

La calidad de los datos y las decisiones empresariales

Mercedes Delgado Fernández¹
José Alberto Vilalta Alonso²

Fecha de recepción: 14-02-08

Fecha de aceptación: 27-06-08

Abstract

In this paper it is commented of the paper of the data like support of the decisions that the directive make and as important resource of the company. So that the decisions are objective it is not enough they take based on the data but rather, also, these they should have the appropriate quality, what means that they are adapted to the use that they are sought to give. Reference is made to the effect that it has more than enough the economy, image and prestige of the organizations has the poor data quality, to the multidimensional character of the data quality and the necessity of taking initiatives that allow to solve these problems that are presented in this field.

Key Words

Data quality, quality, data.

Resumen

En este trabajo se comenta del papel de los datos como soporte de las decisiones que toman los directivos y como importante recurso de la empresa. Para que las decisiones sean objetivas no basta con que se tomen basadas en los datos sino que, además, estos deben tener la calidad adecuada, lo que significa que sean adecuados al uso que se les pretende

1. Doctora en Ciencias Técnicas. Profesora Titular. Ingeniera Industrial. Facultad de Ingeniería Industrial. Instituto Superior politécnico José Antonio Echeverría. La Habana, Cuba. mdelgado@ind.cujae.edu.cu.
2. Máster en Aseguramiento de la Calidad. Ingeniero Industrial. Profesor Auxiliar. Facultad de Ingeniería Industrial. Instituto Superior politécnico José Antonio Echeverría. La Habana, Cuba. jvilalta@ind.cujae.edu.cu.

dar. Se hace referencia al efecto que sobre la economía, imagen y prestigio de las organizaciones tiene la mala calidad de los datos, al carácter multidimensional de la calidad de los datos y a la necesidad de acometer iniciativas que permitan resolver los problemas que se presentan en este campo.

Palabras clave

Calidad de los datos, calidad, datos.

Introducción

Diariamente las personas se enfrentan a la necesidad de tomar decisiones, en todo campo de la vida. En el caso de los directivos de empresas, estas decisiones se toman a cualquier nivel y son de cualquier tipo, lo que cobra matices especiales, por la trascendencia de las decisiones y porque cada vez más se toman bajo la inmensa presión del tiempo y de la incertidumbre que caracteriza al más estable de los escenarios actuales (Gil-Aluja, 2000).

Además, cada vez más se hace evidente la necesidad de que esas decisiones se tomen, no basadas en la intuición sino en los datos (Gil-Aluja, 2000). La falta de una cultura del dato es una de las causas fundamentales que provocan esta situación.

Este artículo tiene como objetivo acercar a los directivos de empresas al concepto de calidad de los datos, al profundo impacto que la mala calidad de los datos tiene para sus organizaciones y a la necesidad de plantearse la realización de iniciativas que, en este campo, contribuyan a la solución de los problemas.

Los datos y el efecto de la mala calidad

Los datos son “un término general para denotar alguno o todos los hechos, letras, símbolos y números referidos a, o que describen un objeto, idea, situación, condición u otro factor” (Maynard, 1982) y constituyen un elemento fundamental para la toma de decisiones objetivas a todos los niveles de una organización (Javed & Hussain, 2003; Levy, 2004; Naveh & Halevy, 2000). Es más, para una organización moderna, los datos constituyen uno de sus recursos estratégicos (Olson, 2002; Tayi & Ballou, 1998).

Sin embargo, no basta con la existencia de datos ni con la voluntad de basar las decisiones que se tomen en ellos (Gil-Aluja, 2000; Javed & Hussain, 2003), se requiere que éstos tengan la calidad adecuada. Es decir que, cuando basados en los datos se favorezca una decisión sobre otra, se tenga la certeza de que los datos estén libres de errores y que, además, posean atributos relevantes (Redman 2001).

Es precisamente la mala calidad de los datos uno de los grandes problemas que han comenzado a enfrentar, en años relativamente recientes, las organizaciones y particularmente aquellas que poseen grandes volúmenes de datos de sus clientes y proveedores, que al ser erróneos, de alguna manera provocan cuantiosas pérdidas económicas, deterioro de la imagen y la pérdida de confianza de los clientes (Javed & Hussain, 2003; Levy, 2004; Redman, 2001).

Aunque esto se hace más evidente en situaciones como la anteriormente mencionada, otras organizaciones han detectado en la mala calidad de los datos generados por sus procesos productivos, de servicios y, fundamentalmente, investigativos, una de las causas fundamentales de la falta de objetividad en las decisiones que a diario toman (López, González & Goiret, 1994).

La literatura técnica relacionada con el tema es profusa en cifras y ejemplos del impacto que, desde los puntos de vista económico y de la calidad de la toma de decisiones en las organizaciones, tiene la pobre calidad de los datos. Son reflejo de esta situación la cantidad de datos erróneos en una organización (Javed & Hussain, 2003), la cantidad de errores en las bases de datos de clientes (Levy, 2004; Klein, 1998), las cuantiosas pérdidas, en dinero y en tiempo, que ocasionan la mala calidad de los datos, tanto por la atención a las quejas y reclamaciones de los clientes, como por la revisión y reparación de los datos (Kovac, Lee & Pipino, 1997; Redman, 1998; Wang, Kon & Madnick, 1993).

Pero el efecto de la mala calidad de los datos no sólo puede medirse en términos económicos (Javed & Hussain, 2003; Klein, 1998). También es considerable el efecto de la mala calidad de los datos en el éxito de muchas de las nuevas aplicaciones informáticas de ayuda a la toma de decisiones, como los Almacenes de Datos (Caro *et al.* 2006), la Minería de Datos y los Sistemas de Gestión de Relaciones con los Clientes (Olson, 2002).

La calidad de los datos

Sin embargo, son muy pocas las empresas que han tomado conciencia de esta situación y han comenzado a actuar en consecuencia, a pesar de que se conoce, por estudios realizados (Javed & Hussain, 2003; Levy, 2004), que si bien la mala calidad de los datos tiene un alto costo para las organizaciones, no es elevado el costo de prevenir los problemas.

Pero la mayoría de las empresas no toman en serio el problema de la calidad de los datos (Tayi & Ballou, 1998). Hay varias razones para ello. Por una parte está el hecho de que las grandes cantidades de datos que se generan en una organización moderna dificulta el proceso de detección y corrección de errores (Javed & Hussain, 2003; Levy, 2004), a pesar de la aparición en el mercado de novedosos sistemas informáticos destinados a la limpieza de datos (Kaposhtunas, 2002; Loshin, 2007). Por otra parte, los datos tienen características que los diferencian considerablemente de los otros recursos, como es el hecho de su intangibilidad, lo que dificulta su medición (Redman, 2001).

Hasta ahora se ha estado hablando de calidad de los datos, pero ¿qué es calidad de los datos? Al igual que ocurre con la definición de calidad de un producto o servicio no hay una definición única (Klein, 1998; Loshin, 2007; Rossel, 2003; Redman, 2001). No obstante hay algunos aspectos comunes en casi todas.

Uno de ellos es que la calidad de los datos significa adecuación al uso (Carliner, 2000; Klein, 1998; Lee & Strong, 2004; Redman, 2001; Tayi & Ballou, 1998; Wang, Kon & Madnick, 1992). Otro aspecto común tiene que ver con el hecho de que esta aptitud para el uso se mantenga a lo largo del ciclo de vida de los datos, desde su captación hasta su uso (Lee & Strong, 2004; Lee, 2004). Pero tal vez lo más significativo es el carácter multidimensional de la calidad

de los datos (Carlzon, 2000; Gendron & D'Onofrio, 2001; Javed & Hussain, 2003; Levy, 2004; Redman, 2001, Wang, 2004).

Esto quiere decir que la calidad de los datos está asociada a un conjunto de dimensiones o atributos que son los que la definen. Un objetivo fundamental de la definición de las dimensiones es poder establecer un lenguaje común y también focalizar los problemas de calidad de los datos y las oportunidades de mejora (Javed & Hussain, 2003; Naveh & Halevy, 2000).

Entre las dimensiones más importantes, pues son las más utilizadas y referenciadas, están la exactitud, la integridad, la consistencia y la coherencia (Cong *et al.*, 2007; Levy, 2004; Olson, 2002; Redman, 2001; Strong, Lee & Wang, 1997); es conveniente señalar que éstas deben ser definidas teniendo en cuenta las características propias de cada sector (Gendron & D'Onofrio, 2001).

A pesar de que se reconoce, al menos en la teoría, la importancia de la calidad de los datos (Javed & Hussain, 2003; Kovac, Lee & Pipino, 1997; Naveh & Halevy, 2000; Redman, 2001), estudios realizados indican que en Estados Unidos de América sólo una de cada cuatro empresas tiene desarrollados programas de calidad de los datos (Javed & Hussain, 2003), mientras que otro plantea que solo el 17% de las empresas encuestadas llevan a cabo iniciativas programáticas sobre la calidad de los datos (Redman, 1997). Se plantea que entre los principales beneficios de la implantación de programas de este tipo está la reducción de los costos operativos, la mejora del desempeño organizacional, la mejora del servicio al cliente y la mejora de las relaciones con éstos (Carlner, 2000, Carlzon, 2000).

Ya se habla de sistemas de calidad de los datos que soporten estas iniciativas y permitan que se integren a la gestión empresarial. Se definen como “la totalidad de los esfuerzos de una empresa para gestionar, controlar y mejorar la calidad de los datos”, que incluye actividades dirigidas a comprender las necesidades de los clientes, a detectar y corregir errores y a evitar errores futuros, así como actividades para construir una infraestructura de gestión para llevar a cabo lo anterior de forma eficaz y eficiente (Redman, 2001). Estos sistemas han ido evolucionando desde la simple detección y corrección de datos erróneos o duplicados hasta la consideración de que la detección y eliminación de errores forma parte importante de la gestión de los flujos de información que soportan los procesos fundamentales de las organizaciones (Loshin 2007).

Como parte importante de las iniciativas para la garantía de la calidad de los datos están los llamados sistemas de evaluación de la calidad de los datos y metodologías o procedimientos de diagnóstico, dirigidos a evaluar el estado actual de los datos con los que las organizaciones toman decisiones y a descubrir las causas que los originan para la toma de medidas que permitan prevenir o reducir su ocurrencia (Carlzon, 2001; Cong *et al.*, 2007; Fajardo & Crispido, 2004; Olson, 2002; Vilalta & Espinosa, 2006). Estos sistemas son muy variados, tienen como premisa la clara definición del uso que se le dará a los datos (Carlner, 2000; Klein, 1998; Loshin 2007; Lee & Strong, 2004; Redman, 2001; Tayi & Ballou, 1998; Wang, Kon & Madnick, 1992) y utilizan diferentes métodos y herramientas para el logro de su objetivo, que van desde técnicas estadísticas, técnicas heurísticas (Cong *et al.*, 2007), el uso de plataformas informáticas (Fajardo & Crispido, 2004, Olson, 2002), hasta herramientas para la identificación y solución de problemas (Lee, 2004; Vilalta & Espinosa, 2006).

Conclusiones

La toma efectiva de decisiones por parte de los empresarios se ve afectada, en buena medida, por la mala calidad de los datos que se utilizan. A pesar de que los esfuerzos por prevenir los problemas de calidad de los datos son menos costosos, desde todo punto de vista, que el efecto de la mala calidad de los mismos, aún no existe una arraigada conciencia por parte de los directivos para revertir esta situación, aunque ya se han ido dando pasos para el establecimiento y mantenimiento de iniciativas programáticas dirigidas a mejorar la calidad de los datos.

Bibliografía

1. Carliner, S. (2000). Physical, Cognitive, and Effective: A three-part framework for Information Design. Technical Communication.
2. Carlzon C. (2001). Toward a Framework for Assessing Data Quality. International Monetary Fund.
3. Cong, G., Fan, W., Geerts, F., Jia, X. & Ma, S. (2007). Improving data quality: consistency and accuracy. Proceedings of the 33 International Conference on Very Large Data Base.
4. Fajardo, F. & Crispido, I. (2004). Data Quality Evaluation, Una herramienta para evaluar la calidad de los datos en un sistema de información multifuente. X Congreso Argentino de Ciencias de la Computación (Cacic'2004), La Matanza, Argentina.
5. Gendron, M. & D'Onofrio, M. (2001). Data Quality in the Healthcare Industry. Data Quality, Vol. 7, No. 1.
6. Gil – Aluja, J. (2000). Las decisiones y la incertidumbre. Barcelona.
7. Javed, B. & Hussain, S. (2003). Data quality – A problem and an approach. Wipro Technologies.
8. Kaposhunas, A. (2002). Data Cleansing: The Foundation of Customer Information Management. www.dnb.com.
9. Klein, B. (1998). Data quality in the Practice of Consumer Product management: evidence from the field. Data Quality, Vol. 4, No. 1.
10. Kovac, R., Lee, Y. & Pipino, L. (1997). Total Data Quality Management: The case of IRI. Second Conference on Information Quality.
11. Lee, Y. & Strong, D. (2004). Knowing – why About Data Processes and Data Quality. Journal of Management Information Systems. Vol. 20, No. 3, pp. 13 – 39.
12. Lee, Y. (2004). Crafting Rules: Context-Reflective Data Quality Problem Solving. Journal of Management Information Systems. Vol. 20, No. 3, pp. 93 – 119.
13. Levy, S. (2004). Model Documents and forms for Organizing and Maintaining a Data Quality Program. www.dataqualitymodeldocument.com

14. López, C., González, E. & Goiret, J. (1994). Análisis por componentes principales de datos pluviométricos. Aplicación a la detección de datos anómalos. Estadística.
15. Loshin, D. (2007). Data Profiling, Data Integration and Data Quality: The Pillars of Master Data Management. White Paper.
16. Maynard, J. (1982). Dictionary of Data Processing. Londres, Inglaterra.
17. Naveh, E. & Halevy, A. (2000). A hierarchical framework for a quality information system. Total Quality Management, Vol. 11, No. 1, p 87-111.
18. Olson, J. (2002). Data Profiling: The Data Quality Analyst's Best Tool. DM Direct, December. DMReview.com
19. Redman, T. C. (1997). Data Quality in the Information Age. Artech House.
20. Redman, T. C. (1998). The Impact of Poor Data Quality on the Typical Enterprise. Communications of the ACM, Vol. 41, No. 2, pp 79 – 82.
21. Redman, T. C. (2001). Sistemas de Calidad de Datos de Segunda Generación. Manual de Calidad, McGraw Hill, Madrid, España.
22. Rossel, R. (2003). Mejoramiento de la Calidad de los datos en un Data Warehouse.
23. Strong, D.M., Lee, Y.W. & Wang R.Y. (1997). 10 Potholes in the Road to Information Quality. IEEE Computer, Vol. 30, No. 8, pp. 38 – 46.
24. Tayi, G. & Ballou, D. (1998). Examining Data Quality. Communications of the ACM. Vol. 41, No. 2.
25. Vilalta, J. & Espinosa, M. (2006). Propuesta Metodológica para el diagnóstico de la calidad de los datos. IV Simposio de Ingeniería Industrial y Afines (SIIA), Cujae, La Habana, Cuba.
26. Wang, R., Kon, H. & Madnick, S. (1992) Data Quality Requirements Analysis and Modeling. Ninth International Conference of Data Engineering.
27. Wang, R. (2004). Data Quality: Theory in Practice. EPA 23rd Annual Conference on Managing Environmental Quality Systems.