

Diseño de Bases de Datos de Calidad



MARIO G. PIATTINI VELTHUIS
Universidad de CASTILLA LA MANCHA
E-mail: mpiattin@inf-crudm.es



ISMAEL CABALLERO
Universidad de CASTILLA LA MANCHA
E-mail: mpiattin@inf-crudm.es

A **BSTRACT:**
En un artículo publicado en el N° 2 de Cuore, se planteaba la necesidad de dotar de calidad a las bases de datos para la obtención de mejores resultados en el uso y explotación de dichas bases de datos. Presentamos en esta ocasión una Metodología para la Creación de Bases de Datos de Calidad.

■ 1 INTRODUCCIÓN

Genero et al. (1999) resaltan la necesidad de que los datos existentes en las bases de datos organizacionales presenten ciertas características que permitan obtener mayor rendimiento en su explotación. Esta circunstancia está causando cambios en las formas de hacer negocios: ha provocado que las empresas se planteen manejar la información como un activo más, como parte indispensable de su día a día, como la forma de resolver problemas y situaciones a nivel operativo. Pero es más, también supone la base para la creación de un conocimiento que permita planificar tácticas y estrategias de mercado (medio y largo plazo) que conlleven beneficios para la organización.

Las organizaciones, para el desarrollo de su actividad, necesitan recopilar, almacenar y procesar una gran cantidad de información proveniente de múltiples fuentes. Esta situación es debida a, entre otros, los siguientes motivos:

- La facilidad y el bajo coste con el que se pueden capturar datos gracias a la

mejora y difusión de las tecnologías de entrada de datos.

- La redundancia incontrolada de datos.
- La existencia de grandes cantidades de datos "históricos" caducados, que ya no sirven para realizar ningún proceso ni obtener ningún tipo de información relevante. Orr (1998) comenta que a los datos les puede pasar lo que a los músculos: cuando no se usan se atrofian y pierden fuerza y sentido.

Además de esta situación de redundancia de datos, es posible encontrarse con otras circunstancias tales como que las operaciones que consultan o actualizan los datos estén mal diseñadas o que el propio Sistema de Gestión de Bases de Datos (SGBD) provoque errores.

Genero et al. (1999) basaron su estudio en la calidad de los modelos conceptuales de datos. En este artículo el objetivo será mejorar la calidad de los datos que se almacenarán en las bases de datos. Para ello nos centraremos en la idea de que los datos deben presentar ciertas características que permitan filtrarlos o etiquetarlos de alguna forma para poder gestionarlos mejor. Estas características, que se denominan **dimensiones de calidad**, deben ser elegidas por el diseñador de la base de datos en función de los requisitos que el usuario haya especificado y en cierto modo de la naturaleza del problema que se está tratando.

■ 2 DISEÑO DE BASES DE DATOS Y TRATAMIENTO DE LA CALIDAD DE LOS DATOS

Como sabemos, cuando se diseña una base de datos se comienza con el análisis, que consiste en crear un esquema conceptual a partir de la semántica de la Realidad (Universo del Discurso) que se quiere modelar (De Miguel et al., 1999). Después se traduce dicho esquema a un Modelo Lógico (por ejemplo al Relacional), y se normaliza el resultado. Por último se implementa en un SGBD Relacional que se ajuste a las necesidades de los usuarios.

En todo el proceso anterior, la calidad que manifiestamente se pretende conseguir es aquella que es intrínseca a los datos, es decir, aquella que se consigue cuando se tratan de evitar defectos importantes que influyen de forma crítica en la aplicación, tales como un diseño poco acorde con el Universo del Discurso que se modela, o la posibilidad de que haya valores prohibidos o incoherentes. Los únicos medios que se tienen son la acumulación de experiencias de diseños anteriores y procurar una buena formación para los diseñadores, haciendo que utilicen correctamente las herramientas y técnicas disponibles (Modelo Entidad / Interrelación, Teoría de Normalización, modelado de integridades, otras propias del SGBD tales como la concurrencia, la recuperación, integridad y control de la seguridad).

Orman et al. (1994) dan tres líneas sobre cómo se puede mejorar la calidad intrínseca de la base de datos:

1. Construir modelos semánticos más ricos que reflejen mejor la realidad. Estos modelos deberían:

- a. No permitir datos que no tengan una semántica correcta.
- b. Requerir la adquisición de datos que no han sido introducidos para garantizar la completación.
- c. Cuantificar y almacenar la calidad de los datos, dentro del modelo de datos.

2. Reforzar las bases de datos con un

mayor número de restricciones para identificar y discriminar datos con problemas y enlazar dichos datos con las aplicaciones apropiadas.

3. Restringir el uso de los datos a procesos predefinidos, no permitiendo que sean modificados por cualquier proceso para que no puedan ser borrados accidentalmente.

De todos los puntos enunciados es preciso resaltar la idea de introducir los términos relativos a las dimensiones de la calidad en el modelo de datos. Para ello hay que buscar una forma que permita almacenar datos sobre la calidad de los datos. Llegados a este punto se tienen que resolver dos problemas:

- qué información de calidad se debe almacenar de los datos y
- cómo almacenarla, para después poder manipularla con el fin de obtener conocimiento de ella.

Sobre el primer punto, se trata de decidir cuáles son las dimensiones de calidad que se consideran necesario resaltar. Una vez decididas dichas dimensiones de calidad es necesario especificar el rango de valores que puede tomar la magnitud que representa la dimensión y las operaciones que con ellos se pueden hacer. Este rango de valores puede ser:

- continuo, si la dimensión se refiere a una característica objetiva medible del concepto que estamos tratando, o
- discreto si la dimensión se refiere a una característica subjetiva cuyos valores pueden ser expresados mediante indicadores lingüísticos, tales como nada, poco, mucho, ... o si se refiere a una característica objetiva cuyos posibles valores se han discretizado.

A la hora de definir las métricas se debe tener en cuenta si la medición la va a hacer una persona o un dispositivo electrónico (por ejemplo, un sensor). La diferencia va a radicar en la rapidez con la que se rellenan los campos de datos y se hacen las mediciones, aunque si bien es cierto que ambos pueden cometer errores al decidir el valor de las medidas objetivas, se reserva para las personas la capacidad de juzgar subjetivamente un determinado valor.

Con respecto a cómo guardar la información, la literatura ofrece distintas soluciones, según el modelo que se utilice: si elegimos el modelo Entidad / Interrelación

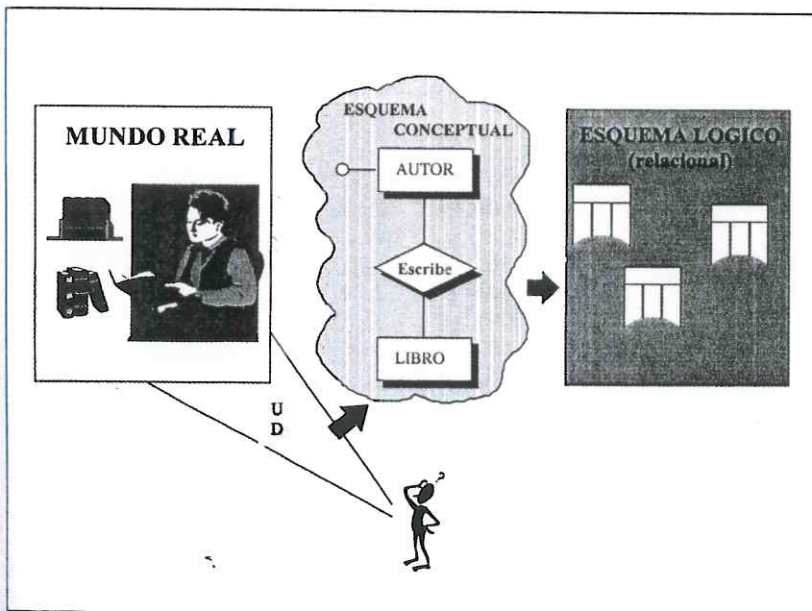


FIGURA 1 • Diseño Clásico de Bases de Datos

para modelar el Universo del Discurso, la idea sugerida en la literatura (Wang et al., 1992; Wang et al., 1993 y Storey y Wang, 1994) es la de considerar las dimensiones de la calidad como entidades. Con ello conseguiremos introducir los datos referentes a la calidad de datos en nuestro diseño como si fuesen datos normales. Por otro lado se podría considerar el modelo relacional y añadir la información a nivel de tupla o a nivel de atributo. Wang et al. (1992) deciden marcar datos a nivel de atributo porque se pueden tener valores distintos para dos atributos distintos de la misma tupla.

■ 3 METODOLOGÍA DE DISEÑO DE BASES DE DATOS DE CALIDAD

Con lo visto anteriormente, se plantea la necesidad de modificar la metodología tradicional de diseño de bases de datos para dar cabida al modelado de calidad de datos.

Una de las novedades consiste en recoger, ya no sólo los requisitos que los usuarios quieren que tenga su aplicación, sino además, los referentes a la calidad de los datos. Wang et al. (1994) hace la siguiente clasificación de los requisitos:

- **Requisitos del producto.** Los requisitos del producto corresponden a los tradicionales requisitos de la aplicación.
- **Requisitos de calidad del producto.** Los requisitos de calidad del producto corresponden a los aspectos de calidad del producto. Dichos requisitos pueden tener en cuenta las fases de entrada, procesamiento y salida asociadas a la realización de un producto. Tales requisitos pueden incluir la comparación del producto final con un conjunto de características de la especificación del diseño.
- **Requisitos de calidad de datos.** Los requisitos de calidad de datos se corresponden a la calidad de los datos en la base de datos que representan tanto los requisitos del producto como los de la calidad del producto. También está aceptado que la calidad de los datos es un concepto multidimensional que abarca dimensiones tales como precisión, compleción, oportunidad y consistencia.

Presentamos a continuación una metodología de desarrollo de Bases de Datos con calidad. Otros antecedentes en la literatura son Wang et al. (1992) y Wang et al. (1994). La diferencia entre ambos es que el segundo incorpora al diseño la recogida de los requisitos de calidad del producto. La siguiente Metodología está basada en las propuestas de De Miguel et al. (1999) y las de Wang et al. (1992).

3.1. Presentación de la Metodología.

Para diseñar una base de datos con calidad vamos a tener en cuenta los siguientes pasos:

1. **Análisis.** El análisis consiste en recoger los requisitos de los usuarios, tanto los relativos a los del producto como a los de calidad de datos. Para esta labor se pueden utilizar cualquiera de las técnicas y herramientas recogidas por Piattini et al. (1996).

Una vez analizados los requisitos del usuario para la aplicación, es necesario concentrarse en los requisitos de la calidad de datos. Para ello se tienen que ver cuáles van a ser las dimensiones respecto a las cuales se van a valorar la calidad de los datos. La elección de dichas dimensiones debería estar supeditada a la naturaleza del producto que se está diseñando, y ser representativas para que los resultados que de ellas se deriven sean lo suficientemente significativos. Para un estudio exhaustivo del significado de las dimensiones de calidad se pueden consultar Ballou et al. (1994), Huang et al. (1999), Redman (1996), Wand y Wang (1994a), Wang et al. (1993), ...

Merece la pena resaltar el modo en que Wang y Wand (1993) recogen las dimensiones de la calidad: empiezan por extraer las **subjetivas (parámetros de calidad)** a partir de los requisitos de los usuarios y a partir de éstas, consideran aquellas que podrían tener medidas **objetivas (indicadores de calidad)**; el objetivo es lograr que la calidad de la base de datos repose en hechos objetivos con medidas objetivas libre de cualquier connotación.

El análisis de requisitos debería también permitir obtener información tanto para las operaciones que los usuarios vayan a hacer con las bases de datos, como para las que vayan a hacer con los datos referentes a la calidad de datos. Con relación a la de calidad de los datos, algunos ejemplos de operaciones podrían ser consultar cuántos y cuáles atributos tienen una calificación de

muy buena en el campo de exactitud, o cuántos valores están caducados, ...

Llegados a este punto se debería tener claro cómo se van a representar en nuestro esquema conceptual los requisitos de calidad de datos: si a nivel de atributo en el modelo Entidad / Interrelación (Wang et al., 1992) o como si fueran entidades (Wang et al., 1994). Dadas las ventajas de poder calificar individualmente un atributo en una determinada dimensión, parece mejor opción la primera. Por tanto, para diseñar la calidad de los datos consideraremos que un atributo va a tener a su vez atributos que serán los correspondientes a las distintas dimensiones de calidad que elijamos para ese atributo en concreto.

Según el razonamiento del párrafo anterior, comprendemos que se hace necesario extender el modelo Entidad / Interrelación para que dé cabida a los atributos de calidad de los atributos "normales". El diseño del esquema conceptual tendrá los siguientes elementos:

- Se tomarán como entidades y atributos de las entidades los que se obtengan al analizar los requisitos del usuario. También habrá, como siempre, que definir las cardinalidades y los dominios de los atributos.
- Como atributos de los atributos se tomarán aquellos que se deriven de los requisitos de calidad recogidos.
- Las interrelaciones entre las entidades, así como sus tipos de correspondencias serán determinados como es habitual por los requisitos de los usuarios.

Para diferenciar a los atributos "normales" de los atributos de calidad en el esquema conceptual, vamos a rellenar su símbolo con líneas diagonales de la forma que aparece en la página siguiente.

"El análisis consiste en recoger los requisitos de los usuarios, tanto los relativos a los del producto como a los de calidad de datos"

BASES DE DATOS

Por simplicidad y por evitar situaciones problemáticas se debería evitar que el atributo que es clave primaria tenga atributos de calidad.

Establezcamos los primeros pasos de la metodología:

Paso 1: Recoger los requisitos de producto del usuario, para datos y operaciones con los datos.

Paso 2: Analizar y escoger las dimensiones de calidad y los atributos a los que calificarán. La elección de las dimensiones debe tener en cuenta los posibles rangos de valores, sus métodos de recogida y las operaciones que se harán con ellos.

Paso 3: Construir el esquema conceptual, utilizando una versión ampliada que permita atributos de calidad para los atributos "normales".

2. Diseño. El diseño consistirá en la traducción del esquema conceptual obtenido anteriormente al correspondiente esquema lógico.

Una vez que tengamos el esquema lógico, habrá que diseñar las restricciones semánticas oportunas y normalizar hasta conseguir el esquema más óptimo para el problema.

Es el momento de decidir la forma en la que serán tratados los atributos de calidad. Existen dos posibilidades:

- Algunas bases de datos objetos-relacionales, permiten almacenar información en tablas que no son planas; esto se consigue definiendo atributos cuyo tipo de dato sea una colección de datos. La limitación puede estar en que dicha colección de datos sea homogénea, es decir todos los elementos de la colección son del mismo tipo de datos. De esta forma el atributo que tenga atributos de calidad se definirá con tipo de dato colección y tendrá tantos elementos como dimensiones con las que vaya a ser calificado más un elemento más para guardar el propio valor del atributo "normal". Así tendríamos tablas como la de la figura 3.

- En bases de datos relacionales "tradicionales", podríamos utilizar subrogados. Estos subrogados, siendo únicos y sin admitir valores nulos, actúan como si fuesen punteros a otra tabla donde se va a

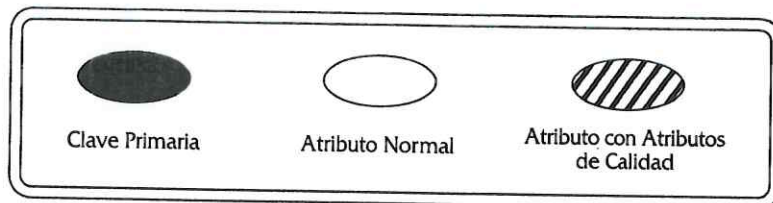


FIGURA 2 • Tipos de atributos

NOTA SELECTIVIDAD				NOTA MEDIA CARRERA		
Alumno	Valor Nota	Fecha	Fuente	Valor Nota	Fecha	Fuente
William Smith	8	30/10/90	MEC	7	30/07/95	ESI
Gene Hackman	9	30/06/90	MEC	8	29/07/94	ESI

FIGURA 3 • Atributos con calidad de tipo de dato colección de elementos

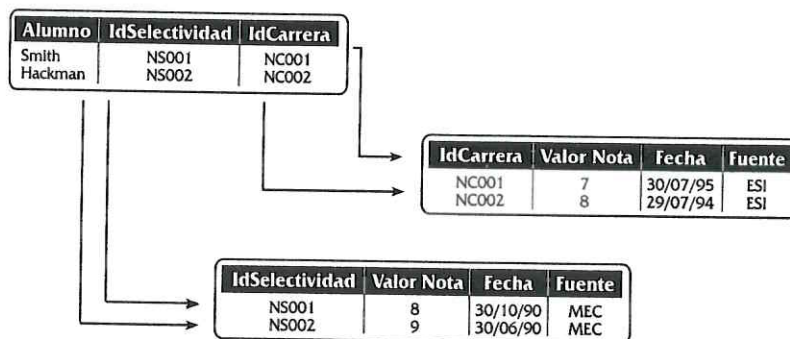


FIGURA 4 • Diseño con subrogados

almacenar el valor del atributo "normal" y los valores de cada una de las dimensiones que califican a dicho atributo. Así la tabla anterior se representaría como se indica en la figura 4.

Paso 4: Transformar el esquema conceptual al lógico. Este paso debería incluir la normalización del esquema lógico y el diseño de las restricciones, tanto para los atributos como para los atributos de los atributos. Además hay que decidir la forma

en la que almacenaremos los atributos de calidad.

3. Implementación. La implementación consistirá en la creación del Esquema Físico en un SGBD determinado según las necesidades del problema.

Paso 5: Implementación del Modelo Lógico en un SGBD.

En la figura 5, que aparece en la página siguiente, se resume la metodología propuesta.

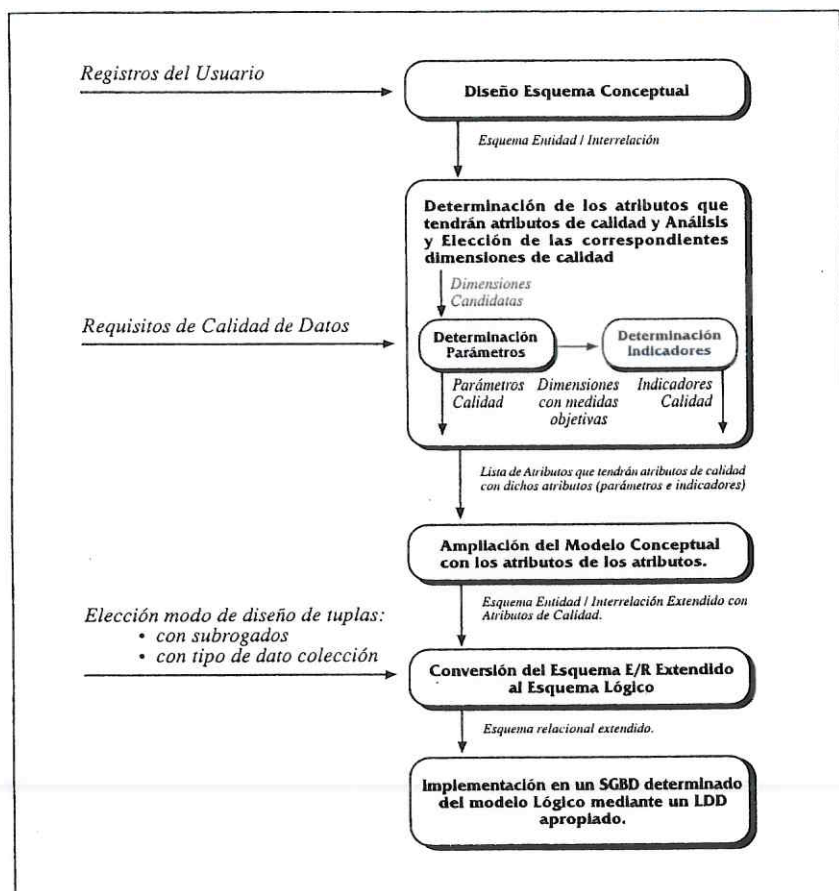


FIGURA 5 • Metodología de Diseño de Bases de Datos con Calidad

3.2. Ejemplo

Se trata de desarrollar un Sistema de Inventario Justo a Tiempo (Just In Time - JIT) que requiere que los inputs y componentes necesarios para la producción sean suministrados al proceso de transformación justo en el momento en el que son necesarios, ni antes ni después, de tal forma que los inventarios de inputs, puedan mantenerse en un mínimo. Los componentes para la fabricación de un producto son puestos en contenedores, y a medida que se van vaciando, los contenedores vacíos son enviados de nuevo al proveedor, con un pedido escrito en una tarjeta pidiendo más componentes para poder seguir fabricando más productos. Todos los datos de los inputs deben mantenerse en una base de datos, que debe estar constantemente actualizada.

Para ello vamos a seguir los pasos de la metodología:

Paso 1: Recogida de los requisitos de los usuarios

Los requisitos de los usuarios se pueden resumir en:

- 1• Necesitamos guardar los datos de los Proveedores, entre los que están su Nombre, Dirección, CIF, y la Valoración que los Directivos tienen de dicho Proveedor.
- 2• Necesitamos guardar los datos de los Productos, entre los que están su Marca, Modelo, el Código asignado a dicho Producto, una valoración sobre el Grado de Utilización de dicho producto en nuestros procesos, la Canti-

dad que tenemos en los almacenes y a qué Proveedor fue comprado.

3• Necesitamos guardar las ofertas que hacen los proveedores, que tendrán un Código de la Oferta, Descripción de la Oferta, un Precio Global. También deberíamos guardar cuánto costó cada producto y la fecha en que fue comprado.

Paso 2: Análisis y Recogida de los requisitos de calidad de datos de los usuarios.

1• Sobre los Proveedores:

a. Se quiere que la valoración de los proveedores pueda ser hecha por votación de varios directivos de la empresa, e interesa conocer el grado de participación en la votación, ya que cuanto más gente haya votado (se supone que cada votante sólo puede emitir un voto) más fiable será la valoración. En principio, la votación realizada por cada una de las personas que votó es algo subjetivo que dependerá de los criterios propios que cada persona tenga. Por el contrario, el índice de votación es un valor objetivo, ya que podemos contar todas las personas que han votado. Este va a ser por tanto un indicador de calidad, porque va a tener un valor objetivo, medible y demostrable.

2• Sobre las Ofertas:

a. Es necesario conocer el grado de interpretabilidad de la oferta, pues no sería agradable que después de hacer un pedido llegaran dificultades por no haber sabido interpretar la oferta. Esta dimensión, que en principio es subjetiva, es susceptible de un análisis más profundo para intentar descubrir o dividir en aspectos objetivos, tales como la definición de un formato para el documento, el símbolo exacto para la moneda, ... Por tanto, y si se asume que el grado de interpretabilidad puede tomar valores tales como poco, mucho, nada, ... entonces se puede considerar que se tendrá un parámetro de calidad.

b. Es interesante saber en todo momento la vigencia y exactitud de la oferta de cada uno de los proveedores. La vigencia va a ayudar a determinar si los precios y caducidad de los productos que constituyen la oferta son todavía válidos o no. Estas medidas serán obje-

BASES DE DATOS

tivas, por lo que se tendrá un indicador. Si se considera la exactitud como el grado de estimación de la concordancia entre el contenido de la oferta y lo que en realidad se pidió o se ofertó, entonces se tendrá un parámetro de calidad. Por otro lado, se puede objetivar la exactitud considerándola como la relación entre el número de productos que se espera que aparezcan (por ejemplo los que se les pidieron a los distribuidores) y los que en realidad aparecen, se obtiene otro indicador.

3• Sobre los Productos:

• a. Interesa conocer el grado de exactitud de la cantidad de productos y si coincide exactamente lo que se cree que se tiene con lo que realmente se tiene en el momento actual. Esta dimensión tiene un análisis similar al anterior. Sin embargo se la puede considerar objetiva, ya que en cualquier momento es posible contar o medir la cantidad de productos almacenados y compararse con la que indica la base de datos. Se tiene por tanto otro indicador.

• b. También interesa conocer la fiabilidad del grado de utilización de los productos, porque si se compran productos y no se utilizan o se sobreutilizan, puede que alguno de los procesos esté mal especificado o que exista algún fallo en las máquinas. Para medir la fiabilidad se puede hacer un estudio de lo que se necesita en los procesos de fabricación y de lo que en realidad se usa. Para ello se compara en el análisis de los procesos con lo que la base de datos dice que se tiene. De aquí se obtiene otro indicador.

Paso 3: Construcción del modelo conceptual

De la especificación de los requisitos se desprenden la existencia de las siguientes entidades:

■ **Proveedor:** que tendrán los siguientes atributos: CIF (como Clave Primaria), Nombre, Dirección, Valoración Directivos.

■ **Producto:** CódigoProducto (como Clave Primaria), Marca, Modelo, GradoUtilización y Cantidad.

■ **Oferta:** CódigoOferta (como Clave Primaria), Descripción y Precio.

Para satisfacer la especificación de los requisitos de calidad de datos, los siguientes atributos "normales" van a tener estos atributos de calidad, representado a las distintas dimensiones de calidad que se han analizado anteriormente:

■ **Proveedor.** ValoraciónDirectivos tendrá el atributo ÍndiceVotación.

■ **Oferta.** Descripción tendrá los atributos Interpretabilidad, Actualidad y Exactitud.

■ **Producto.** Cantidad estará calificado por los atributos Precisión y Actualidad.

■ **Producto.** GradoUtilización tendrá el atributo Fiabilidad.

Todos estos datos se recogen en el esquema conceptual de la figura 6.

"Se puede objetivar la exactitud considerándola como la relación entre el número de productos que se espera que aparezcan y los que en realidad aparecen"

Nos quedaría por definir cuáles son los rangos de valores de cada uno de los atributos:

■ Proveedor

- CIF, Nombre y Dirección pueden ser cadenas de texto.
- ValoraciónDirectivos puede tomar valores reales entre 0 y 10
- ÍndiceVotación pueden representarse en tanto por ciento

■ Producto:

- CódigoProducto, Marca, Modelo pueden ser cadenas de texto.

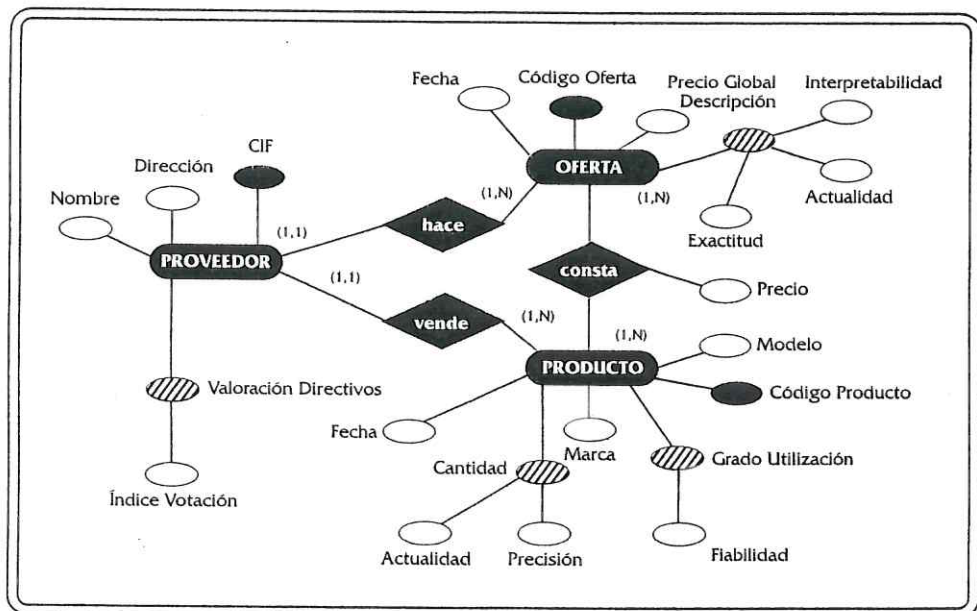


FIGURA 6 • Esquema conceptual del ejemplo

BASES DE DATOS

- Cantidad será un valor en una determinada magnitud: unidades, kilos, metros, ...

- Actualidad puede indicar los minutos de desfase que hay entre el valor almacenado y el momento en el que en realidad había esa cantidad.

- Precisión puede medirse en tanto por ciento.

■ **Oferta:**

- Código Oferta es una cadena de texto

- Precio será un entero largo.

- Descripción es una cadena de texto
 - Interpretabilidad se puede medir con indicadores lingüísticos tales como poca, normal o total

- Exactitud se puede en tanto por

ciento.

- Actualidad se puede medir en días desde que se hizo la oferta.

Paso 4: Transformación del modelo conceptual al lógico.

El esquema lógico resultante de la traducción del conceptual y del uso de subrogados da como resultado el siguiente esquema relacional de la figura 7

Paso 5: Implementación del Modelo Lógico en un SGBD

En el lenguaje apropiado de definición de datos (por ejemplo en SQL) se crearían todas las tablas, y con el correspondiente lenguaje de manipulación de datos se implementaría todas las operaciones de consultas, actualización y disparadores

necesarios para la gestión correcta de la Bases de Datos.■

Este trabajo se encuadra dentro del proyecto CALIDAT, que desarrolla la empresa Cronos Ibérica, S.A. en colaboración con la Universidad de Castilla-La Mancha con financiación de la Consejería de Educación y Cultura de la Comunidad de Madrid (Ref: 09/0013/1999).

“Es necesario que las organizaciones sometan sus bases de datos a análisis dentro de marcos de referencias para la valoración de la madurez de éstos”

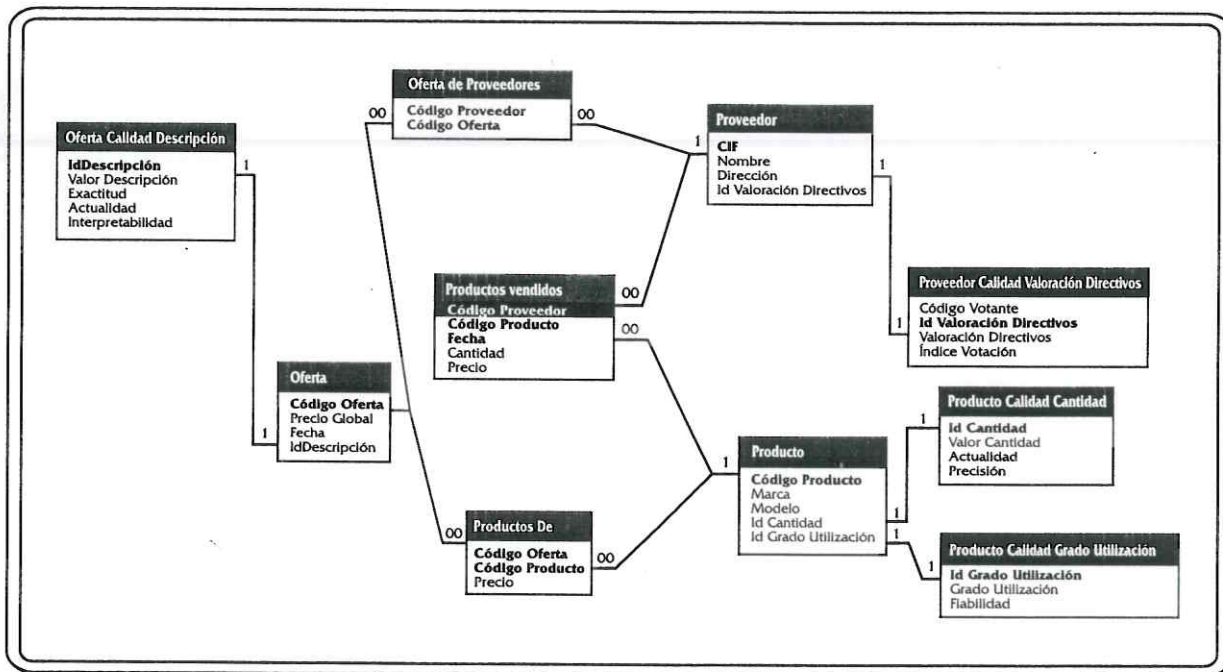


FIGURA 7• Esquema relacional del ejemplo

CONCLUSIONES

Los datos se están convirtiendo en el activo más importante de las empresas, pues en función de la información que se pueda obtener de esos datos, bien por analistas de mercados, bien por métodos automáticos (data mining), las organizaciones pueden obtener importantes beneficios. Pero para tomar este tipo de decisiones, necesitamos que la base, los datos, estén en condiciones óptimas para poder fiarnos de ellos. Esta condición se manifiesta buscando calidad en esos datos. Esta calidad se valora en unas determinadas dimensiones. La elección de estas dimensiones depende de la naturaleza de la aplicación y de los requisitos de calidad de datos de los usuarios.

Para mantener bases de datos con calidad de datos, además de la calidad intrínseca que pudiera tener dicha base de datos, hay que integrar los datos propiamente dichos con los datos relativos a la calidad de los mismos. Para ello hemos presentado una metodología de diseño de bases de datos con dimensiones de calidad. Con estos datos de calidad, que acabarán en tablas relacionales, podremos usar las operaciones normales de consultas y actualización sobre ellos para extraer conocimiento de la bases de datos. Este conocimiento va a permitir detectar defectos en los procesos de recogida, almacenamiento y procesado de datos y obtener conclusiones e indicadores.

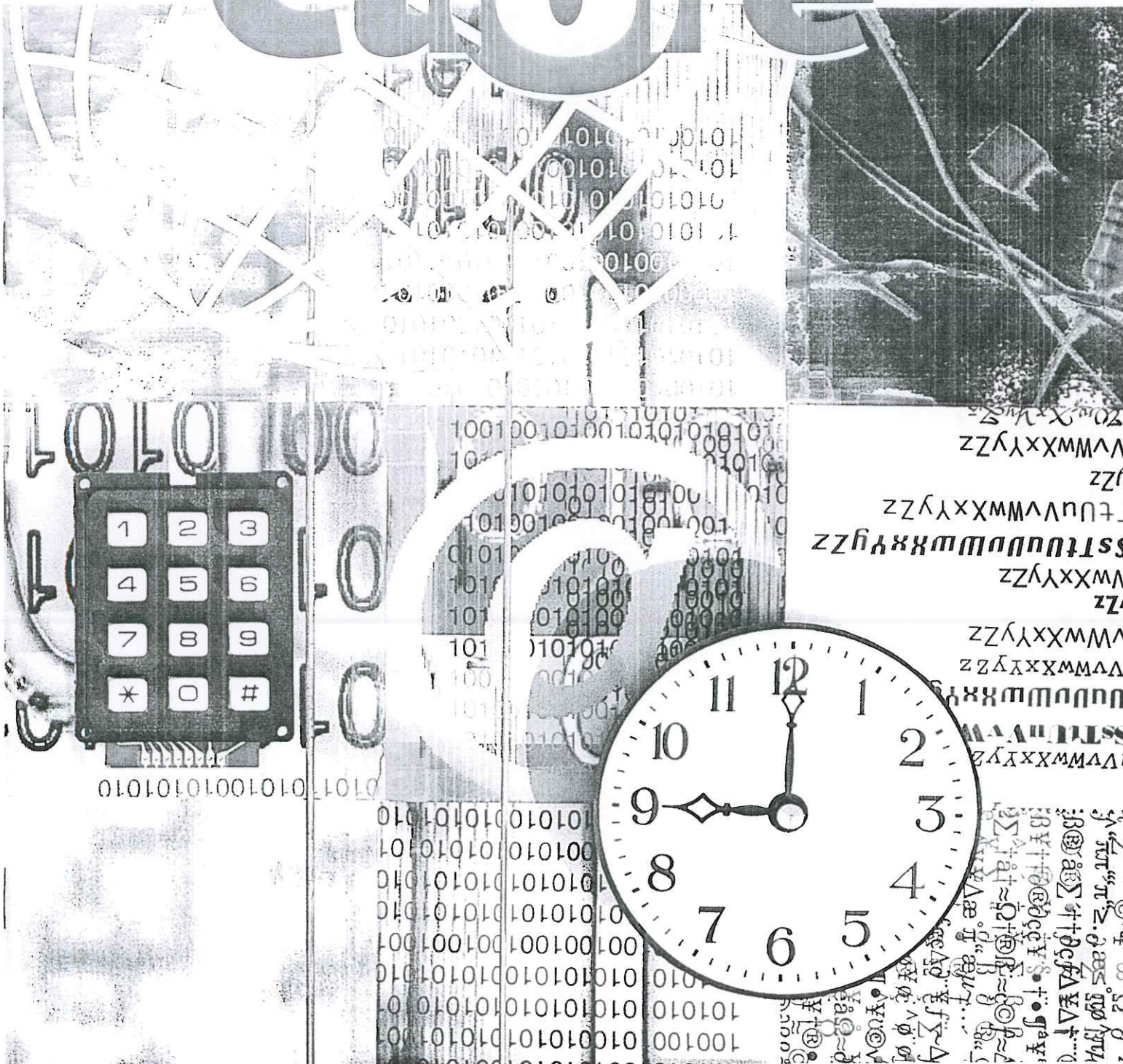
La metodología presentada no es suficiente por sí misma para garantizar la calidad de los datos existentes en las bases de datos de las organizaciones. Para ello sería necesario que las organizaciones sometieran sus bases de datos a análisis dentro de marcos de referencias para la valoración de la madurez de los datos al estilo de los modelos de valoración de procesos existentes, tales como SPICE, CMM, BootStrap,...

BIBLIOGRAFÍA

- Ballou, D., Wang, R., Pazer, H., Tayi, G.K., (1994). *Modeling Information Manufacturing Systems to Determine Information Product Quality*. *Management Science*, 44(4), 1998 pp.462-484
- De Miguel A., Piattini, M. Y Marcos E. (1999) *Diseño de Bases de Datos Relacionales*. Ed. Ra-ma
- English, L. (1999) *Improving Data Warehouses and Business Information Quality*. John Wiley & Sons.
- Genero, M., Calero, C. Piattini, M. (1999) *Asegurar la calidad de las Bases de Datos... un reto para el año 2000*. *Cuore N°2*, Julio 1999, pp 28-35
- Huang, K.T., Lee, Y., Wang, R. (1999) *Quality Information and Knowledge*. Prentice Hall, Upper Saddle River
- ISO (1998) *Software Product Evaluation – Quality Characteristics and Guidelines for their Use*. ISO/IEC Standard 9126, Geneva
- Madnick, S. y Wang, R. (1992). *Introduction to TDQM research*. (TDQM – 92- 01) *Total Data Quality Management (TDQM) Research Program*, MIT Sloan School of Management, Cambridge, MA. Disponible en <http://web.mit.edu/tdqm/www/papers/92/92-01.html>
- Orman, L., Storey, V. Y Wang, R. (1994) *Systems Approaches to Improving Data Quality*(TDQM-94-05, August 1994. Disponible en <http://web.mit.edu/tdqm/www/papers/94/94-05.html>
- Orr, K.(1998) *Data Quality and System Theory*. *Communication of the ACM*, 41 (2), pp 66-71.
- Piattini, M., Calvo-Manzano, J. A., Cervera, J., Fernández, L. (1996) *Análisis y Diseño de Aplicaciones Informáticas de Gestión*. Ra-Ma
- Pierce, E. (1997) *Modeling Database Error Rates*. *Web de Data Quality*. Septiembre 1997. www.dataquality.com
- Redman, T.C.(1996) *Data Quality for the Information Age*. Artech House Publishers, Boston.
- Storey, V., Wang, R. (1994) *Modeling Quality Requirements in Conceptual Database Design*. (TDQM-94- 02) <http://web.mit.edu/tdqm/www/papers/94/94-02.html>
- Wand, Y., Wang, R. (1994) *Anchoring Data Quality Dimensions in Ontological Foundations*. *Communications of the ACM (CACM)*, 39, (11), pp 86-95
- Wang, R., Kon, H. y Madnick S. (1993) *Data Quality Requirements Analysis and Modeling*. Published in the Ninth International Conference of Data Engineering Vienna, Austria. pp 670 – 677
- Wang, R., Reddy, M.P., Kon, H. (1992) *Toward Quality data: An Attribute-based approach*. *Journal of Decision Support Systems (DSS)* 13 pp 349-372.
- Wang, R., Strong, D. y Guarascio, L., (1994a) *Data Consumers' Perspectives of Data Quality*. *Information Systems Research (ISR)* (TDQM-94-01). Disponible en <http://web.mit.edu/tdqm/www/papers/94/94-01.html>
- Wang, R., Strong, D., y Guarascio, L. (1994b) *Beyond Accuracy: What Data Quality Means to Data Consumer*. (TDQM-94-10). Disponible eb <http://web.mit.edu/tdqm/www/papers/94/94-10.html>
- Willshire, M. J. *A process for Improving Data Quality*. *Web de Data Quality*. Septiembre 1997. www.dataquality.com

Círculo de Usuarios Oracle de España

Cuore



- Diseño de Bases de Datos de Calidad.
- La firma electrónica.
- ¿Está listo para la crisis del éxito?.
- Oracle Internet File System.