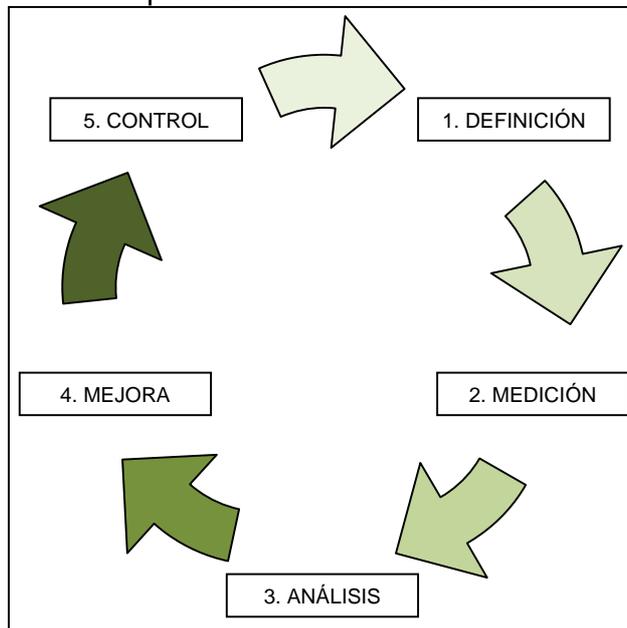
Una vez que la organización ha decidido iniciar un proceso de calidad de datos y se han definido las herramientas que se utilizarán para llevar a cabo el proyecto, es momento de iniciar un proceso cíclico e iterativo que permitirá medir y analizar el nivel de la calidad de los datos dentro de la entidad.

El proceso de calidad de datos se compone de cinco pasos: 1. Definición, 2. Medición, 3. Análisis, 4. Mejora y 5. Control. Cada uno de estos pasos es dependiente uno del otro, por tanto, no se recomienda iniciar con un paso hasta tanto el anterior se encuentre completamente finalizado.

Adicionalmente, el proceso es cíclico, por lo que pueden realizarse todas las iteraciones que sean necesarias hasta alcanzar un nivel aceptable de calidad en los datos que permita cumplir con los objetivos organizacionales.

En la Figura 1 podemos apreciar el ciclo de mejor manera.

**Figura # 1**  
**El proceso de calidad de datos**



Fuente: Autor

A continuación se describe cada uno de los pasos del proceso, así como sus principales componentes:

1. **Definición:** Como la palabra lo indica, en esta fase se definen los componentes más relevantes del proceso de calidad de datos que se llevará a cabo. En este punto se puntualizan los siguientes aspectos:
  - a. Cuadro del proyecto: Se establece el alcance y limitaciones del proyecto.
  - b. Definir el set de datos: Se establece el conjunto de datos que serán sujetos de análisis y en los cuales la mala calidad tiene un impacto negativo y significativo para la organización.
  - c. Definir los interesados: Se identifican todos los participantes del ciclo de información, es decir, quienes generan los datos, quienes los utilizan y quienes se ven afectados de una u otra forma por la poca calidad de la información. Entre estos actores podemos mencionar a los siguientes: dueños de procesos, usuarios de sistemas, administradores, entidades reguladoras y clientes externos.
  - d. Crear el mapa del proceso: Se determinan los procesos del negocio y las aplicaciones que intervienen en la creación o manipulación del set de datos definido y que son utilizados por el grupo de interesados identificado. En este punto se deben definir las bases de datos que serán analizadas, los procesos del negocio y programas que crean, actualizan, transforman y transfieren los datos a otros repositorios o suprimen datos.

2. Medición: En este punto se debe establecer qué tipo de mediciones se realizarán sobre el set de datos. Es importante identificar el grupo de información en el que los costos de las fallas o los errores son altos. Las mediciones deben determinar los aspectos de calidad que diferencian al negocio del resto, así como aquellos que determinan la satisfacción del cliente. Las mediciones deben realizarse en el lugar correcto, es decir, en el almacén de datos que contenga la información que regularmente se utiliza para los procesos del negocio. Adicionalmente, las mediciones deben gestionarse en los momentos adecuados; por lo general este aspecto está relacionado con el factor sorpresa, pues se realizan mediciones sin que los usuarios estén enterados y por consiguiente, en momentos de operación regular. En este punto se deben identificar los parámetros de medición por utilizar, es decir, establecer los componentes más importantes de las dimensiones de calidad de datos requeridas para los sistemas de información que se estén analizando.
  
3. Análisis: En esta fase se analizan los procesos identificados bajo los criterios de medición establecidos. Las siguientes acciones forman parte de esta fase:
  - a. Determinar los procesos y archivos por evaluar: El objetivo de este punto es identificar para cada proceso, el punto en el cual los datos deben ser evaluados, con la finalidad de llevar a cabo una medición eficaz.
  - b. Evaluar la calidad de la información: Su objetivo es medir el grupo de datos contra una o varias dimensiones de calidad, con el fin de determinar su nivel de fiabilidad y así poder establecer el tipo y grado de inconsistencias que se presenten. En este punto son sumamente útiles las funciones de *profiling* de las herramientas de calidad de los datos, las cuales permiten conocer, entre otras cosas, si los datos son conformes de acuerdo con las reglas del negocio, si los valores registrados son válidos y si se encuentran dentro de los límites razonables dentro del contexto del negocio. Como se ha visto, estas herramientas son capaces de validar situaciones como: rangos y dominios de valores, unicidad de llaves primarias, integridad de referencia (relaciones padre-hijo en modelos relacionales), reglas de dependencia y consistencia de formatos.
  - c. Interpretación y reporte sobre información de calidad: El objetivo de este punto es comunicar el estado de calidad de los datos, de manera que se identifiquen los procesos que requieren mejoras basados en el impacto de los defectos que se han detectado. Esta información debe ser consolidada y entregada a la alta gerencia, para que conozca el diagnóstico inicial y pueda tomar decisiones a corto y mediano plazo en relación con la implementación de mejoras integrales.

4. Mejora: En esta fase del proceso se plasman las acciones correctivas necesarias para lograr una mejor calidad de los datos. En este punto se deben realizar las siguientes acciones:
  - a. Estandarización de datos: El objetivo primordial es lograr una estandarización de la información para facilitar el proceso de limpieza y consolidación de los datos. Como se mencionó en una sección anterior, los valores de datos no estandarizados incluyen datos en los cuales dos o más valores diferentes significan lo mismo.
  - b. Corrección y completado de datos: Este paso mejora la calidad de los datos existentes al más alto nivel, corrigiendo valores de datos inexactos y encontrando y capturando valores de datos perdidos para que reúnan los requerimientos de los usuarios. Para llevar a cabo este paso, se deben tomar los datos con valores faltantes, errores y datos sospechosos, en procura de identificar los valores correctos para realizar las correcciones del caso. La corrección de los datos se puede realizar de forma automática o manual, dependiendo del tipo de corrección que se gestione y las posibilidades de gestión de errores que se tengan en la organización. Es sumamente importante en esta fase documentar todos los cambios de datos que se realicen, así como priorizar los tipos de datos sujetos de modificación.
  - c. Consolidación de datos: El objetivo de esta fase es examinar los datos para encontrar registros duplicados, para luego consolidar la información y crear un registro único. Para una mayor efectividad del proceso, la consolidación debe realizarse después de aplicar las rutinas de estandarización y completado de datos. En este punto es vital tener presente el impacto que podría generar una consolidación de datos incorrecta, por lo que resulta muy recomendable mantener una copia de los datos originales antes de la consolidación, como un respaldo de recuperación ante cualquier error en el proceso.
  - d. Análisis de los tipos de error detectados: En este paso se analizan los patrones de errores de datos para mejorar los procesos y eliminar los problemas más significativos causados por las inconsistencias de los datos. En esta fase se debe realizar una clasificación de los errores, estimar su frecuencia y costo para finalmente obtener el impacto del defecto encontrado.
  - e. Transformación y mejora de los datos: En este paso se deben preparar los datos para ser trasladados al repositorio de datos final. Los datos limpios y consolidados deben ser formateados y convertidos para poder mapear la información al destino final.
5. Control: En esta fase se realiza un control de calidad final sobre todo el proceso en general, con el fin de garantizar la correcta aplicación de todos los pasos anteriores. En este punto se incluyen las siguientes tareas:

- a. Auditar y controlar la extracción, transformación y carga de datos: El objetivo de esta tarea es garantizar que los datos correctos son extraídos de las fuentes correctas; que el proceso de transformación de datos se realiza de conformidad con las especificaciones; y que los datos son trasladados a los campos correctos del repositorio destino. Para llevar a cabo esta tarea se pueden establecer controles manuales o automáticos, tomando en cuenta factores como:
  - i. ¿Están contemplados todos los datos que deben extraerse?
  - ii. ¿Los datos son transformados según las especificaciones funcionales?
  - iii. ¿Se encuentran identificados los posibles errores y las excepciones en los procesos?
  - iv. ¿Los procesos de extracción, transformación y carga se ejecutaron correctamente?
  - v. ¿Todas las entradas de datos generaron su correspondiente salida?
- b. Desarrollar un plan de mejoramiento: Busca identificar las causas raíz de un problema y establecer las acciones correctivas para eliminarlas o minimizarlas.
- c. Presentación del informe final: En este punto se debe realizar un informe sobre el proceso en general, que muestre el estado inicial de la calidad de los datos, los hallazgos detectados durante el proceso de *profiling*, estadísticas de corrección de datos, detalle de los procesos que fueron afectados, el nivel de calidad de los datos alcanzado y las expectativas a futuro sobre el proceso de calidad a nivel organizacional.

Para que cada uno de los pasos anteriores tenga éxito es vital que sea aplicado por expertos en la materia, y que se lleve un estricto control documental sobre cada una de las fases mediante elementos como formularios, listas de chequeo, reportes de procesos y matrices. Igualmente, es requerido mantener un control documental sobre la información capturada mediante cuestionarios, catálogos de datos, diccionarios de datos, matrices de aplicaciones, matrices de procesos, organigramas, gráficos o cualquier otro que haya sido utilizado como insumo en el proceso de calidad de datos. También es importante resaltar que este proceso es independiente de las herramientas de análisis de datos que pueda utilizar la organización.

Igualmente, el proceso se puede aplicar sin la necesidad de contar con una herramienta de calidad. No obstante, es claro que el tiempo por invertir para completar las fases de medición, análisis y mejora será mucho mayor de no utilizarse software especializado, con lo cual los resultados de una iniciativa de calidad no serían tan inmediatos como se esperaría, lo que podría convertirse en un inconveniente para la alta gerencia.

## **Beneficios de los procesos de calidad de datos**

Los reportes y análisis de los clientes, el mercado, los productos o cualquier otro factor similar posibilitan una ventaja competitiva en el mundo de los negocios; sin embargo, esta información es útil para la toma de decisiones siempre y cuando haya sido generada con datos correctos. El riesgo de que la información contenida en estos reportes no sea de calidad genera un dilema para quienes deben interpretar los resultados.

La integración y el intercambio de información entre organizaciones resultan productivos únicamente si los datos que se intersecan entre los sistemas cumplen con ciertos estándares de calidad. Alguna vez se ha preguntado si las empresas estarían dispuestas a cargar en sus sistemas información incorrecta de clientes, productos o transacciones generadas en otros sistemas; la respuesta es fácil: NO. Tanto la empresa que genera la información como la que la recibe deben garantizar que los datos son de calidad para poder distribuirse o recibirse.

De esta forma, el principio fundamental para el intercambio de información se podría resumir en algo así: antes de aceptar los datos para utilizarlos en un proceso de toma de decisión, lo almacene en un sistema de la organización o lo transfiera a un tercero, demuestre que la información es de calidad y que sirve para los propósitos establecidos. Esto es sumamente difícil de alcanzar, ya que aún en industrias que se rigen por regulaciones estrictas como es el caso de las finanzas, es conocido que mientras sea necesario obtener información de los clientes, las compañías o las transacciones que se realizan, no existe garantía de que esta información sea de suficiente calidad como para cumplir con todos los objetivos propuestos.

### *E-business*

La tecnología ha dado nacimiento al comercio electrónico. Todos los productos que ingresen al mercado en el futuro cercano tendrán una relación con el comercio electrónico (en inglés, *e-commerce*) o los negocios electrónicos (en inglés, *e-business*). El *e-business* posibilita a las compañías ofrecer una gran cantidad de productos y servicios a costos sumamente bajos.

En la era digital de los negocios, la rapidez y la eficiencia son conceptos críticos para el éxito. Con el comercio electrónico, los clientes han aumentado sus expectativas sobre los servicios. Es más, la forma de hacer negocios ha cambiado como resultado del *e-business*. Por ejemplo, los clientes desean realizar sus transacciones en sitios web de forma rápida y simple.

El impacto de la tecnología en los negocios es enorme. La tecnología, usando Internet, posibilita que aun el negocio más pequeño pueda interactuar con clientes alrededor del mundo. La Internet permite que las empresas compitan a nivel local, regional, nacional y mundial, por lo que se hace necesario identificar claramente

los clientes y dirigir los servicios y productos de acuerdo con las necesidades específicas de estos.

Unificar la información facilita separar los nuevos clientes de los ya existentes. Esta separación permitirá adaptar el contenido de la web a la medida del cliente y realizar campañas publicitarias de acuerdo con sus preferencias. Un ejemplo que puede ayudar a entender este concepto es el siguiente: imagine que es el dueño de una tienda virtual de música en la que pone a disposición de los usuarios música de todos los géneros. Un cliente ingresa y da sus datos personales y realiza una compra de un CD de música clásica. Luego, en un par de visitas posteriores, realiza otra compra de música del mismo género (el cliente ya fue identificado cuando proporcionó información de la primera compra). Con base en la información de las compras realizadas, se crea un "perfil" para el cliente y la próxima vez que ingrese al sitio a buscar información lo primero que visualizará será información sobre los nuevos CD del estilo de música que se "intuye" le gusta, es decir, el género de la música clásica, no música de otro tipo. Igualmente, podríamos aplicar una publicidad dirigida mostrando las promociones en música de este género no solo a este usuario, sino al grupo específico de clientes que cumplan con ciertas características, datos que se tienen registrados gracias a la información que se ha obtenido del sitio web.

Es claro entonces que, si se filtra información incorrecta a través de los puntos de captura de datos, estos se trasladarán a los sistemas empresariales y a los de toma de decisiones, lo que le generará un problema a la organización.

Anteriormente, el éxito de un negocio de estas características se medía por la cantidad de *hits* o visitas al sitio, así como la rapidez para atraer nuevos clientes. No obstante, conforme ha transcurrido el tiempo, los expertos han empezado a modificar el concepto anterior y a concentrarse más en mantener a los clientes que se tienen en vez de preocuparse por la cantidad de tráfico que pase por el sitio, esto debido a la gran variedad de servicios que se encuentran en la Internet.

La precaución en este sentido es vital, ya que el *e-business* también conlleva un riesgo. Los pedidos por Internet son propensos a errores si no existe una verificación de la información referente a destinatarios, por lo que los datos del cliente se convierten en un aspecto relevante de la relación empresa-proveedor. Información inexacta puede llevar a perder clientes por insatisfacción. Las herramientas de calidad de datos pueden ayudar a prevenir esta situación en tiempo real.

A largo plazo, garantizar la calidad en la información de los clientes le permite a la organización maximizar los índices de recuperación de inversión mediante procesos de Sistemas de Planificación de Recursos<sup>4</sup> (en inglés ERP, *Enterprise*

---

<sup>4</sup> Según Wikipedia (2007), los ERP "son sistemas de gestión de información que integran y automatizan muchas de las prácticas de negocio asociadas con los aspectos operativos o productivos de una empresa".

*Resource Planning*), Administración de la Relación con los Clientes<sup>5</sup> (en inglés CRM, *Customer Relationship Management*), Inteligencia Empresarial<sup>6</sup> (en inglés BI, *Business Intelligence*) y Almacenes de Datos<sup>7</sup> (en inglés, *Data Warehouse*).

Conocer al cliente es la clave. Este conocimiento permite estrechar los vínculos con el cliente, con lo cual, a pesar de que otros ofrezcan los mismos productos y servicios que usted ofrece por el mismo precio, los clientes se mantendrán fieles en cuanto se personalice su interacción con la empresa y se mantengan aspectos como la facilidad de acceso, el tiempo de respuesta, la exactitud, veracidad e integridad.

Desde esta perspectiva resulta fácil entender que los procesos de calidad de datos se convierten en una herramienta para aumentar la competitividad en la economía global. Un plan estratégico para garantizar la calidad de la información de los clientes contempla que los procesos de captura de la información sean sumamente robustos para que, una vez cargada en los sistemas operacionales, los datos puedan ponerse a disposición de toda la organización, lo que genera un beneficio para los procesos de ERP, CRM, *Data Warehouse* y BI.

La información de los clientes es el activo más valioso de las organizaciones. Convertir la calidad de los datos de prioridad garantiza que los esfuerzos e inversiones realizados en los datos rendirán frutos a corto y largo plazo.

Si se conocen los beneficios que tiene emprender un proceso de calidad de datos en nuestras empresas u organizaciones, ¿por qué resulta tan difícil llevarla a cabo? Una iniciativa de calidad de datos, como cualquier otro proyecto, requiere recursos, en este caso, tanto financieros como humanos. Requiere que existan responsables de realizar los análisis, tomar decisiones sobre los problemas detectados, generar las correcciones del caso y realizar un seguimiento oportuno de la gestión en aras de mantener los estándares de calidad para la nueva información que se almacene en los sistemas. Es responsabilidad de los profesionales en tecnologías de información concienciar a la alta gerencia sobre la importancia que tienen para la organización los datos y gestionar lo necesario para garantizar su calidad.

---

<sup>5</sup> Según Wikilepdia (2007), CRM “es un modelo de gestión para toda la organización, basada en la orientación al cliente”.

<sup>6</sup> Según Wikipedia (2007), BI “es el conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa”.

<sup>7</sup> Según Wikipedia (2007), un *Data Warehouse*, en el contexto de la Informática, “es una colección de datos orientadas a un dominio, integrado, no volátil y variable en el tiempo que ayuda a la toma de decisiones de la empresa u organización”.

## Determinando el grado de recuperación de inversión en proyectos de calidad de los datos

Al igual que con cualquier otro tipo de proyecto que se emprenda en la empresa u organización, todo depende del beneficio que este le genere al negocio y en este punto, uno de los factores que se deben tomar en cuenta es el factor de recuperación de la inversión, es decir, cómo y en cuánto tiempo recuperará la organización los recursos invertidos en el proyecto. Determinar el valor de recuperación de la inversión por una iniciativa de calidad de datos es sumamente difícil. En alguna medida, una mínima parte puede ser estimada midiendo el impacto que tiene actualmente la mala calidad de los datos en los sistemas organizacionales; sin embargo, sabiendo que los datos son utilizados por varios participantes tanto dentro como fuera de la organización, basar los cálculos únicamente a nivel del sistema podría subestimar el verdadero valor de una iniciativa de calidad de datos.

Según Wong (2007), para que una organización pueda estimar el valor total de recuperación por un proceso de calidad de datos de forma integral, deben contestarse dos preguntas clave:

1. ¿Cuán valiosos son los datos para la empresa?, asumiendo que la información almacenada tiene una calidad del 100%.
2. ¿Cuánto decrece ese valor si la calidad disminuye?

Mientras estas preguntas parecerían incontestables para quienes se desenvuelven en el campo de la Tecnología de la Información, en el mundo de las finanzas son comunes aspectos relacionados con el valor y la depreciación de los bienes, como las acciones o las propiedades.

Desafortunadamente, los datos son uno de esos bienes intangibles en que los métodos regulares de asignación de valores como es el costo, son inservibles. No obstante, existen otros conceptos intangibles como las relaciones cliente-proveedor, para las cuales existen técnicas que permiten asignarles un valor.

Lo anterior demuestra que, por más extraño que parezca, es posible que en un futuro se establezcan metodologías que permitan asignar un valor real a los datos de las empresas y con ello lograr calcular de manera más exacta, el valor de una iniciativa de calidad de datos.

## Conclusiones

En la actualidad, la información de las compañías puede provenir de diferentes fuentes, entre ellas, viejos sistemas operacionales, bases de datos propietarias, proveedores externos y la misma Internet. Como resultado de esto, es muy probable que existan diferencias de formato entre los orígenes, lo que compromete la calidad y la integración de los datos, factores que pueden poner en riesgo la organización.

Por ende, los procesos de calidad de datos son necesarios para determinar el grado de calidad de la información en los sistemas y poder gestionar las acciones correctivas requeridas, ya sea a nivel de los datos en sí, los procesos, los sistemas, repositorios de datos u otros, en aras de mantener un estándar de calidad de la información. La calidad de los datos definida desde ese espectro de conformación y utilidad es un concepto jerárquico y multidimensional. Las iniciativas de calidad de datos en las organizaciones no deben ser implementadas lejos de los procesos que generan los datos ni del contexto en el que son utilizados.

Las herramientas de calidad de datos facilitan los análisis de grandes volúmenes de información a partir del concepto de dimensiones definido por los expertos; sin embargo, para aplicar un proceso de calidad de datos es recomendable utilizar un marco procedimental compuesto por cinco fases dependientes e iterativas. Este procedimiento es totalmente independiente de las herramientas de calidad que se utilicen y puede ser adaptado a prácticamente cualquier proceso.

Es sumamente importante que en los procesos de calidad de datos se incorporen tanto aspectos técnicos como funcionales y operativos, se involucre de lleno a la alta gerencia, se dé un compromiso por parte de los usuarios involucrados y se establezcan claramente los alcances del proyecto para de esta forma poder obtener resultados óptimos que permitan completar otros procesos de calidad dentro de la organización. La calidad de datos no debe considerarse como un proceso independiente de la operativa regular, por el contrario, debe buscarse la integración de las políticas y rutinas de calidad dentro del marco regular de aplicación, con el objetivo de mantener los estándares de calidad en el tiempo y evitar que las inconsistencias se propaguen.

## Bibliografía

- Abate, M., Diegert, K. y Allen, H. (1998). A Hierarchical Approach to Improving Data Quality. *Data Quality, setiembre de 1998, volumen 4, número 1*. Recuperado el 14 de julio de 2007, de <http://www.dataquality.com/998abate.htm#1>
- Crosby, Philip. (1980). *Quality is free: The art of making quality certain*. Nueva York: Mentor.
- Fisher, T. y Marinos, G. (2003). *Better Decisions Through Better Data Quality Management*. PricewaterhouseCoopers
- University of Minnesota. (1997) *GIS Glossary*. Recuperado el 14 de julio de 2007, de <http://fwcb.cfans.umn.edu/courses/FW5620/glossary.htm>
- Government of British Columbia. (2001). *BC Government Information Resource Management Glossary*. Recuperado el 14 de julio de 2007, de [http://www.cio.gov.bc.ca/other/daf/IRM\\_Glossary.htm](http://www.cio.gov.bc.ca/other/daf/IRM_Glossary.htm)
- Juran, J. M. (1989). *Juran on leadership for quality: An executive handbook*. Nueva York: Free Press.
- Kahn, B., Strong, D. y Wang, R. (2002). Information Quality Benchmarks: Product and Service Performance. *Communications of the ACM, April 2002*. pp. 184-192.
- National Center for Environmental Research and Quality Assurance. (1997). *Glossary of Quality Assurance Terms And Related Acronyms Quality Assurance Division*.
- Redman, T.C. (1996). *Data Quality for the Information Age*. Norwood, MA: Artech House.
- Tayi, G. y Ballou, D. P. (1998). Examining Data Quality. *Communications of the ACM, 41(2)*, pp. 54 - 57.
- Wang, R., Strong, D. y Guarascio, L. (1994). *An empirical investigation of data quality dimensions: A data consumer's perspective*. *Total Data Quality Management (TDQM) Research Program, MIT Sloan School of Management*.
- Wong, E. (2007). *Valuing Data Quality*. Recuperado el 14 de julio de 2007, de [http://blogs.informatica.com/announcementstest/2007/02/valuing\\_data\\_quality\\_1.html](http://blogs.informatica.com/announcementstest/2007/02/valuing_data_quality_1.html)

Wikipedia. (2007). *Almacén de Datos*. Recuperado el 14 de julio de 2007, de [http://es.wikipedia.org/wiki/Almac%C3%A9n\\_de\\_datos](http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos)

Wikipedia. (2007). *CRM*. Recuperado el 14 de julio de 2007, de <http://es.wikipedia.org/wiki/CRM>

Wikipedia. (2007). *Inteligencia Empresarial*. Recuperado el 14 de julio de 2007, de [http://es.wikipedia.org/wiki/Inteligencia\\_empresarial](http://es.wikipedia.org/wiki/Inteligencia_empresarial)

Wikipedia. (2007) *Organización Internacional para la Estandarización*. Recuperado el 14 de julio de 2007, de [http://es.wikipedia.org/wiki/Organizaci%C3%B3n\\_Internacional\\_para\\_la\\_Estandarizaci%C3%B3n](http://es.wikipedia.org/wiki/Organizaci%C3%B3n_Internacional_para_la_Estandarizaci%C3%B3n)

Wikipedia. (2007). *Planificación de Recursos Empresariales*. Recuperado el 14 de julio de 2007, de [http://es.wikipedia.org/wiki/Planificaci%C3%B3n\\_de\\_recursos\\_empresariales](http://es.wikipedia.org/wiki/Planificaci%C3%B3n_de_recursos_empresariales)