
Perspectiva epistemológica de la calidad de datos

Dr. Silverio Bustos Díaz^{1,2}

¹Facultad de Ingeniería de Sistemas e Informática
Universidad Nacional Mayor de San Marcos

² Facultad de Ingeniería
Universidad Ricardo Palma

sbustosd@unmsm.edu.pe

RESUMEN

El artículo hace una presentación de la calidad de datos en el ciclo de vida de los sistemas de información, haciendo énfasis en la importancia que tienen los datos para la operación de los sistemas de información y durante esta. Por ello se describen, analizan y presentan las perspectivas de la calidad de datos en términos de los esfuerzos de la calidad de datos, la evolución de los datos, los sistemas de información heredados, la implantación de los sistemas de información, los problemas de la calidad de datos, las dimensiones de la calidad de datos, el desgaste de los sistemas de información, el control de la calidad de datos, los procesos inherentes a la calidad de datos y la gestión de datos maestros. Finalmente, se concluye que las organizaciones mejorarán la credibilidad en sus datos mientras sean conscientes del potencial problema de calidad de sus datos y la posibilidad de una solución oportuna; además, se recomienda desarrollar un Plan de Acción desde las fuentes hasta los resultados de información pasando por los diversos procesos inherentes a su evolución.

Palabras clave: Calidad de datos, aseguramiento de calidad de datos, procesos del negocio, sistemas de información, inteligencia de negocios, base de datos

ABSTRACT

The article gives a presentation of the data quality in the life cycle of information systems, emphasizing the importance of data to and during the operation of information systems. As described, analyzed and presented the perspectives of data quality in terms of the efforts of the quality of data, data evolution, legacy information systems, information systems implementation, data quality problems, data quality the dimensions, the wear of information systems, data quality control, the processes inherent in the quality of data, and master data management.

Finally, the organizations will improve the credibility of their data while aware of the potential problem of data quality and the possibility of a timely resolution; also recommends developing an "Action Plan" from sources to the results of information through the various processes involved in its evolution.

Key words: data quality, data quality assurance, business process, information systems, business intelligence, database, MDM, ETL, OLAP.

1. INTRODUCCIÓN

La proliferación y aumento del volumen de las fuentes de información hacen más imperativos la integración y calidad de los datos. Esta comprende los procesos siguientes:

- La recuperación de datos integrados para el análisis multidisciplinario en la organización;
- La eliminación de las redundancias, incoherencias y errores en los datos; y
- La limpieza de datos y la transferencia en los nuevos sistemas de información.

Hoy, con el advenimiento y uso cada vez más intenso de las Tecnologías Informáticas (TI) en las organizaciones, la calidad de datos tiene más importancia por su trascendencia estratégica en la toma de decisiones; sin embargo, es notorio el alto porcentaje de ejecutivos que dudan de los informes, consultas o reportes que se obtienen de los sistemas de información (SI). Por ello, muchos ejecutivos fundamentan sus decisiones en el sentido común, por la falta de credibilidad de la información que se les proporciona.

Los ejecutivos dudan de la información proporcionada por los SI, en particular de su capacidad para reflejar el mundo real de la organización en un momento determinado; esa falta de confianza en los datos es bastante común porque dificulta la gestión de procesos por no disponer de información confiable y oportuna para las reuniones estratégicas de la organización –entre otros– para:

- La planificación financiera y del marketing de clientes;
- La dirección estratégica, incluyendo los recursos humanos y materiales;
- La gestión de la producción (productividad y efectividad) y de costos (directos e indirectos) en los proyectos (departamentos); etc.

De esta manera, hay un problema de credibilidad de la información en términos de valor y disponibilidad del dato, debiendo prestarse atención a la integración de los procesos del negocio que pueden ser: automáticos, semiautomáticos o manuales.

Esto es explicable porque en el mundo real, el modelo de procesos *“puede ser descrito mediante un modelo dinámico. Estos modelos muestran cómo se comporta el sistema en el transcurso del tiempo, mediante la ejecución de un proceso de negocio o como resultado de*

los acontecimientos que se producen dentro del ámbito de la solución” [4]; entretanto, los SI representan las actividades mediante el procesamiento que comprende etapas tales como: la creación de datos, recopilación o importación de datos, almacenamiento de datos, integración de datos y análisis de datos.

Al respecto, indistintamente a las etapas mencionadas, es importante sintetizarlas en tres funciones básicas a considerar en la calidad de datos, éstas son: recuperación, producción y custodia; tanto para los usuarios como para los datos propiamente dichos. Al respecto, debemos mejorar la gestión de la información y ejecutar los tres pasos siguientes:

- a. Orientar a tener una sola fuente, si existen diversas fuentes siempre se va dedicar inútilmente recursos para que la información cuadre, o un esfuerzo adicional para mantener actualizados los SI, tanto los formales como los informales, es conveniente tener una sola fuente de información, con buena o mala información, pero que al menos sea la misma para todos;
- b. Asegurar que la fuente de información sea única, correcta y refleje la realidad de lo que está pasando en la organización, esto comprende el sistema formal de información tenga la información correcta, expresado de otra forma, que la información de nuestro sistema; y
- c. Lograr esto no es nada fácil, ya que el personal no siempre percibe como parte de su trabajo mantener una información confiable. Por ello es frecuente oír decir a los usuarios: “a mí me pagan por acomodar, cuidar y surtir los productos, no por procesar (capturar, actualizar, etc.) la información”.

2. ¿QUÉ ES LA CALIDAD DE DATOS? [1]

En este contexto surge el concepto de calidad de datos como la acción para mejorar el contenido de los datos; en búsqueda de una representación pertinente (oportuna, exacta y confiable) de la información para cualquier forma de procesamiento. Al respecto, surge la necesidad de lograr un aseguramiento de la calidad de datos (*DQA = Data Quality Assurance*), que se convierte en un problema común y permanente de cualquier SI.

Como cualquier problema, si éste es atacado y resuelto oportunamente, es posible obtener ventajas competitivas importantes y se mitigará una situación no desea-

da, especialmente por uso de información indebida, incorrecta o inoportuna.

La falta de calidad de datos se encuentra directamente asociado desde la recuperación de información (RI) que se disponga a través del software de aplicación (programas); en este contexto los objetivos generales de la RI están basados en:

- Que la información (contenida) de las fuentes de información se almacena en campos estructurados de una base de datos (BD).
- Que la creación de la BD corresponde a la recuperación de la información desde fuentes heterogéneas contenidas en:
 - Formatos manuales,
 - Bases de datos de usuario final, o
 - Información digital en archivos planos o tablas de BD de otros SI.
- Que los datos a ser ingresados a la BD deben ser racionalizados y solo los necesarios para poder obtener la información necesaria.

2.1. Esfuerzos para la calidad de datos [1]

Desde la aparición de la TI en las organizaciones, se tiene especial interés para mejorar la calidad de datos; así podemos mencionar los innumerables estudios sobre los errores en la programación y en los procesos de carga de datos, que han permitido desarrollar métodos y técnicas para mitigar su impacto en la interpretación de valores, disponiéndose de herramientas que ayudan en la validación de la semántica y de la sintaxis; sin embargo, es difícil eliminar posibles inconsistencias en la etapa de registro (carga), tales como los errores de deletreo (“*spelling errors*”) en donde se puede omitir o intercambiar la secuencia de caracteres alfanuméricos y que obligan a re-procesos.

Cualquier proyecto de TI está orientado a optimizar los procesos (estratégicos, tácticos y operativos) de la organización mediante una reducción del esfuerzo y una optimización en el uso de sus recursos; éste debe ser el criterio fundamental para la implantación y operación del SI.

Por esta razón debemos reafirmar que la calidad de datos está directamente relacionada con el funcionamiento de la organización porque a través de los datos se darán los servicios de información necesarios para la ejecución de sus procesos. Evidentemente, la organización no es ajena a esta afirmación porque “*malos*

datos conducen a malas decisiones”, y pueden afectar o conducir a:

- Objetivos incumplidos de la organización;
- Pérdidas de oportunidades de negocio;
- Estrategias inadecuadas (no óptimas) de la organización.

Debemos entender que una mala decisión conduce a acciones inaceptables (altos costos, impacto en la seguridad o en la gestión de la organización, pérdida de imagen, etc.), en tanto que una buena decisión conduce a acciones aceptables que generan rentabilidad y confianza en la organización, reduciendo costos como consecuencia de una disminución del esfuerzo y la racionalización en el uso de recursos. Asimismo, una mala decisión trae como consecuencia un debilitamiento en la organización, especialmente en su funcionamiento; por ejemplo, podría impactar directamente –entre otras– en:

- el aumento de costo en la operación (producción) del servicio (producto);
- la oportunidad en la entrega del servicio (producto), afectando la imagen;
- el incremento de costos de operación de los SI's (manuales o automatizados).

2.2. La evolución de los datos

Los datos requeridos por la organización deberán ser completos y válidos para sus necesidades de información, por lo que debe tomarse en cuenta los procesos inherentes a su evolución, en el mismo SI como en su integración con otros SI incluyendo la BI (Business Intelligence, Inteligencia de negocios); esto es, los procesos de actualización de datos [1]:

- *Entrada de los datos*: comprende carga de datos, archivos y tablas por ingresar, complementados con controles (cruzados, semánticos y heurísticos) basados en patrones y que debe evaluarse el momento de su aplicación, por ejemplo, la reducción de consistencias (controles) en las aplicaciones Web;
- *Gestión de los datos*: en donde las “reglas del negocio” no siempre cubren la diversidad de situaciones reales que se presentan, generando problemas semánticos de coherencia e integridad; y
- *Salida de los datos*: pantallas, reportes, archivos o tablas a extraer, etc., en donde cualquier validación (control) está sujeta a la subjetividad del personal

que lo realiza, en donde se busca transcribir la información almacenada en la BD; los problemas se detectan al momento del uso de la información, así por ejemplo, en las interfaces con otros SI.

Los datos, como recursos de la organización, están expuestos a diversas situaciones, entre las que podemos citar:

- a. Corrupción de datos, en cualquier momento de su procesamiento, puede ser originada por una manipulación indebida o por un acto delictivo en su procesamiento, lo que puede ocurrir en cualquiera de los siguientes momentos:
 - Ingreso (carga o recuperación) de datos,
 - Migración de datos,
 - Integración de datos.
- b. Limpieza y transformación durante la conversión. Generalmente durante la creación o mantenimiento de la BD se debe realizar una serie de conversión de valores, generando situaciones que deben ser depuradas. De esta manera identificamos los siguientes momentos:
 - Identificar anomalías; eliminar inconsistencias y detectar datos incorrectos;
 - Desarrollar una estrategia y metodología para mantener una forma de detección y limpieza de datos;
 - Método de procesamiento:
 - Usar software para acceder y manipular los datos de orígenes diversos para su transferencia a múltiples plataformas;
 - El software debe ser personalizado para la detección automática de anomalías, inconsistencias, limpiar y transformar los datos de acuerdo a reglas del negocio preestablecidas por la organización.

Esta situación contribuye al riesgo de obsolescencia de los datos ya sea por la ocurrencia de los riesgos expuestos como por la integración de los SI. Estas deben ser tomadas en cuenta para los procesos de BI sobre DW (Data Warehousing, Almacén de Datos) para que no se afecte la toma de decisiones de la organización.

2.3. Los sistemas de información heredados (“*Legacy information systems*”)

La convivencia del SI con los sistemas heredados tiene como principal dificultad la vigencia de la información

almacenada en la BD de SI heredados con la BD; es así que es posible establecer dos momentos: [1]

- a. Futuro (SI nuevo): en donde se establecerán los controles automáticos acordes con las necesidades actuales de la organización, pudiendo introducir reglas precisas;
- b. Antiguo y Presente (SI heredado): en donde los datos que contienen la información histórica han debido evolucionar a diversas situaciones, entre las que podemos considerar:
 - El marco jurídico del momento de la acción, en base a las normas legales que han tenido su validez jurídica, y que han ido evolucionando progresivamente;
 - Formatos de manual o pantalla, acordes a cada disposición legal, no necesariamente con los mismos datos.
 - Uso de formatos antiguos (anteriores) luego de entrar en vigencia los nuevos dispositivos legales con otros formatos.
 - Falta de uniformidad en el uso de formatos y pantallas como consecuencia de hábitos (costumbres) de los responsables, a pesar de la posible existencia de algún “Manual de Llenado”, que genera un apreciable número de registros con criterios dispares (falta de unicidad de criterios en el llenado del formato).
 - Capacitación insuficiente de los responsables, especialmente cuando ha existido alta rotación de personal.

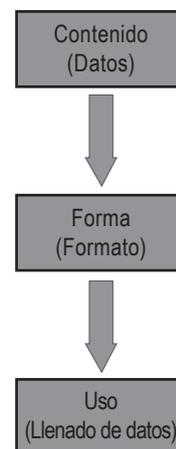


Figura N.º 1. Documentos fuente para la recuperación de información.

Por lo tanto, los documentos fuente para la RI tienen las siguientes características generales (Ver Figura N.º 1):

- A. *Contenido* (Datos): se cuenta con la información mínima para cada necesidad de información, en concordancia con el marco de la organización y legal vigente.
- B. *Forma* (Formato): se cuentan con diversos formatos, acordes con el marco jurídico del momento. La información contenida en los formatos no siempre es la misma, y no necesariamente se cuenta con toda la información exigida en la ley vigente.
- C. *Uso* (Llenado de datos): los responsables no han contado o podrían haber obviado los instructivos específicos para el llenado de los formatos.

Al respecto cada organización desarrolla factores sociológicos que conducen a problemas de interpretación e integración de datos, por lo que se puede afirmar que **cada organización crea su propia cultura o clima, con sus propios tabúes, costumbres y usos**. Esta cultura refleja tanto las normas y valores formales e informales en donde las modalidades de comunicación pueden haber sido inefectivas e ineficientes, y que se reflejan en la fuente de información.

2.4. La implantación de los sistemas de información

Dependiendo del nivel de informatización de la organización, el SI a implantar puede ser nuevo o heredado; en ambos es imprescindible el conocimiento de las reglas del negocio (antiguas y vigentes) [2]:

- Para un SI heredado, hay que poner énfasis en el pasado sobre:
 - Las estrategias y formas de ejecución desde la carga inicial (migración de datos para los sistemas de información heredados).
 - Las premisas usadas por la organización y encontrar formas de equivalencias con las vigentes.
- Para un SI nuevo se considerarán las premisas básicas acordes con el presente (misión) y futuro (visión) de la organización.

De cualquier manera, la implantación del SI se inicia con la carga inicial de datos mediante procesos manuales o automatizados –normalmente bien definidos– en donde se requiere precisar la identificación, definición y ejecución de procesos estandarizados en el manejo de los datos; es decir, al iniciar la implantación y durante la operación se debe contar con reglas estandarizadas y bien definidas en el uso de los datos:

- en el proceso de carga inicial y los subsiguientes procesos de actualización;
- la información histórica, normalmente en la BD o en archivos planos;
- el procesamiento de datos, que incluye las interfaces internas y externas con otros SI's; y
- la obtención o emisión de resultados, entre otros, para la BI desde el DW.

Adicionalmente, puede ser necesario aplicar una estrategia progresiva de implantación en donde exista un período de convivencia entre el pasado y lo nuevo; esta estrategia puede llevarnos a situaciones con diferentes grados de dificultad, no necesariamente excluyentes, y que pueden combinarse para la implantación progresiva de:

1. las funcionalidades del SI, en términos de módulos o subsistemas;
2. la operatividad del SI, vinculado a la infraestructura de hardware y comunicaciones.

2.5. Los problemas de la calidad de datos

Hay una diversidad de problemas de la calidad de datos, a los que podemos agruparlos de la siguiente manera:

- a. Procesos manuales: con datos no necesariamente estandarizados y con riesgos de interpretación y transcripción;
- b. Procesos automatizados: con datos inconsistentes por la desincronización en el procesamiento de los SI operacionales o SI legados.
- c. Errores en el contenido o valor de los datos: faltantes, incorrectos u obsoletos;
- d. Incoherencia de los datos: diferentes formatos para los mismos datos, diferentes interpretaciones para el mismo valor de código (antónimos), varios valores de código que tienen un mismo sentido (sinónimos), uso excesivo de campos para fines no previstos;
- e. Problemas de organización y estructuración de datos: que pueden afectar el tiempo de respuesta en los SI; y
- f. Procesos de migración: que generan datos inválidos por problemas de estandarización que genera incoherencia de datos, datos duplicados, datos perdidos, etc.

Normalmente, los problemas de calidad de datos se propagan muchas veces de manera silenciosa, generando dificultades en el uso adecuado de la información de los SI en perjuicio de la organización.

2.6. Las dimensiones de la calidad de datos

La calidad –en general– tiene un costo intrínseco, y la calidad de datos no es ajena a ella, de esta manera se dice que *“la falta de calidad de los datos es una cuestión que está completamente asumida como el precio que se paga por disponer de la Tecnología de Información”*. Al respecto, es importante hacer referencia a las 6 dimensiones que establece el Sistema Estadístico Europeo [3] sobre la calidad de los datos, ellas son [2]:

1. *Relevancia o pertinencia*: Que la información le sirva para su fin.
2. *Exactitud*: Que la información refleje lo que se desea.
3. *Oportunidad y puntualidad*: Que la información esté disponible cuando se requiere para tomar una decisión.
4. *Accesibilidad y claridad*: Que corresponde a la facilidad de obtener los datos.
5. *Comparabilidad o nivel de detalle*: Que la información tenga el nivel de detalle requerido para su buen entendimiento.
6. *Consistencia o coherencia*: Que sea fiable de diferentes maneras y para distintos usos, incluyendo la totalidad de lo requerido.

Estas seis dimensiones de la calidad de datos, no siempre coexisten armoniosamente, a menudo hay un equilibrio entre las diferentes dimensiones, de modo que las decisiones basadas en el máximo de la calidad de los datos pueden favorecer una sobre otra dimensión. Existen propuestas que definen otras dimensiones, las mismas que no son tan relevantes para la generalidad de casos; entre ellas podemos citar las propuestas de Diane Strong (1997), Leo Pipino (2002), y Cinzia Capiello (2003) [7].

2.7. El control de la calidad de datos

El control de la calidad de datos involucra tres actividades:

- **Aseguramiento de la calidad**: Actividades para asegurar la calidad de los datos antes de su recuperación. El aseguramiento comprende una estandarización del proceso y la elaboración de manuales de operación del SI.

- **Manejo de datos**: Manejo y proceso de datos en todo el proceso, comprende el diseño del proceso de recuperación, manipulación y presentación de los datos.
- **Control de calidad**: Monitoreo y mantenimiento de la calidad de los datos durante el proceso, comprende el seguimiento y protección de los datos durante toda su manipulación.

De esta manera, [7] se afirma que *“la calidad de los datos y de la información viene determinada tanto por la calidad de la BD como de la calidad de la presentación de los datos”*. Por lo tanto, es necesario construir un sistema integrado de datos de calidad a lo largo de toda la vida de un SI y garantizar el rastreo de la calidad en todo momento. Al respecto, Lee & Strong [5] plantean *“un modelo de datos de calidad multicapas de datos de calidad para apoyar el control de la calidad de los productos de software en el ciclo de vida y en curso de su evolución; dicho modelo consta de seis capas: inventario general (GBOM = General Bill Of Material), el soporte de datos de calidad, los procesos de calidad, las actividades de calidad, los objetos de calidad y los datos físicos”*.

2.8. El desgaste de los sistemas de información

Los SI como cualquier ente orgánico sufre un “desgaste natural” producto de su “envejecimiento vegetativo” en el tiempo porque el valor del SI se reduce progresivamente. Esta situación genera la necesidad de reemplazar el SI antiguo; sin embargo, mediante análisis detallados de proyectos se pueden observar lo siguiente:

- Que un buen número de requisitos para el SI nuevo son necesidades ya atendidas por el SI actual, quiere decir que había un desgaste del SI actual.
- Que hay funciones disponibles en el SI actual, nunca se habían implementado.
- Que muchas funciones requeridas en el SI nuevo lo tiene el SI actual, lo significa que se ha tomado una ruta de reemplazo en lugar de trabajar con su SI actual, que en muchos de los casos ha resultado ser más cara y difícil.

El desgaste de los SI se debe a muchos factores, y casi todos son importantes; pero, la realidad en el ciclo de vida de los SI afecta las dimensiones de la calidad de datos, y puede ser esquematizado (véase Tabla N.º 1):

Tabla N.º 1. Esquema de causa y efecto que influyen en la calidad de datos.

Causas	Efectos
Los usuarios olvidan las funcionalidades que no usan encontrando sus propias secuencias de procesos, obviando tácitamente aquellos que “no son relevantes” para sus actividades cotidianas y por lo tanto no validan los datos que podrían ser relevante para otros usuarios.	Relevancia y accesibilidad
La rotación de personal genera que los usuarios antiguos capaciten a los usuarios nuevos, quienes a pesar de sus mejores intenciones no logran transmitir el 100% de sus conocimientos; por lo tanto, la capacitación se va degradando progresivamente.	Comparabilidad y consistencia
La organización evoluciona, por lo general, lentamente a tal punto que el plan de mantenimiento no refleja el 100% de los cambios; en cuyo caso los usuarios optan por usar sus propias soluciones que se llevan a cabo con herramientas de ofimática o de forma manual.	Exactitud y oportunidad

2.9. Los procesos inherentes a la calidad de datos

Para aumentar la credibilidad o confiabilidad de sus datos, las organizaciones deben recurrir a actividades permanentes y eventuales, en donde se combina [2]:

- *El mejoramiento de la producción*, en base a procesos permanentes de corrección, validación y normalización de datos en el mismo SI.
- *La integración*, mediante la comparación, fusión, cruce de datos tablas o BD internos y externos con otros sistemas de información.
- *La supervisión y vigilancia*, mediante procesos periódicos no estructurados de aseguramiento de la calidad de datos, incluye referencias cruzadas.
- *La auditoría*, con el objeto de revisar los datos en búsqueda de errores, inconsistencias, redundancias e información incompleta o inexistente.

En estas actividades se pueden identificar hasta cuatro actores que deben participar en el proceso:

- El patrocinador: define los planes de acción, el control de calidad de datos, organiza las reuniones y hace el seguimiento por encargo de la gerencia general.
- El analista de datos: colaborador que ejecuta los planes de análisis que genera métricas de calidad de datos que son guiados por el patrocinador.
- El usuario: corrige, valida, enriquece los datos faltantes, elimina los datos obsoletos o los consolida. Además, certifica la información de su responsabilidad.
- El informático: ejecuta los procesos de integración de datos para preparar los datos que deben ser usados por el analista de datos.

Tanto el analista de datos como el informático disponen de productos de software genéricos y software espe-

cializado para auditoría y para calidad de datos; éstos comprenden hasta cuatro actividades:

1. *Analizar*: entender o comprender las reglas,
2. *Normalizar*: alinear las fuentes con los servicios de información,
3. *Limpiar*: actualizar las reglas de calidad, y
4. *Mantener*: soluciones tácticas combinando la sociología de la organización con la infraestructura TI.

2.10. Gestión de maestros de datos

La gestión de maestros de datos (MDM = Master Data Management) es un concepto evolucionado para solucionar el problema de la integración de datos desde fuentes heterogéneas; este tipo de herramientas están teniendo demanda en las organizaciones porque el MDM se convierte en el único punto de gestión y de acceso a los datos de referencia para los SI, y de esta manera los SI reformulados y nuevos usan los servicios de datos estandarizados.

Los nuevos SI deberán mitigar sus riesgos porque no hay aplicaciones SOA de calidad sin datos de calidad. Al respecto, el MDM es la base de la gestión de procesos del negocio (BPM = Business Process Management) para lo cual se requiere un conjunto de reglas; es decir, no es posible pensar en procesos flexibles sin el uso de reglas y no hay reglas flexibles sin datos de calidad unificados y controlados.

En consecuencia, la oportunidad para disponer datos de calidad está sujeto a la adecuada ejecución de (véase Figura N.º 2):

- a. La gestión de maestros de datos (MDM);
- b. El sistema gestión de reglas del negocio (BRMS = Business Rules Management System); y
- c. La gestión de procesos del negocio (BPM).

Esta estructura, sustentada en un MDM y otras herramientas de software descritas precedentemente, permiten mejorar tanto la agilidad en el servicio de información como la coherencia de datos.

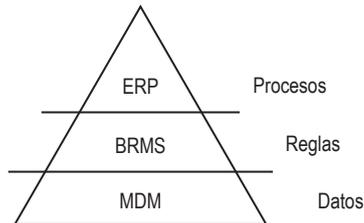


Figura N.º 2. Gestión de Maestros de Datos.

3. CONCLUSIONES

En conclusión, las organizaciones mejorarán la credibilidad en sus datos mientras sean conscientes del potencial problema de calidad de sus datos y la posibilidad de una solución; es decir, que el objetivo no sea tan solo proveer información de calidad, sino el de apoyar a la organización a hacer buen uso de la información para mejorar el uso de sus recursos y sus actividades (operacionales y de decisión) que influyen directamente en sus costos operativos.

Evidentemente, cualquier solución posible es consecuencia de un esfuerzo constante del personal de la organización y en especial de la TI, en donde se valide y se trabaje progresivamente sobre las dimensiones de calidad de datos (relevancia, exactitud, oportunidad y puntualidad, accesibilidad y claridad, comparabilidad o nivel de detalle, consistencia o coherencia, totalidad o complitud); cada una de las cuales está asociado directamente a la confiabilidad de la BD como del DW que se obtenga para cualquier nivel de decisión en la organización.

4. RECOMENDACIONES

Para ello se requiere un Plan de Acción, que será ejecutado por el personal de TI, desde las fuentes (carga) de información hasta los productos (salidas) de información pasando por los diversos procesos inherentes a la evolución de la información en la ejecución de los diversos procesos consistentes con las reglas del negocio de la organización.

Hoy en día existen herramientas especializadas para apoyar la mejora en la calidad de datos y que facilitan

dicha labor; algunas herramientas son específicas y otras que pueden ser adaptadas y cuentan con otras funcionalidades. Las herramientas se encuentran bajo diferentes denominaciones y acrónimos, entre otros, tales como: "Limpieza de Datos" (*Data Profiling, Data Validation, Data Auditing, and Data Cleansing*) o "Calidad de Datos" (*Data Quality*) o "Auditoría" (*Auditing*) o "Estadística" o "ETL" (*Extract, Transform, Loading*) o "Minería de Datos" (*Data Mining*) o "OLAP" (*On-Line Analytical Processing*) o "Inteligencia de Negocios" (*Business Intelligence*) o "Auditoría" (*Audit Tools*).

Finalmente, es necesario precisar que la buena calidad de datos en el ciclo de vida del SI puede ser usada por la organización en la inteligencia de negocios (análisis operativo, gestión de rendimiento, etc.), migración de datos (conversión y disponibilidad de datos), integración de los SI's (asegurar coherencia), aplicaciones SOA (acceso a datos y conversión), gestión de contenidos (unificación de datos estructurados y no estructurados), etc.

5. REFERENCIAS

- [1] Bustos Díaz S. 2007. Calidad de datos: un desafío necesario y permanente (1ra. parte). *Revista del Programa de Apoyo a la Competitividad de la Industria del Software (PACIS) de la Asociación Peruana de Productores de Software (APESOFT)*, Año 2, N.º 4.
- [2] Bustos Díaz S. 2007. Calidad de datos: un desafío necesario y permanente (2da. parte). *Revista del Programa de Apoyo a la Competitividad de la Industria del Software (PACIS) de la Asociación Peruana de Productores de Software (APESOFT)*, Año 2, N.º 7.
- [3] Eurostat. 2003. *Software Quality Framework*.
- [4] IIBA. 2007. *A guide to business analysis body of knowledge*, International Institute to Business Analysis, version 1.6, USA,
- [5] Lee Yang W, Strong DM. "Knowing-why about data processes and data quality", *Journal of Management Information Systems*. Vol. 20, N.º3. M. E. Sharpe, Inc. Armonk, NY, USA, pp. 13-39.
- [6] Olson JE. 2003. *Data Quality: The Accuracy Dimension*. Morgan Kaufmann,
- [7] Piattini, Miguel *et al.* 2007. *Calidad de sistemas informáticos*. Alfaomega, México.