



APLICACIÓN DE UN PROCEDIMIENTO PARA EL DIAGNÓSTICO DE LA CALIDAD DE LOS DATOS EN UNA EMPRESA DE PRODUCCIÓN

Resumen / Abstract

La calidad de los datos es un factor que se tiene en cuenta cada vez más en la empresa moderna, lo cual está motivado sobre todo por los costos que se han asociado a la pobre calidad de los datos, es decir, los costos por la detección y corrección de errores en los datos, la insatisfacción de los clientes, las decisiones erróneas, etc. Para tratar el tema de la calidad de los datos han surgido diversos productos informáticos que brindan una gran ayuda, pero que son insuficientes, ya que se centran en la corrección de errores y no en la búsqueda de las causas de los problemas. Justamente el objetivo del presente trabajo es evidenciar la utilidad del Procedimiento de Diagnóstico de la Calidad de los Datos, el cual está sustentado en técnicas de gestión, a partir de mostrar los resultados obtenidos con su aplicación en una empresa de producción.

Data quality is a factor that is been considering more and more in modern enterprises, which is motivated mainly because of costs associated to poor data quality, for instance, data error detection and correction costs, costumer dissatisfaction, wrong decisions, etc. In order to deal with data quality topic, several information technology products have been developed which help out a lot although are not sufficient due to their primary intention is correction of data errors instead of detecting quality problem causes. Indeed the goal of this paper is to show the usefulness of The Methodology for Diagnosing Data Quality, which is based on management methods, taking as point of starting the results obtained while getting into practice this Methodology within a production enterprise.

Palabras clave / Key words

Datos, Calidad de datos, Procedimiento para el Diagnóstico de Calidad de Datos.

Data, Data quality, Methodology for Diagnosing Data Quality.

José A Vilalta Alonso, Doctor en Ciencias Técnicas, Ingeniero Industrial, Profesor auxiliar, Departamento de Ingeniería Industrial, Facultad de Ingeniería Industrial, Instituto Superior Politécnico "José A. Echeverría", CUJAE, Ciudad de la Habana, Cuba.
e-mail: jvilalta@ind.cujae.edu.cu

Jobany J. Heredia Rico, Ingeniero Industrial, Profesor Adiestrado, Departamento de Ingeniería Industrial, Facultad de Ingeniería Industrial, Instituto Superior Politécnico "José A. Echeverría", CUJAE, Ciudad de la Habana, Cuba.
e-mail: jherediar@ind.cujae.edu.cu

Recibido: 05/12/2008

Aprobado: 19/12/2008

INTRODUCCIÓN

Los grandes avances que se han producido desde hace algunos años en las esferas de la informática y las telecomunicaciones le han valido el calificativo a nuestros tiempos de "La era de la Información". Esta ha sido una de las razones por las cuales actualmente se ha enfatizado la importancia de gestionar la información como vía para optimizar el uso de los demás recursos con que cuenta una empresa. La información ha sido tradicionalmente entendida como un resultado que se obtiene al agrupar, clasificar y en general, agregarle valor a los datos [1]. En efecto, los datos son la materia prima de la información. Esta estrecha relación explica el por qué los directivos de empresas, como vía para obtener una información más amplia acerca de los procesos desarrollados en la organización, tratan de generar mayor cantidad de datos relacionados con estos procesos valiéndose de las posibilidades que hoy día brindan las tecnologías de captura y almacenamiento de datos.

No obstante, para lograr obtener una información apropiada y válida a partir de los datos, éstos tienen que poseer determinadas características deseables, es decir, deben tener calidad [2]. En este sentido se han realizado diversas investigaciones desde hace algunos años enfocadas al tema de la adecuación o calidad de los datos. Entre los factores que han motivado este enfoque respecto a los datos se encuentran los diversos sistemas que se alimentan de grandes volúmenes de datos, como los almacenes de datos, el enfoque de Gestión de Relaciones con los Clientes, la minería de datos, etc. Además, el hecho de que los datos están asociados con los clientes, los productos y servicios, el mercado y las cuestiones financieras, y son usados para conducir cualquier operación relacionada con estos elementos, desde las más comunes hasta las de planificación y toma de decisiones [2; 3].

La significación de la calidad de los datos también ha aumentado por los diferentes costos que han sido asociados a los problemas de mala calidad en los datos, como por ejemplo, el costo de la detección y corrección de errores en los datos, la insatisfacción de los clientes, la insatisfacción de los empleados, las decisiones erróneas y los fallos que se producen en la aplicación de nuevas tecnologías [4]. A pesar de que en la actualidad se ha creado cierta alerta, muchas empresas continúan sin enfrentar el problema de la calidad de los datos, lo cual está dado fundamentalmente por la intangibilidad de los datos, elemento éste que dificulta la medición de su calidad, así como la evaluación del beneficio esperado, en términos económicos, de los programas de mejora de calidad de los datos [2; 4].

Como es lógico, para tratar el tema de la calidad de los datos han surgido una amplia gama de productos informáticos que proporcionan una gran ayuda en este sentido. Las funcionalidades de estos productos van desde la corrección y detección de errores en las bases de datos, hasta la prevención de errores en la entrada de datos¹. No obstante a las ventajas que brinda la informática en este sentido, para mejorar definitivamente la calidad de los datos se deben utilizar técnicas de gestión, ya que para encontrar las causas de los problemas e incluso para mejorar la calidad de los datos, se necesita enfocarse en cuestiones no sólo relacionadas con el uso de tecnología de punta [2]. Pretender mejorar la calidad de los datos sólo a partir de buscar y corregir errores sería actuar sobre los efectos del problema, cuando realmente lo más importante es detectar las causas de estos errores.

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 originan los problemas de calidad con el fin de tomar medidas que permitan prevenir o reducir su ocurrencia [5; 6].

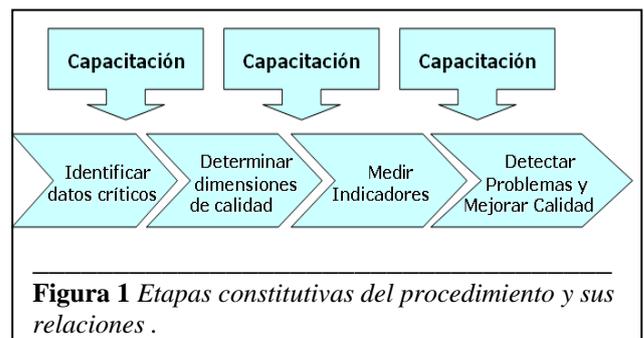
¹ La detección y corrección de errores en los datos son las tareas fundamentales realizadas por los “Software de Calidad de los Datos”. La prevención de errores en la entrada de datos se garantiza con un adecuado diseño de la entrada de datos.

Dada esta panorámica, en nuestro país se hace necesario incorporar a la práctica empresarial la filosofía y conceptos relacionados con la calidad de los datos, con el objetivo de minimizar los problemas derivados de un mal comportamiento en este sentido, lo cual sin duda ayuda a maximizar los beneficios que reporta la actividad industrial a nuestra sociedad.

Esta situación dictó la necesidad de diseñar una metodología de Diagnóstico de la Calidad de los Datos. Al aplicar esta metodología en una organización se pretende determinar los problemas de calidad de los datos que regularmente afectan en la misma y en función de éstos, establecer planes de mejora. Precisamente, el objetivo de este artículo es mostrar la utilidad del Procedimiento de Diagnóstico de la Calidad de los Datos a partir de los resultados obtenidos con su aplicación en una empresa productora de envases.

PROCEDIMIENTO PARA EL DIAGNÓSTICO DE LA CALIDAD DE LOS DATOS

El procedimiento diseñado para el diagnóstico de la calidad de los datos, tiene como base el uso de herramientas de gestión. El objetivo general de este procedimiento es diagnosticar la calidad de los datos, brindando como resultados más importantes los principales problemas que la afectan, así como un plan de acciones encaminado a eliminar y prevenir éstos. El procedimiento está constituido por diferentes etapas o actividades relacionadas entre sí, de forma que todas tributan al logro de los principales objetivos del mismo pues en cada una de ellas se obtiene un resultado necesario para las subsiguientes etapas o para completar el logro de algún objetivo parcial [6]. A continuación en la Figura 1 se muestra una representación de las etapas constitutivas del procedimiento y la relación entre las mismas:



Seguidamente se comentan los objetivos individuales y resultados a obtener en cada una de las etapas.

Etapa 1: Identificación y ordenamiento de datos críticos

Dado que para la aplicación del procedimiento se cuentan con recursos limitados (financieros, tiempo, etc), es importante determinar sobre cuáles, del gran conjunto de datos que existen en una organización, se va a trabajar. Es bastante probable que detectando las causas que provocan la mala calidad de una parte del gran conjunto de datos, se logren establecer medidas que eliminen este problema de manera general. Es por estas razones que en esta etapa se seleccionan, teniendo en cuenta el uso previsto de los datos bajo análisis y la importancia que éstos tengan

APLICACIÓN DE UN PROCEDIMIENTO PARA EL DIAGNÓSTICO DE LA CALIDAD DE LOS DATOS EN UNA EMPRESA DE PRODUCCIÓN

respecto a este uso, los tipos de datos a los que se les realizará el proceso de medición de calidad.

La selección de los datos más críticos se desarrolla a partir de considerar las evaluaciones dadas por los usuarios de los datos y en general, de los trabajadores ligados a los procesos de producción de datos (especialistas)².

Etapas 2: Determinación de las dimensiones de calidad de los datos

Las dimensiones de calidad de los datos no son más que las características o propiedades que deben poseer los datos para ser considerados como adecuados [7]. La selección de estas dimensiones es una actividad muy importante, pues equivale a decidir qué características de calidad del producto, en este caso un dato, serán analizadas. Como el grado de adecuación o aptitud de un producto (dato) depende del cliente (en este caso los usuarios de los datos), se hará necesario definir las dimensiones de calidad a analizar, a partir de los requerimientos que tengan los usuarios respecto a los datos.

Después de definidas las dimensiones, será importante ponderar las mismas según su importancia, para lo cual también se tendrá en cuenta el criterio de los usuarios de los datos. Finalmente será necesario en esta etapa definir indicadores que permitan evaluar, en una etapa posterior, el grado de adecuación del dato atendiendo a cada dimensión.

Etapas 3: Medición de indicadores

Esta etapa tiene como objetivo medir la calidad de los datos críticos que hayan sido definidos anteriormente. Como indicadores de la calidad de los datos se puede calcular el porcentaje de datos correctos ($1 - \text{tasa de error}$) para cada uno de los atributos seleccionados, el valor real de los indicadores asociados a cada dimensión definida y la proporción del total de errores encontrados que son graves, de gravedad media y menos graves [8].

Otro aspecto importante en esta etapa es comprender el comportamiento de los indicadores, es decir, analizar los mismos. Para esto es muy ventajoso haber calculado los indicadores en alguna aplicación previa y así poder contar con la base comparativa suficiente para enriquecer el análisis.

Específicamente para el análisis ha sido propuesto el uso de gráficos de pastel o diagramas de barras para resaltar el comportamiento de indicadores respecto a determinadas clases (por ejemplo la gravedad de los errores); la comparación entre los indicadores obtenidos para cada dimensión y entre los tipos de datos; los gráficos de control, específicamente los de p o np para analizar el comportamiento del porcentaje defectuoso ya sea de un tipo de dato o de las dimensiones asociadas a éste y la comparación, mediante cualquier método, entre los indicadores de diferentes períodos. En este caso será muy útil el cálculo y comparación en diferentes períodos de un indicador que englobe el valor real de los indicadores asociados a las dimensiones y la ponderación de las

mismas, para lo cual se diseñó el Índice Ponderado de Calidad del Dato (IPCD).

Una vez que se utilicen las herramientas para el análisis del comportamiento de los indicadores, debe procederse al análisis, propiamente dicho, de este comportamiento. Para ello deben considerarse los indicadores que provocan un mayor efecto en la mala calidad de los datos, los tipos de errores más frecuentes, su gravedad y los tipos de datos más afectados por estos errores. También debe tenerse en cuenta el comportamiento cronológico de los indicadores.

Etapas 4: Análisis de los problemas y mejora de la calidad del dato

Esta etapa tiene como primer objetivo determinar las causas raíces de los problemas de calidad de los datos encontrados. Para la detección de estas causas ha sido propuesta la aplicación del enfoque de gestión del error humano denominado “Enfoque del Queso Suizo” [9], teniendo en cuenta la importante influencia de este factor. Para una búsqueda más organizada de las causas utilizando este enfoque, se comenzó estableciendo las posibles barreras, que en el caso del análisis con datos, podían ayudar a prevenir la existencia de errores. Las barreras definidas fueron: los procesos de producción de datos, los trabajadores involucrados en estos procesos, la tecnología informática, el ambiente de trabajo y las normas y regulaciones [9].

A pesar de ser difícil el establecimiento de un límite claro entre estas barreras, pues por ejemplo, la barrera constituida por los “procesos de producción de datos” podría incluir alguna de las otras, se entendió oportuna su individualización ya que esto facilitará la detección de algunas causas de problemas de calidad de datos que pudieran estar afectando de manera muy específica una de estas barreras. En general, esta individualización posibilita el establecimiento de planes de mejora más efectivos. A partir de las barreras definidas contra errores en los datos, se necesitará buscar, en el estudio específico que se haga, las causas que afectan cada barrera. Para esto es recomendable la utilización de un diagrama causa-efecto donde las causas se desglosan quedando así más evidente donde se encuentran las raíces.

El otro objetivo de esta etapa es realizar planes de acción encaminados a la mejora de la calidad de los datos. No obstante a que la mejora de la calidad de los datos tiene sus peculiaridades según el proceso que se analice, pueden ser definidas algunas buenas prácticas en sentido general, por ejemplo, el diseño y descripción adecuada de los procesos de producción de datos, el uso de técnicas estadísticas en estos procesos, el diseño adecuado de la base de datos, la entrada doble datos, etc.

Etapas 5: Capacitación del personal

Con esta etapa se pretende dar una capacitación a trabajadores relacionados con los procesos de producción de datos y al personal directivo, enfocada en brindarle conocimientos sobre calidad de los datos (relacionados específicamente con la aplicación del procedimiento y conocimientos generales sobre esta temática), en aras de que este personal pueda contribuir a la correcta aplicación del mismo.

² Para más detalles sobre el procedimiento consultar referencia [6]

APLICACIÓN DEL PROCEDIMIENTO DE DIAGNÓSTICO DE LA CALIDAD DE LOS DATOS EN UNA EMPRESA PRODUCTORA DE ENVASES DE TABACO

La investigación fue desarrollada en una unidad de producción dedicada a la confección de envases de madera, de disímiles modelos. Las producciones se realizan con una tecnología tradicional, destacándose entre los equipos más utilizados las sierras, las máquinas de lijado y las máquinas de impresión y pintura.

Dentro de esta unidad el objeto de estudio lo constituyó el flujo de información que se produce desde el proceso productivo hacia el departamento de economía de la unidad, cuyo objetivo es registrar los datos que se utilizan para calcular el costo de las diferentes órdenes de producción. Los datos que se reciben en el departamento de economía se registran en una hoja de datos Excel (Excel de Costo) con el fin comentado. Esta hoja de datos se alimenta con datos de diversos documentos provenientes directamente de la producción.

Resultados de la Etapa 1

De todos los tipos de datos registrados en la hoja Excel que constituían un total de once, se realizó una selección de tres tipos de datos como los más críticos. Para esto se utilizó la opinión de tres usuarios de los datos y de dos trabajadores considerados como especialistas, debido a su relación con el proceso de producción de datos. Para realizar la selección de los tipos de datos se definió como dato más crítico a aquel dato que con más necesidad se requería que estuviera libre de error para lograr un cálculo adecuado del costo de producción de una determinada orden. Después de definido esto, se le pidió a usuarios y especialistas que ponderan los datos considerados según la importancia que ellos les asignaban.

Después de analizar la existencia de concordancia entre las calificaciones dadas por el personal seleccionado, se realizó un Diagrama Pareto con el total de puntuación que tenía cada tipo de dato y se determinó que tres de estos datos eran los más importantes al acumular el 50,3 % de la puntuación total dada a los once atributos. Sobre los valores de estos tres tipos de datos³ se realizaron las mediciones en etapas posteriores.

Resultados de la Etapa 2

A partir de los requerimientos o expectativas que declararon tener los usuarios y especialistas respecto a los valores de los datos, se definieron para los tres tipos de datos considerados las siguientes dimensiones:

- 1) Integridad: Un dato se consideró íntegro, si se reportaban en la hoja de datos Excel, todas las cantidades que se reportaban en los registros primarios.
- 2) Exactitud: Un dato se consideró exacto, si los valores de las cantidades que se reportaban en la hoja de datos Excel coincidían con los valores de los registros primarios. Nótese de la definición de exactitud que aunque faltara algún valor por reportar en la hoja Excel, si los valores que se reportaban, coincidían con los de los registros primarios, pues el dato se consideró exacto.

- 3) Consistencia: Un dato se consideró consistente si no se reportaban más cantidades en la hoja de datos Excel que en la fuente primaria.

Después de definir estas dimensiones es importante notar que un dato para ser considerado correcto, o sin error, no debía tener problemas, ni de exactitud, ni de integridad, ni de consistencia.

Una vez definidas las dimensiones se utilizó el criterio de los cinco trabajadores participantes en la identificación del dato crítico para asignar un nivel de importancia a las mismas. Después de que este personal realizó la ponderación y se determinó que existió concordancia en el ordenamiento dado, se obtuvo como resultado que la dimensión más importante fue la integridad con una importancia relativa del 47%, después la exactitud con una importancia del 30 % y finalmente la consistencia con una importancia del 23 %.

En esta etapa también se definieron los indicadores que se utilizarían en la evaluación de las dimensiones de calidad de los datos. Para las tres dimensiones se definieron indicadores que reflejaban el porcentaje de datos que cumplían con la definición dada para cada dimensión. Además, dado que se tiene la ponderación de cada dimensión, se planteó calcular el Índice Ponderado de Calidad del Dato (IPCD). Con este índice se tendrá en cuenta no sólo el valor de las dimensiones obtenidas para el dato, sino también la importancia de las mismas, y será muy importante la comparación de los valores de este indicador en diferentes períodos para saber si ha habido mejoras respecto a la calidad de los datos. El IPCD se calcularía según la expresión 1:

$$IPCD_i = \sum W_i D_i \quad (1)$$

Donde:

i: tipo de dimensión (Integridad, Exactitud y Consistencia)

D: Dimensión.

W: Importancia relativa de la Dimensión

Resultados de la Etapa 3

Después de ubicar los registros primarios con los cuales se compararían los valores de los datos almacenados en la hoja de datos Excel, se procedió al cálculo de los indicadores. La Figura 2 muestra una representación del porcentaje de datos correctos en cada uno de los tipos de datos analizados.

Los resultados más desfavorables se obtienen con los datos relacionados con el "Consumo de Acetato", ya que el porcentaje de datos correcto es inferior al 80 %. No obstante, la tasa de error (100 – porcentaje correcto) en cada uno de los tipos de datos es superior a 5%, lo que evidencia, en general, la existencia de problemas.

Por otra parte se calcularon indicadores referentes a la gravedad de los errores detectados en cada uno de los tipos de datos. En la Figura 3 se muestran tres gráficos de pastel con la distribución de gravedad de los errores ocurridos en cada uno.

³ Los tipos de datos seleccionados fueron "Cantidad de Cajas Terminadas", "Consumo del Material Plywood" y "Consumo de Material Acetato".

APLICACIÓN DE UN PROCEDIMIENTO PARA EL DIAGNÓSTICO DE LA CALIDAD DE LOS DATOS EN UNA EMPRESA DE PRODUCCIÓN

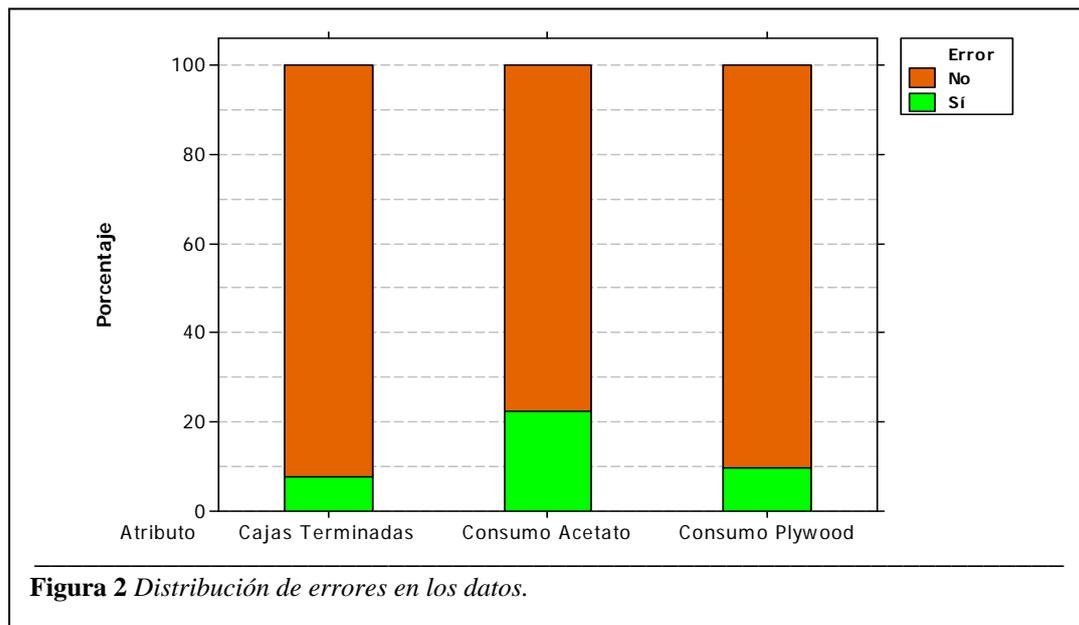


Figura 2 Distribución de errores en los datos.

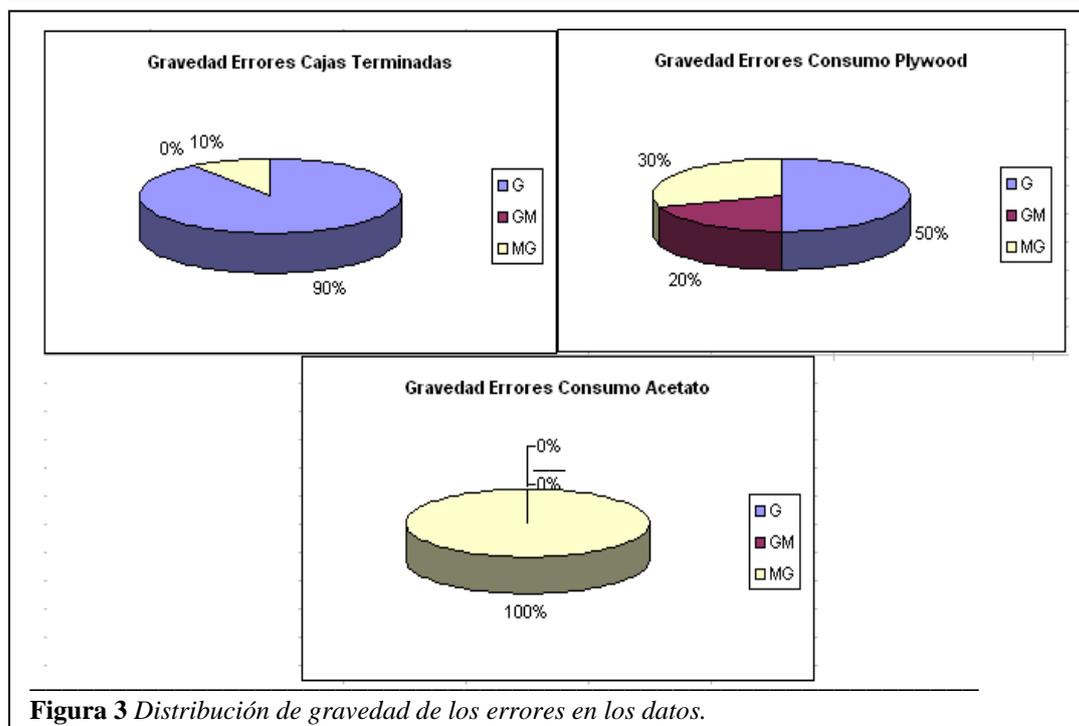


Figura 3 Distribución de gravedad de los errores en los datos.

Como se observa, en el caso de los errores en los datos de “consumo de plywood”, la gravedad está distribuida en las tres categorías, aunque predominan los errores graves, mientras que en los datos de “Cantidad de Cajas Terminadas” y “Consumo de Acetato”, predominan marcadamente los errores graves y menos graves respectivamente.

Este resultado debe ser considerado en el momento de asignar recursos para mejorar la calidad de los datos, ya que, aunque en los datos de “consumo de acetato” es donde existe una mayor tasa de error, las consecuencias de los errores no son tan grandes como en el caso de los errores en los datos de “Cantidad de Cajas Terminadas”, donde según los resultados, es altamente probable que cada vez que ocurra un error sea de consecuencias considerables para la toma de decisiones.

Otros indicadores calculados fueron los relacionados con las dimensiones. En este caso se realizó un análisis de la distribución de gravedad de los errores asociados a cada dimensión, obteniéndose como resultado que todos los errores de integridad fueron graves o de gravedad media, en el caso de la exactitud un 33% de los errores fueron menos graves y un 67% graves, y los errores de consistencia todos resultaron menos graves.

También se obtuvo como resultado respecto al análisis de las dimensiones que ocurrieron 24 problemas de integridad, 17 de exactitud y 8 de consistencia, es decir, parece suceder que los problemas de integridad producen las consecuencias más graves y también son los más frecuentes. Este aspecto debe considerarse en el momento de la búsqueda y eliminación de las causas de los

problemas de calidad, pues se debe tratar, indefectiblemente, de minimizar la ocurrencia de las causas asociadas a los problemas de integridad.

Otro resultado importante obtenido fue el relacionado con la afectación que producían los errores en los datos de “Consumo de Plywood” y “Consumo de Acetato” sobre el indicador económico “Gasto de Material Mensual”. En el gráfico de la Figura 4 se muestra este efecto. El gráfico se realizó considerando la diferencia en cada mes, entre el valor que se reportó de más y el valor que se dejó de reportar debido a los errores en los datos relacionados con estos materiales.

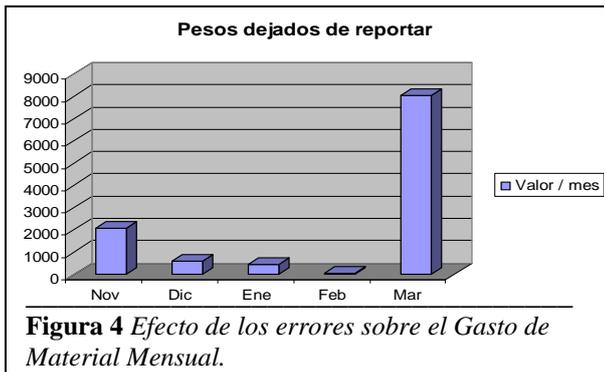


Figura 4 Efecto de los errores sobre el Gasto de Material Mensual.

Como se observa en todos los meses la mala calidad de los datos produjo como efecto global que se dejaron de reportar valores en el “Gasto de Material Mensual”. El valor promedio de pesos dejados de reportar fue de \$2249. En resumen este comportamiento es evidencia de que en cada mes ocurren gastos reales en la empresa de los cuales el departamento de economía nunca llega a saber. Esto implica, en principio, decisiones profundamente erróneas en relación a cualquier análisis que se pretenda realizar respecto al comportamiento del indicador “Gasto de Material Mensual”.

Finalmente se detectó que la mala calidad de los datos provoca pérdidas de tiempo, ya sea el tiempo que dedican los trabajadores a la corrección de errores en los datos o a la espera por la llegada de datos para procesar. Para medir las implicaciones de la mala calidad de los datos en el aprovechamiento del tiempo de trabajo se aplicó una encuesta a nueve trabajadores⁴. Los resultados más relevantes obtenidos fueron que 6 de los 9 trabajadores encuestados plantearon dedicar entre “mucho” y “bastante tiempo” a la rectificación de errores en los datos. Además 5 trabajadores expresaron recibir “casi nunca” los datos en tiempo. Quizás el resultado más importante es el que muestra que a todos los trabajadores encuestados los afecta en su desempeño, al menos “bastante”, las pérdidas de tiempo por estos conceptos.

Resultados de la Etapa 4

Para encontrar las causas de los problemas de calidad de los datos se definieron como efectos, los problemas asociados a las dimensiones. Partiendo de esto se construyeron tres diagramas causa-efecto donde se reflejaron las causas y subcausas que provocaban estos

⁴ Los nueve trabajadores consistían en los cinco que participaron en la primera etapa, más cuatro trabajadores relacionados con los datos analizados, pero más ligados al área productiva.

problemas. Algunas de las causas más importantes detectadas fueron:

- Desmotivación respecto a la calidad de los datos, provocada por la falta de cultura existente en los trabajadores respecto a estos temas.
- Carencia de procedimientos de trabajo estándares para capturar, almacenar y utilizar los datos.
- Falta de control sobre la manera en que se ejecutan los procesos de producción de datos lo que conlleva fundamentalmente a indisciplinas de trabajo que a su vez provocan problemas de calidad de datos.
- Poca habilidad de los trabajadores relacionados con los procesos de producción de datos para el trabajo con computadoras.
- Documentos primarios mal diseñados, es decir, la estructura espacial de los datos en estos documentos es inadecuada. Esta dificultad se acentuó, pues los datos se registraban de manera ilegible (escritos a mano o impresos con una impresora deficiente).
- Los trabajadores de los procesos de producción de datos laboraban en condiciones ergonómicas desfavorables sobre todo desde el punto de vista biomecánico y de iluminación, aunque algunos trabajadores eran afectados también por el ruido.
- La aplicación de entrada de datos estaba mal diseñada pues se hacía poco uso de las ventajas informáticas, además, era desordenada y tendiente a la confusión en sí misma.

Para la detección de estas causas fue favorable utilizar el enfoque del Queso Suizo ya que en muchas ocasiones la detección partió de asociar los diferentes efectos con alguna posible debilidad en las barreras de defensa definidas a partir de este enfoque. En la Figura 5 se muestra una representación de las principales causas que afectan cada una de las barreras de defensa contra errores en los datos.

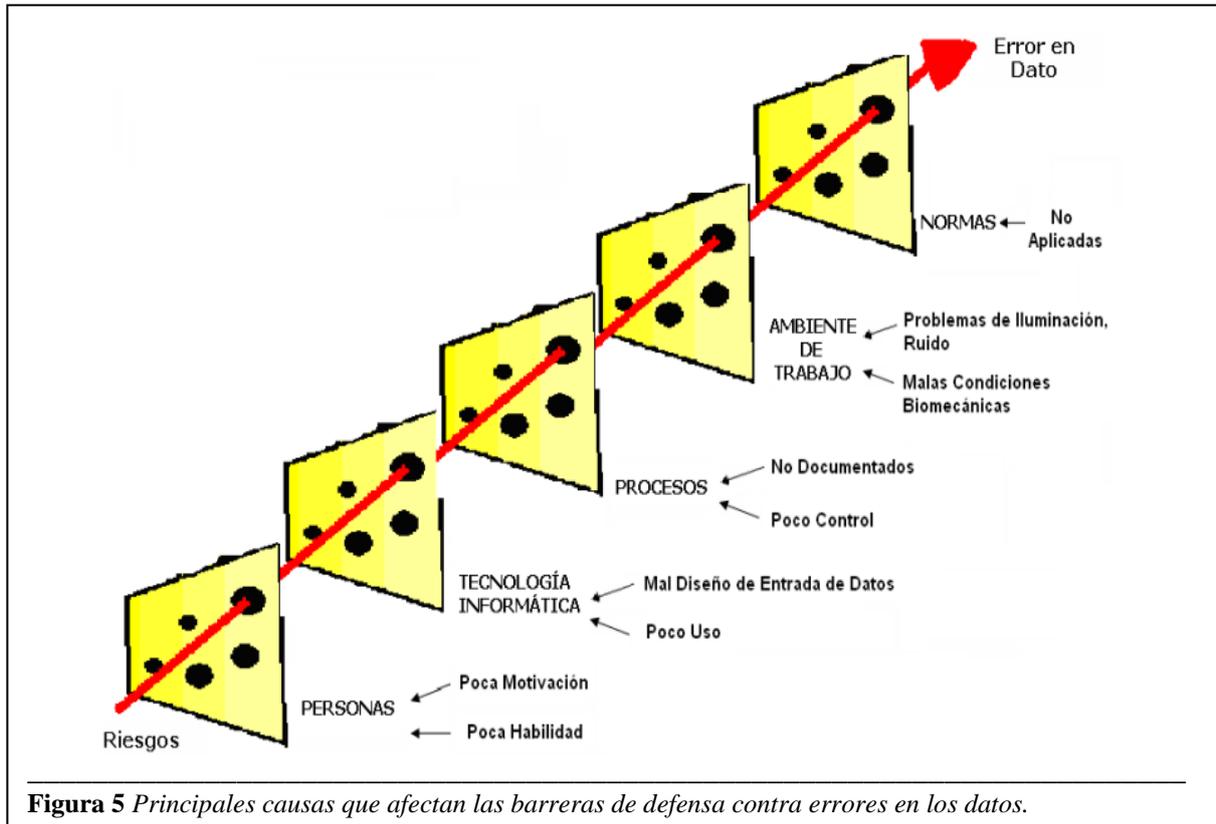
Concluyendo la aplicación del procedimiento se realizaron planes de mejora enfocados en eliminar o minimizar la influencia de las causas detectadas. Los planes comprendieron:

- Mejorar las condiciones ergonómicas.
- Mejorar el diseño de entrada de datos.
- Establecer el uso de técnicas estadísticas para la detección de valores atípicos (posibles errores).
- Incrementar habilidades y motivación de los trabajadores.
- Rediseñar y documentar procesos de producción de datos.

CONCLUSIONES

- 1- La aplicación del Procedimiento de Diagnóstico de la Calidad de los Datos en la empresa productora bajo análisis, posibilitó la detección tanto de problemas de calidad de los datos, como de las causas de éstos.
- 2- La calidad de los datos estudiados se puede calificar como inadecuada, ya que entre otras cosas el porcentaje de error que presenta cada tipo de dato es superior al 5 %.
- 3- La mala calidad de los datos estudiados no afecta sólo la toma de decisiones, sino también el desempeño de los trabajadores involucrados en los procesos de producción de datos.

APLICACIÓN DE UN PROCEDIMIENTO PARA EL DIAGNÓSTICO DE LA CALIDAD DE LOS DATOS EN UNA EMPRESA DE PRODUCCIÓN



4- Los errores en los datos son consecuencia no sólo del carácter falible de las personas que trabajan en la producción de los mismos, sino además del mal uso de las ventajas informáticas y los problemas organizativos existentes.

5- Los planes de mejora derivados de la aplicación del procedimiento promueven el correcto diseño de los procesos de producción teniendo en cuenta no sólo el adecuado uso de la tecnología informática, sino también la creación de condiciones ergonómicas adecuadas, la correcta selección del personal para trabajar en estos procesos, así como otros aspectos que minimicen la posibilidad de ocurrencia de errores en los mismos.

REFERENCIAS

1. PONJUÁN, Gloria. *Gestión de Información en las Organizaciones*. [s.l.]: Editorial Félix Varela, 2006.
2. REDMAN, Thomas. "Sistemas de calidad de datos de segunda generación". En: *Manual de Calidad de Juran*. [s.l.]: McGraw Hill, 2001.
3. LOSHIN, David. "Integration and the Data Quality Imperative. The Data Quality Monitor". [en línea]. 2001, Disponible en: <http://www.datajunction.com>
4. REDMAN, Thomas. "Data: An unfolding quality disaster". [en línea]. 2004, Disponible en: <http://www.dmreview.com/portals/dataquality>
5. CONG, G. et al. "Improving data quality: consistency and accuracy". En: *Proceedings of the 33 International Conference on Very Large Data Base* ([s.l.]: [s.n.], 2007)
6. VILALTA, José. "Procedimiento para el Diagnóstico de la Calidad de los Datos. Una nueva versión". En: *14 Conferencia de Ingeniería y Arquitectura* (Cujae, La Habana, Cuba: [s.n.], 2008)
7. PIPINO, Leo; LEE, Yang y WANG, Richard. "Data Quality Assessment". [en línea]. 2002, Disponible en: <http://web.mit.edu/tdqm/www/tdqmpub/PipinoLeeWangCACMApr02.pdf>.
8. HEREDIA, Jobany. "Procedimiento para el diagnóstico de la calidad de los datos. Aplicación en la empresa productora de envases de tabaco ARCA". Trabajo de Diploma (Ingeniero Industrial). Ciudad Habana: CUJAE, 2008.
9. HEREDIA, Jobany y VILALTA, José. "Procedimiento para el Diagnóstico de la Calidad de los Datos". En: *Sexto Taller de Calidad* (Universidad de la Habana: [s.n.], 2008)



Instituto Superior Politécnico
José Antonio Echeverría
cujae