

# I+D Computación

Una publicación de la Academia de Posgrado de Ciencias Computacionales



Julio de 2002, Vol. 1 No.1

ISSN 1665-238X

# Calidad de Bases de Datos

Mario Piattini, Marcela Genero, Coral Calero  
Grupo de Investigación Alarcos, E.S. Informática,  
Paseo de la Universidad, 4, 13071 Ciudad Real (España)  
{Mario.Piattini, Marcela.Genero, Coral.Calero}@uclm.es

**Abstract.**– With this paper we begin a series in which the main work recently published in the area of database quality is presented, taking into account from data model quality up to the quality of the data themselves.

**Keywords:** database quality, conceptual modeling, databases.

## 1. INTRODUCCIÓN

La calidad es un tema que se ha impuesto en la actualidad no sólo en los ámbitos económicos y organizativos sino también en el mundo de las bases de datos, que se han convertido en uno de los componentes más importantes de los sistemas de información. En un mercado cada vez más competitivo y globalizado, la información hoy en día, juega un papel fundamental en las empresas, y la calidad de dicha información tiene una gran influencia sobre la toma de decisiones que permite a los directivos plantear diferentes estrategias para alcanzar una mayor competitividad en el mercado.

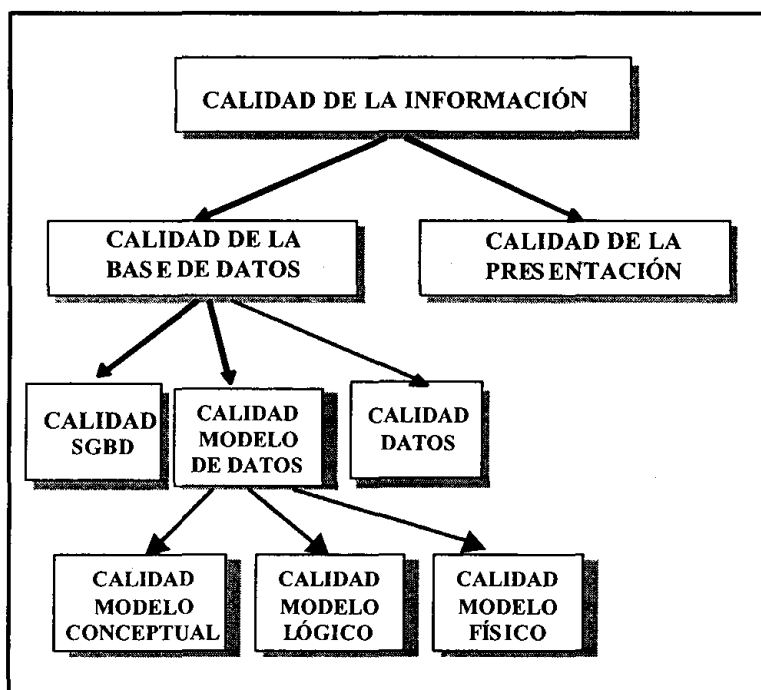
Debido a la creciente complejidad de los sistemas de información, es necesaria una continua atención en la evaluación de la calidad de las bases de datos a lo largo del proceso de desarrollo, para tener conocimiento de la calidad de los sistemas. El compromiso de la calidad en el desarrollo de cualquier producto de software es esencial, no solo para satisfacer a los clientes sino también para mejorar la productividad del proceso.

Existe un rango desconcertante de definiciones formales e informales con respecto al concepto de calidad [20]. La mayoría de las definiciones han surgido de ámbitos económicos y de fabricación. No debemos olvidar que la calidad es un concepto relativo, dado que la importancia de los diferentes factores que influyen en la calidad de un producto varía entre los diseñadores a lo largo del tiempo [16]. Por consiguiente es importante que cualquier método de evaluación de la calidad reconozca estas diferencias de puntos de vista, y permita que los diferentes aspectos tengan diferentes pesos de acuerdo al contexto o situación que se esté considerando.

La mayoría de los trabajos realizados con respecto a la calidad se refieren a la calidad del software. La calidad de las bases de datos es un tema que ha sido descuidado en el ámbito de la investigación. Incluso en el diseño tradicional de las bases de datos, los aspectos referidos a la calidad no se consideran explícitamente. Aunque el trabajo con las bases de datos tradicionalmente no ha estado centrado en temas referidos al manejo de la calidad en sí misma, muchas de las herramientas desarrolladas le han dado importancia al manejo de la calidad de los datos. En efecto, la investigación ha estado dirigida a temas referidos a cómo prevenir las inconsistencias de los datos (restricciones de

integridad y teoría de la normalización) y a cómo prevenir la corrupción en los datos -manejo de transacciones [42].

Nosotros opinamos que en la calidad de las bases de datos se tienen que considerar diferentes aspectos [31]. La calidad de la información de una base de datos viene determinada por la calidad tanto del sistema de bases de datos como la calidad de la presentación de los datos (véase la figura 1). De hecho, es muy importante que los datos de la base reflejen correctamente el mundo real, pero es también muy importante que los datos sean interpretados correctamente. En la calidad de la base de datos, se deben considerar tres aspectos: la calidad del SGBD (Sistema Gestor de Base de Datos) relacional, activo, objeto-relacional, orientado a objetos o multidimensional de que se disponga, la calidad del modelo de datos (tanto conceptual, lógico como físico) y la calidad de los propios datos contenidos en la base.



**Figura 1.** Calidad de la información y la base de datos

Para asegurar la calidad del SGBD, podemos usar un estándar internacional como el ISO 9126 [21], o alguno de los estudios comparativos de productos existentes. Este tipo de calidad debe ser asegurada en la etapa de análisis y selección de los productos dentro del ciclo de vida de una base de datos [12].

En este artículo y en el próximo abordaremos la calidad en el modelado conceptual, mientras que en los siguientes trataremos la calidad del modelado lógico y la de los propios datos, la calidad del modelo físico es muy dependiente del propio SGBD utilizado, por lo que no la trataremos.

## 2. CALIDAD EN EL MODELADO CONCEPTUAL

El proceso del modelado conceptual juega un papel relevante [11] ya que proporciona el enlace entre las necesidades del usuario y la solución del software que las satisface; aún cuando representa sólo una porción relativamente pequeña del esfuerzo total de desarrollo de un SI, su impacto en el resultado final es probablemente mucho mayor que el de cualquier otra fase. Por ello, el modelado conceptual se ha convertido en una tarea clave en las primeras etapas del ciclo de vida de los SI. Por lo tanto, centrarse en la calidad del modelado conceptual es un aspecto clave para asegurar el éxito del desarrollo de los SI. Tal y como sugieren la aproximación de la Gestión de Calidad Total (TQM), es más rápido y barato concentrar el esfuerzo durante las primeras fases del desarrollo de un producto, y detectar y corregir defectos lo más pronto posible en el ciclo de vida del producto [13]. Esto significa que la calidad del SI puede estar altamente influenciada por las primeras decisiones tomadas en el desarrollo.

Cuando hablamos del modelado conceptual, debemos distinguir entre la calidad del producto y la calidad del proceso. La calidad del producto está relacionada con las características del modelo conceptual (el producto), mientras que la calidad del proceso está relacionada con cómo se desarrollan los modelos conceptuales (el proceso). Nosotros enfocamos este artículo a la calidad del producto, mientras que los asuntos relacionados con la calidad del proceso están parcialmente tratados en [27] y [26].

Algunos autores han definido la calidad de los modelos conceptuales como una lista de propiedades "deseables" que debe satisfacer el modelo de datos (ver tabla 1) y han propuesto una serie de transformaciones con el objetivo de mejorar la calidad de los mismos.

**Tabla 1.** Propiedades "deseables" para el modelo conceptual de datos

<b>Autores</b>	<b>Propiedades</b>
Batini et al. (1992)	compleción, corrección, minimalidad, expresividad, legibilidad, autoexplicación, extensibilidad y normalidad.
Reingruber M. y Gregory W. (1994)	corrección conceptual, compleción conceptual, corrección sintáctica, compleción sintáctica, conocimiento de la empresa.
Boman et al. (1997)	facilidad de comprensión, corrección semántica, estabilidad, compleción, enfoque conceptual.

Aunque estas listas de propiedades proveen un punto de partida útil para entender y mejorar la calidad en el modelado conceptual, la mayoría no son estructuradas, usan definiciones imprecisas, a menudo se solapan, se mezclan características de la especificación con las del método y del lenguaje de modelado, se presupone la existencia de diseño/implementación, y se presentan objetivos poco realistas o imposibles de alcanzar [25].

Una estrategia más adecuada para abordar la calidad es definir marcos de referencia que organicen y estructuren los conceptos claves y características en el modelado conceptual de datos. Además, se han publicado algunos marcos de referencia que permiten abordar la calidad en el modelado conceptual de una manera más sistemática ([28],[29], [25], [22], [23], [39]). En la figura 2 se presenta el marco de Krogstie et al. [23], en el que se propone distinguir varios tipos de calidad:

- **Calidad sintáctica:** que viene determinada por la correspondencia entre el modelo y el lenguaje, y cuyo objetivo es la corrección sintáctica.
- **Calidad semántica:** (percibida), que comprende tanto la validez (en el sentido de que lo expresado en el modelo es correcto y relevante para el problema) como la compleción (el modelo contiene todas las sentencias acerca del dominio que son correctas y relevantes).
- **Calidad pragmática:** cuyo objetivo es que el modelo sea comprendido.
- **Calidad social:** que persigue distintos tipos de acuerdo tanto en la interpretación del modelo como respecto al conocimiento del dominio.

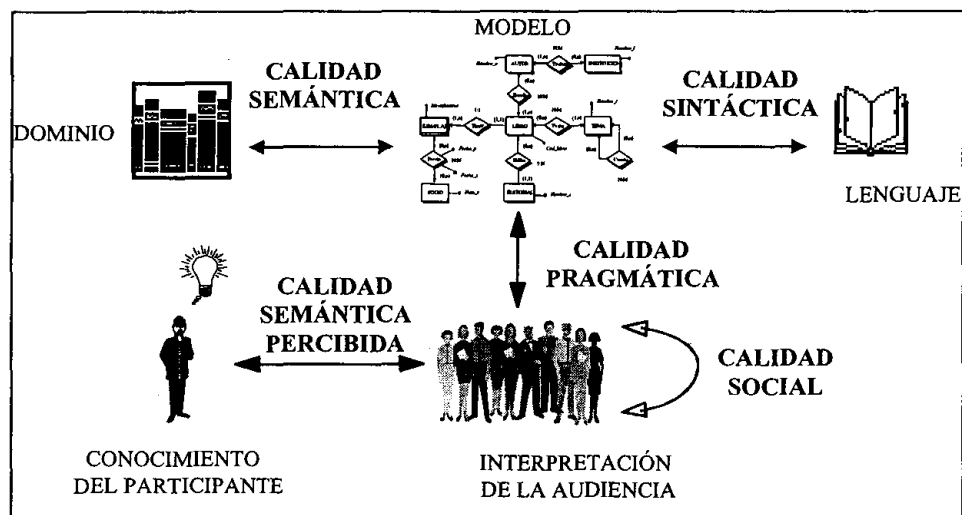


Figura 2. Marco para la calidad del modelado conceptual, Krogstie et al. [23]

Como es natural, consideraremos que un modelo tiene una mayor calidad semántica cuanto mejor sea la correspondencia entre el modelo conceptual que hemos diseñado y el dominio que pretendemos representar. Sin embargo, como señalan los autores de este marco, es imposible establecer o verificar directamente esta correspondencia, ya que para diseñar el modelo se debe acudir al conocimiento que tiene el público sobre el dominio y para verificarlo se debe emplear la interpretación que el público hace del modelo.

Aunque estos marcos son relevantes en el sentido de que contribuyen a un mejor entendimiento sobre el tratamiento de la calidad de los modelos conceptuales, carecen de medidas cuantitativas que permitan evaluar la calidad de los modelos conceptuales de datos de forma objetiva. Resulta evidente que definir sólo “propiedades deseables” no es suficiente para evaluar la calidad, ya que diferentes personas pueden dar diferentes interpretaciones sobre el mismo concepto, por lo que es necesario contar con medidas que permitan evaluar la calidad de los modelos conceptuales de datos de forma cuantitativa y objetiva, evitando así los sesgos en el proceso de evaluación de su calidad [29].

Como sabemos, existen diferentes tipos de modelos conceptuales:

- Modelos conceptuales tradicionales. Como el modelo Entidad-Interrelación (ER), introducido por Chen [10] y después ampliado por Teorey [40], De Miguel et al. [12], Thalheim [41] y otros. Como sabemos el modelo ER proporciona una serie de constructores capaces de describir los

requerimientos de datos de un SI de una manera fácil de entender, independiente del criterio seguido para la gestión y organización de los datos de un SI.

- Modelos conceptuales orientados a objetos (OO). Se usan para representar no sólo los datos, sino también el comportamiento y la funcionalidad del SI que está siendo modelado. Los modelos conceptuales OO abarcan tanto diagramas estáticos (diagramas de clases), como diagramas dinámicos (diagramas de transición de estados, diagramas de actividad, etc.)

Aunque los modelos conceptuales OO son los más apropiados para representar los requerimientos de los SI que se usan hoy en día, el modelo ER continúa siendo muy usado en el mundo del diseño de bases de datos [30]. En el próximo epígrafe resumimos las propuestas más importantes de métricas para modelos ER.

### 3. PROPUESTAS DE MÉTRICAS PARA MODELOS CONCEPTUALES TRADICIONALES

En esta sección presentamos algunas propuestas de métricas para evaluar la calidad de los modelos conceptuales tradicionales, tales como el ER.

#### 3.1. Métricas de Kesh (1995)

Kesh [22] desarrolló un método para asegurar la calidad de los modelos ER. Este autor considera que la calidad en modelos de datos se determina por componentes tanto ontológicos como del comportamiento. Dicho método puede resumirse en tres pasos:

1. Calcular el valor de cada componente ontológico en forma individual, considerando componentes tanto de estructura (que se refieren a las relaciones entre los elementos que forman el modelo) tanto como de contenido (que se refieren a los atributo de las entidades). Los componentes de estructura son adecuación al problema ( $o_1$ ), validez ( $o_2$ ), consistencia ( $o_3$ ) y concisión ( $o_4$ ). Los componentes de contenido son completitud ( $o_5$ ), cohesión ( $o_6$ ) y validez ( $o_7$ ).
2. Combinar los valores de los componentes ontológicos relevantes para cada componente de comportamiento: usabilidad (desde el punto de vista del usuario) ( $s_1$ ), usabilidad (desde el punto de vista del diseñador) ( $s_2$ ), mantenibilidad ( $s_3$ ), precisión ( $s_4$ ) y rendimiento ( $s_5$ ).
3. Combinar los valores de los componentes de comportamiento para calcular el valor de la calidad del modelo.

Así, la calidad del modelo de datos puede calcularse de la siguiente manera:

$$Q = w_1 s_1 + w_2 s_2 + w_3 s_3 + w_4 s_4 + w_5 s_5$$

donde  $w_1, w_2, w_3, w_4$  y  $w_5$  son los pesos de los factores de comportamiento, y  $s_1, s_2, s_3, s_4$  y  $s_5$  son los valores de los factores de comportamiento. El valor de los pesos puede ser determinado o bien por la importancia del factor para la organización o por un acuerdo acerca de los valores (comúnmente no existen).

Los valores  $s_1, s_2, s_3, s_4$  y  $s_5$  se basan en la combinación de valores de los factores ontológicos:

$$\begin{aligned} s_1 &= (o_1 + o_3 + o_4 + o_5) / 4 \\ s_2 &= (o_2 + o_3 + o_5 + o_6 + o_7) / 5 \\ s_3 &= (o_2 + o_4 + o_6) / 3 \end{aligned}$$

$$s_4 = (o_3 + o_5) / 2$$

$$s_5 = (o_4 + o_5) / 2$$

donde  $o_1, \dots, o_7$  son los factores ontológicos.

Los factores ontológicos se basan en valores que los usuarios dan entre 1 y 5, o en fórmulas más complejas, como se muestra a continuación:

- $o_1$  (adecuación del modelo al problema): los usuarios son entrevistados para determinar si el modelo se adecua al problema, y lo puntúan con una escala de 1 a 5.
- $o_2$  (validez del modelo): calculado de forma similar a  $o_1$ , entrevistando a un grupo técnico que no esté directamente involucrado con el proyecto.
- $o_3$  (consistencia del modelo): puede calcularse deduciendo una cierta cantidad de puntos por cada inconsistencia que contenga el modelo ER, de la siguiente manera:  $o_3 = M - D_1$ , siendo  $M =$  máximo puntaje (en este caso 5) y  $D_1$  está basado en el ratio  $R =$  (número de inconsistencias/número de implicaciones). El número de implicaciones es  $4n_1$ , siendo  $n_1$  es el número de relaciones en el modelo E, ya que dos implicaciones son posibles para cada entidad en una relación.
- $o_4$  (concisión del modelo): Si  $n$  es el número de entidades en un modelo ER, el mínimo número de relaciones necesarias es  $(n-1)$ . Cualquier modelo ER que tenga  $(n-1)$  relaciones tendrá un valor del factor  $o_4$  de  $M$  (en este caso 5). El peor caso es cuando todas las entidades están relacionadas entre sí, el número de relaciones en este caso es  $C\binom{n}{2}$ . En este último caso el valor de  $o_4$  es 0. Para un modelo ER con  $n_1$  relaciones el valor de  $o_4$  se calcula de la siguiente manera:

$$o_4 = M((C\binom{n}{2} - n_1) / C\binom{n}{2} - (n-1))$$

- $o_5$  (completitud de contenido): debe chequearse el modelo ER con respecto a las lista de consultas e informes que se desean obtener de la base de datos. Por cada falta que se observe se debe deducir la puntuación de acuerdo a la importancia de dicho informe o consulta.
- $o_6$  (cohesión del contenido): la cohesión para cada entidad es el tamaño del identificador primario (atributo/s clave). Si el identificador primario está formado por un solo atributo, los atributos tienen el mayor grado de cohesión, por lo tanto  $o_{6i}$  toma el máximo valor  $M$ , donde  $i$  es el número de entidad. Si el identificador primario está formado por todos los atributos de la entidad,  $o_{6i} = 0$ . Si  $n_e$  es el número de atributos de la entidad y  $n_p$  el número de atributos del identificador primario, entonces:

$$o_{6i} = M((n_e - n_p) / (n_e - 1)) \text{ y } o_{6i} = \sum o_{6i} / n$$

- $o_7$  (validez del contenido): se le asigna el valor  $M$  si todos los atributos para todas las entidades son válidos. Se le asigna el valor cero si todos los atributos  $n_e$  son considerados inválidos. Si  $n_i$  es el número total de entidades inválidas, entonces  $o_7$  puede calcularse de la siguiente manera:

$$o_7 = M(1 - n_i / \sum n_e)$$

- Si los valores desde  $o_1$  hasta  $o_7$  caen por debajo de los requerimientos organizacionales el modelo ER debe ser modificado, antes de proceder a calcular el valor de calidad.

A causa de la naturaleza de las métricas se requiere que los diseñadores interactúen con otros usuarios para obtener las medidas, de ese modo se incrementa la realimentación de los usuarios y de los directores.

Kesh [22] sugirió que la determinación de la calidad de un modelo ER es subjetiva, por lo tanto el valor que propuso  $Q$  debe considerarse como un indicador de la calidad del modelo ER más que una estimación precisa.

Estas medidas no fueron validadas teóricamente. Después de aplicar este modelo de la evaluación de la calidad a casos del mundo real, Kesh [22] concluyó que su modelo proporcionaba un marco de trabajo útil para analizar y hacer revisiones a los modelos ER. Sin embargo, no explicó cómo funciona realmente su propuesta en la práctica y también sugirió que tanto el modelo como la metodología deberían aplicarse a modelos ER de mayor magnitud.

### 3.2. Métricas de Moody (1998)

Moody [29] ha definido un conjunto de métricas (ver tabla 2) para evaluar algunos factores de calidad de los modelos de datos ([28], [29]). Algunas de ellas son métricas objetivas (p. ej. el número de entidades o el número de atributos) mientras otras son el resultado de una evaluación subjetiva (p. ej. la habilidad de los usuarios para interpretar el modelo correctamente).

**Tabla 2.** Métricas para evaluar la calidad de modelos ER [29]

<b>Factor de calidad</b>	<b>Métricas</b>
Compleción	<ul style="list-style-type: none"> <li>- N° de elementos del modelo de datos que no corresponden con requisitos de usuario.</li> <li>- N° de requisitos de usuario no representados en el modelo de datos.</li> <li>- N° de elementos de datos que corresponden a requisitos de usuario pero definidos de forma inexacta.</li> <li>- N° de inconsistencias con el modelo de procesos.</li> </ul>
Integridad	<ul style="list-style-type: none"> <li>- N° de reglas del negocio que no se hacen cumplir por el modelo de datos.</li> <li>- N° de restricciones de integridad incluidas en el modelo de datos que no corresponden a políticas del negocio.</li> </ul>
Flexibilidad	<ul style="list-style-type: none"> <li>- N° de elementos en el modelo que están sujetos a cambios en el futuro.</li> <li>- Costes estimados de los cambios.</li> <li>- Importancia estratégica de los cambios.</li> </ul>
Comprensibilidad	<ul style="list-style-type: none"> <li>- Valoración de los usuarios sobre la comprensibilidad del modelo.</li> <li>- Capacidad de los usuarios de interpretar el modelo correctamente.</li> <li>- Valoración de los desarrolladores de aplicaciones sobre la comprensibilidad del modelo.</li> </ul>
Corrección	<ul style="list-style-type: none"> <li>- N° de violaciones de las convenciones de modelado de datos.</li> <li>- N° de violaciones a las formas normales.</li> <li>- N° de instancias de redundancia en el modelo.</li> </ul>
Simplicidad	<ul style="list-style-type: none"> <li>- N° de entidades.</li> <li>- N° de entidades y relaciones.</li> <li>- N° de constructores (<math>aN^E + bN^R + cN^A</math>).</li> </ul>
Integración	<ul style="list-style-type: none"> <li>- N° de conflictos con el modelo de datos corporativo.</li> <li>- N° de conflictos con los sistemas existentes.</li> <li>- Valoración de los representantes de todas las áreas de negocio.</li> </ul>
Implementabilidad	<ul style="list-style-type: none"> <li>- Valoración de riesgo técnico.</li> <li>- Valoración de riesgo de planificación.</li> </ul>



	<ul style="list-style-type: none"> <li>- Estimación del coste de desarrollo.</li> <li>- N° de elementos físicos incluidos en el modelo de datos.</li> </ul>
--	---

Estas métricas no fueron validadas ni teórica ni empíricamente. Moody [29] sugirió utilizar la investigación en acción (*action research*) ([9], [2]) como un método en el que tanto profesionales como investigadores trabajen juntos para demostrar la utilidad de tales métricas.

### 3.3. Métricas de Piattini (2001)

La idea de estos autores se centra en la medición de la mantenibilidad de los modelos ER. Como un atributo externo de la calidad, la mantenibilidad sólo puede medirse cuando el producto está finalizado o próximo a ello, por lo que la idea es definir medidas para la complejidad estructural de los modelos ER (un atributo interno) y basándose en dichas métricas predecir la mantenibilidad de los modelos ER, que lógicamente tiene una alta influencia sobre la mantenibilidad de la base de datos que finalmente se implementa.

La complejidad de un modelo ER puede estar altamente influenciada por los diferentes elementos que lo componen, tales como entidades, atributos, relaciones, generalizaciones, etc. Por lo tanto no es aconsejable definir una medida general para su complejidad [14]. Siguiendo este razonamiento [32] han propuesto un conjunto de medidas para medir la complejidad estructural de un modelo ER (ver tabla 3), siguiendo la noción de complejidad estructural o del producto de Henderson-Sellers.

**Tabla 3. Métricas para modelos ER [32]**

Nombre	Definición
NE	Número total de Entidades dentro de un modelo ER.
NA	Número total de Atributos en un modelo ER, teniendo en cuenta los atributos de las relaciones como los de las entidades. En este número se incluyen atributos simples, compuestos y multivaluados, cada uno de los cuales toma el valor 1.
NDA	Número total de Atributos Derivados en una modelo ER.
NCA	Número total de Atributos Compuestos en un modelo ER.
NMVA	Número total de Atributos Multivaluados en un modelo ER.
NNR	Número total de Relaciones en una modelo ER, teniendo en cuenta solamente relaciones comunes.
NM:NR	Número total de Relaciones M:N en un modelo ER.
N1:NR	Número total de Relaciones 1:N (incluyendo también relaciones 1:1) en un modelo ER.
NBinaryR	Número total de Relaciones Binarias en un modelo ER.
NN-AryR	Número total de Relaciones N-arias (no binarias) en un modelo ER.
NIS_AR	Número total de Relaciones Es_Un (generalización /especialización) que existen en un modelo ER. En este caso, se considera una relación por cada par padre-hijo, dentro de la relación Es_Un.
NRefR	Número total de Relaciones Reflexivas que existen en un diagrama ER.
NRR	Número de Relaciones Redundantes en un Modelo ER.

Las métricas NE, NA, NCA, NDA y NMVA fueron validadas como métricas de tamaño, y finalmente NR, N1:NR, NM:NR, NBinaryR, NN-AryR, NIS\_AR, NRefR, NRR como métricas de complejidad [32], utilizando la aproximación basada en propiedades propuesta por Briand et al. [6].

También fueron validadas teóricamente en Genero et al. [19], siguiendo el marco formal basado en la teoría de la medida, propuesto por Zuse [44], con el objetivo de evaluar qué tipo de escala caracteriza a cada métrica. Las métricas NR, N1:NR, NBinaryR fueron caracterizadas por encima de la escala ordinal y NE, NA, NCA, NDA, NMVA, NN-AryR, NIS\_AR, NRefR, NRR en la escala de ratio.

Estos autores han realizado la validación empírica de estas métricas por medio de un caso de estudio [17] y dos experimentos controlados ([18], [19]).

#### 4. CONCLUSIONES

En este artículo hemos planteado la importancia de abordar el tema de la calidad en las bases de datos y hemos mostrado algunos trabajos de investigación realizados en esta área, que creemos han tenido una mayor repercusión. Si bien todas estas propuestas han significado un gran contribución, creemos que es necesaria una mayor investigación en los aspectos relacionados con la calidad de las bases de datos y, especialmente, con la elaboración y refinamiento de métricas, tanto desde el punto de vista teórico como práctico.

Dado que el modelado conceptual se ha convertido en una etapa clave del desarrollo de SI, está surgiendo la necesidad de evaluar la calidad de los modelos conceptuales, en un sentido amplio, considerando tanto aquellos que tienen que ver con las bases de datos como los relacionados con los SI.

Como Moody [27] argumenta, para que el modelado conceptual pase de ser un “arte” a una disciplina ingenieril, se necesita que se definan explícitamente criterios de calidad y métricas para medirlos. Conocer esta lista de criterios de calidad nos ayudará a identificar aquellas áreas a las cuales debemos prestar atención. Las nociones subjetivas de la calidad del diseño no son suficientes para asegurar la calidad en la práctica, porque diferentes personas tendrán diferentes interpretaciones del mismo concepto (por ejemplo de flexibilidad). Por esta razón, la necesidad de validar las métricas es inminente. Las métricas pueden actuar como una guía para mejorar el modelo y explorar alternativas. Por ello, la calidad del modelo conceptual, debe ser abordada con métodos y técnicas rigurosas y cuantitativas.

La medición de las características de calidad del modelo conceptual es un área aún no consolidada y frecuentemente ignorada. Hasta ahora, la mayoría de los esfuerzos en la medición del software se han centrado en la medición de programas o del diseño avanzado. Esto corrobora el hecho de que comparado con el entendimiento generalizado del concepto de calidad en ingeniería del software, el concepto de calidad en el modelado conceptual no está claramente definido [1].

En efecto, siguiendo los comentarios realizados por Briand et al. ([5], [7]) y Calero et al. [1]:

- Las métricas se deben definir persiguiendo objetivos claros.
- Las métricas deben ser validadas teóricamente, respondiendo a la pregunta: ¿Está midiendo la medida el atributo que se pretende medir?
- Las métricas deben ser validadas empíricamente, respondiendo a la pregunta ¿Es la medida útil, en el sentido de si está relacionada con otras variables de la forma esperada?
- El cálculo de las métricas debe ser sencillo e incluso es mejor si su extracción es automática mediante una herramienta.

La tabla 4 resume las características más importantes de las principales propuestas sobre métricas para modelos conceptuales ER existentes en la literatura. La primera columna hace referencia a la principal fuente de las métricas. En la segunda, se presenta el enfoque de las métricas. La tercera columna se refiere al alcance de las métricas: el modelo conceptual completo o un elemento simple del modelo. La cuarta columna muestra si las métricas son objetivas o subjetivas (cuando se miden atributos de entidades, nos esforzamos en mantener nuestras medidas objetivas. Aun así, es importante reconocer que las medidas subjetivas pueden ser útiles, siempre y cuando entendamos la imprecisión [15]), por ejemplo, si se calculan mediante un método objetivo o mediante uno subjetivo (normalmente puntuaciones dadas por los usuarios o participantes). La quinta y sexta columnas reflejan si existen estudios publicados en los que se haya realizado la validación teórica o empírica de las métricas. La última columna refleja si existe una herramienta automática para el cálculo de las métricas.

**Tabla 4.** Resumen de las propuestas de métricas para modelos ER

Autores	Enfoque	Ámbito	Objetiva/ Subjetiva	Validación Teórica	Validación Empírica	Herramienta
Kesh (1995)	Ontológico y de comportamiento	Modelo ER	Objetivas y subjetivas	NO	NO	NO
Moody (1998)	Varios factores de calidad	Modelo ER	Objetivas y subjetivas	NO	NO	NO
Piattini et al., 2000)	Complejidad	Modelo ER	Objetivas	SI	Parcial	SI

De la tabla y los epígrafes anteriores podemos concluir que:

- Aunque parece que están definidas persiguiendo un objetivo claro, la lista completa de propiedades deseables para obtener un “buen” modelo conceptual, no está aún definida claramente.
- La mayoría de las métricas para modelos ER son subjetivas.
- La mayoría de las métricas para modelos ER no están soportadas por herramientas automáticas.
- Es necesario seguir trabajando en la validación empírica y teórica de las métricas.

#### REFERENCIAS

- [1] P. Assenova. y P. Johanneson, “Improving Quality in Conceptual Modelling by the Use of Schema Transformations”, *15<sup>th</sup> Int’l Conf. of Conceptual Modelling (ER’96)*, Cotbus, Germany, pp. 277-291, 1996.
- [2] D. Avison, F. Lan, M. Myers y A. Nielsen, “Action Research”, *Communications of the ACM*, 42(1), pp. 94-97, 1999.
- [3] C. Batini, S. Ceri y S. Navathe, *Conceptual Database Design. An Entity Relationship Approach*, Benjamin Cummings Publishing Company, 1992.
- [4] M. Boman, J. Bubenko, P. Johannesson y B. Wangler, *Conceptual Modelling*, Prentice Hall, 1997.
- [5] L. Briand, K. El Emam y S. Morasca, *Theoretical and Empirical Validation of Software Product Measures*, technical report ISERN-95-03, International Software Engineering Reserach Network, 1995.
- [6] L. Briand, S. Morasca y V. Basili, “Property-Based Software Engineering Measurement”, *IEEE Transactions on Software Engineering*, 22(1), pp. 68-86, 1996.

- [7] L. Briand, S. Morasca y V. Basili, *An Operational Process for Goal-Driven Definition of Measures*, technical report University of Maryland, CS-TR-3343, version 2, 1999.
- [8] C. Calero, M. Piattini y M. Genero, "Empirical Validation of Referential Integrity", *Information and Software Technology*, 43, pp. 949-957, 2001.
- [9] P. Checkland y J. Scholes, *Soft Systems Methodology in Action*. Wiley, Chichester, 1990.
- [10] P. Chen, "The Entity-Relationship Model: Toward a Unified View of Data", *ACM Transactions on Database Systems*, 1(1), pp. 9-37, 1976.
- [11] P. Chen, B. Thalheim y L. Wong, "Future Directions of Conceptual Modelling", *Conceptual Modeling: Current Issues and Future Directions*, eds. P. Chen, J. Akoka, H. Kangassalo, B. Thalheim, LCNS 1565, pp. 258-271, 1991.
- [12] A. De Miguel, M. Piattini, y E. Marcos, *Diseño de Bases de Datos Relacionales*, Ra-Ma, Madrid, 1999.
- [13] W. Deming, *Out of the Crisis*, MIT Center for Advanced Engineering, Cambridge, 1986.
- [14] N. Fenton, "Software Measurement: A Necessary Scientific Basis", *IEEE Transactions on Software Engineering*, 20(3), pp. 199-206, 1994
- [15] N. Fenton y S. Pfleeger, *Software Metrics: A Rigorous Approach*, 2nd. edition. London, Chapman & Hall, 1997.
- [16] R.A. Freund, "Definitions and Basic Quality Concepts", ed. M. Sepehri, *Quest for Quality: Managing the Total System*, Atlanta, Georgia, Industrial Engineering and Management Press, 1985.
- [17] M. Genero, M. Piattini y C. Calero, "An Approach to Evaluate the Complexity of Conceptual Database Models", *2nd European Software Measurement Conference (FESMA-AEMES 2000)*, Madrid, 2000.
- [18] M. Genero, J. Olivas, M. Piattini y F. Romero, "Knowledge Discovery For Predicting Entity Relationship Diagram Maintainability", *SEKE 2001, Knowledge Systems Institute*, pp. 203-211, 2001.
- [19] M. Genero, G. Poels y M. Piattini, "Defining and Validating Measures for Conceptual Data Model Quality", *CAISE 2002*, Toronto, Canada, LNCS 2348, Springer-Verlag, pp. 724-727, 2002.
- [20] A. Gilles, *Software Quality- Theory and Management*, Chapman & Hall Computing. London, 1992.
- [21] ISO/IEC 9126-1.2, *Information Technology- Software Product Quality – Part 1: Quality Model*, 1999.
- [22] S. Kesh, "Evaluating the Quality of Entity Relationship Models" *Information and Software Technology*, Vol. 37, No.12, pp. 681-689, 1995.
- [23] J. Krogstie, O. Lindland y G. Sindre, "Towards a Deeper Understanding of Quality in Requirements Engineering", *CAISE 1995*, Jyvaskyla, Finland, pp. 82-95, 1995.
- [24] S. Liddle, W. Stephen y S. Woodfield, "A Summary of the ER'97 Workshop on Behavioural Modeling", *Conceptual Modeling: Current Issues and Future Directions*, eds. P. Chen, J. Akoka, H. Kangassalo, B. Thalheim, LCNS 1565, pp. 258-271, 1999.
- [25] O. Lindland, G. Sindre y A. Solvberg, "Understanding Quality in Conceptual Modelling", *IEEE Software*, Vol. 11, No. 2, pp. 42-49, 1994.
- [26] R. Maier, "Organizational Concepts and Measures for the Evaluation of Data Modeling", Capítulo 1 en *Developing Quality Complex Databases Systems: Practices, Techniques and Technologies*, ed. S. Becker, Idea Group Publishing, pp. 1-27, 2001
- [27] D. Moody, "Metrics for Evaluating the Quality of Entity Relationship Models", *Proceedings of the Seventeenth International Conference on Conceptual Modelling (ER '98)*, Singapore, pp. 213-225, Nov. 1998.

- [28] D. Moody y G. Shanks, "What Makes a Good Data Model? Evaluating the Quality of Entity Relationships Models", *Proceedings of the 13<sup>th</sup> International Conference on Conceptual Modelling (ER '94)*, Manchester, England, pp. 94-111, Dic. 1994.
- [29] D. Moody, G. Shanks, y P. Darke, "Improving the Quality of Entity Relationship Models – Experience in Research and Practice", *Proceedings of the Seventeenth International Conference on Conceptual Modelling (ER '98)*, Singapore, pp. 255-276, 1998.
- [30] R. Muller, *Database Design for Smarties. Using UML for Data Modelling*, San Francisco, Morgan Kaufmann, 1999.
- [31] M. Piattini, M. Genero, C. Calero, M. Polo y F. Ruiz, "Database Quality", *Advanced Database Technology and Design*, eds. O. Diaz, y M. Piattini, Londres, Artech House, 2000.
- [32] M. Piattini, M. Genero y L. Jiménez, "A Metric-Based Approach for Predicting Conceptual Data Models Maintainability", *International Journal of Software Engineering and Knowledge Engineering*, 11(6), World Scientific Publishing Company, pp. 703-729, 2001
- [33] M. Piattini, C. Calero, y M. Genero, (eds) *Information and Database Quality*, Kluwer Academic Publishers, Norwell, EEUU, 2002.
- [34] G. Poels. y G. Dedene, *DISTANCE: A Framework for Software Measure Construction*, research report DTEW9937, Dept. Applied Economics, Katholieke Universiteit Leuven, Belgium, 1999.
- [35] G. Poels y G. Dedene, "Distance-Based Software Measurement: Necessary and Sufficient Properties for Software Measures", *Information and Software Technology*, 42(1), pp. 35-46, 2000.
- [36] G. Poels y G. Dedene, "Measures for Assessing Dynamic Complexity Aspects of Object-Oriented Conceptual Schemes", *Proceedings of the 19th International Conference on Conceptual Modeling (ER 2000)*, Salt Lake City, USA, pp. 499-512, 2000.
- [37] K. Pohl, "The Three Dimensions of Requirements Engineering: A Framework and its Applications", *Information Systems*, Vol. 19, pp. 243-258, 1994.
- [38] M. Reingruber y W. Gregory, *The Data Modelling Handbook. A Best-Practice Approach to Building Quality Data Models*, John Wiley & Sons, Inc, 1994.
- [39] R. Schuette y T. Rotthowe, "The Guidelines of Modeling – An Approach to Enhance the Quality in Information Models", *Proceedings of the Seventeenth International Conference on Conceptual Modelling (ER '98)*, Singapore, pp. 240-254, 1998.
- [40] T. Teorey, *Database Modeling and Design: The Entity-Relationship Approach*, Morgan Kaufmann, 1990.
- [41] B. Thalheim, *Entity-Relationship Modeling*, Springer-Verlag, 2000.
- [42] R. Wang, H. Kon y S. Madnick, "Data Quality Requirements Analysis and Modeling", *Proc. of the 9<sup>th</sup> International Conference on Data Engineering*, Viena, IEEE Computer Society, pp. 670-677, 1993.
- [43] E. Weyuker, "Evaluating Software Complexity Measures", *IEEE Transactions Software Eng.*, Vol. 14 No. 9, pp. 1357-1365, 1998.
- [44] H. Zuse, *A Framework of Software Measurement*, Berlin, Walter de Gruyter, 1998.