

Identification of Terminology Used by Business People to Express Data Quality

L. Ortega, A. Caro, A. Rodríguez, and I. Velásquez

Abstract—Problems associated with data quality cost companies millions of dollars every year because of bad decision-making based on inaccurate information, noncompliance with regulations, or failure to timely deal with customer problems. Nowadays, data have become one of the most valuable assets of organizations; data quality is no longer a task involving only information technology managers, business managers are increasingly more implicated. In the context of business management processes, the early specification of data quality requirements can be crucial. This article reveals the results of a research study conducted in specialized sites, blogs, forums, and literature related to how business people (not specialized in information technology) express data quality needs. Additionally, how the result was validated is shown through the participation of business experts and analysts. The outcome was lists with 18 types of problems associated with data quality and 35 characteristics that data must comply with in relation to quality.

Index Terms—Data Quality, Business People, Business Process, Information Technologies People, Data Quality Problems, Data Quality Terminology.

I. INTRODUCCIÓN

LOS PROCESOS de Negocio (Business Process, BP) son un conjunto de actividades coordinadas para alcanzar un objetivo de negocio en un contexto tecnológico y organizacional [1]. Las organizaciones se encuentran inmersas actualmente en un ambiente competitivo y cambiante, volviendo esencial el tener BPs claramente visibles para así estudiarlos, mejorarlos y optimizarlos continuamente.

Los BPs se vuelven visibles a través del modelado, donde se utiliza un lenguaje o notación típicamente en base a gráficos; el Lenguaje Unificado de Modelado (Unified Modelling Language, UML) [2] y la Notación de Modelado de Procesos de Negocio (Business Process Modelling Notation, BPMN) [3] son dos de estos lenguajes. Debido a que es el más utilizado en la industria, BPMN es considerado el estándar *de facto* para modelar BPs [4].

Los procesos de negocio son útiles tanto para personas del área de negocios como para las del área de las Tecnologías de Información (TI): para las primeras, apoyan en la realización de

sus objetivos estratégicos; para las segundas, poseen un rol importante en el diseño y creación de sistemas de información flexibles [1].

Sin embargo, tanto la gente de negocio como la de TI en organizaciones tienen distintos puntos de vista sobre cómo abordar problemas, normalmente usando terminologías distintas para referirse a un mismo concepto, lo que puede generar malentendidos e interpretaciones erróneas [5-7]. Las diferencias entre ellos es una situación real que persiste hace décadas [8]. El problema es serio pues interpretar erróneamente un conjunto de requerimientos podría llevar a implementar una infraestructura de TI que no satisfaga completamente las necesidades de negocio de una organización.

Existen múltiples métodos para identificar y especificar requerimientos no funcionales por medio de BPs [9-12]. Uno de estos es la Calidad de Datos (Data Quality, DQ) [13]. La DQ se define como el “grado en el que las características de los datos satisfacen necesidades declaradas e implícitas cuando son utilizados en condiciones específicas” [14] y es una de las actividades más importantes en las organizaciones [15]. La especificación de requerimientos de DQ debería considerar la participación tanto de personas de negocios como de TI [12]. Sin embargo, esto se ve dificultado por la situación explicada anteriormente, lo que causa el riesgo de malinterpretaciones entre las terminologías usadas por ambos grupos de personas.

Dada esta situación, este trabajo examina la terminología usada en el ambiente de negocios para expresar necesidades de DQ (características y/o problemas) con el objetivo de que, a partir de ésta, se pueda a futuro establecer una correspondencia con la terminología usada por la gente de TI, reduciendo así el riesgo de malinterpretaciones. La estrategia utilizada fue revisar sistemáticamente tanto la literatura como los sitios web orientados a la gente de negocios (sitios especializados, blogs y foros entre otros) y luego validar la terminología identificada por medio de encuestas. Se utilizó la metodología de revisiones sistemáticas de la literatura dado que permite identificar la información relevante en relación a un tema seleccionado [16] y es ampliamente utilizada en la ingeniería de software [17]. Existe evidencia, también, del uso de revisiones sistemáticas de la web, como en [18]. El resultado de esta investigación es una lista de términos usados por gente de negocios para referirse a características y problemas de DQ. Esta lista puede ser de utilidad para contextos donde la terminología de DQ usada por la gente de negocios sea de relevancia para la gente de TI.

El resto del artículo se organiza de la siguiente manera: la Sección 2 presenta trabajos relacionados a esta investigación; la Sección 3 describe el proceso realizado para identificar la terminología. La Sección 4 muestra la validación de ésta. Y la

Esta investigación es parte de los siguientes proyectos: DIUBB 144319 2/R y BuPERG (DIUBB 152419 G/EF).

L. Ortega es titulado del Master en Ciencias de la Computación de Universidad del Bío-Bío, Chillán, Chile (e-mail: luortegar@gmail.com).

A. Caro es profesora en la Facultad de Ciencias Empresariales de la Universidad del Bío-Bío, Chillán, Chile (e-mail: mcaro@ubiobio.cl).

A. Rodríguez es profesor en la Facultad de Ciencias Empresariales de la Universidad del Bío-Bío, Chillán, Chile (e-mail: alfonso@ubiobio.cl).

I. Velásquez es ayudante de investigación en la Facultad de Ciencias Empresariales de la Universidad del Bío-Bío, Chillán, Chile (e-mail: ignaciowl@outlook.com).

Sección 5 define las conclusiones y el trabajo futuro.

II. TRABAJOS RELACIONADOS

Existen múltiples métodos que permiten la identificación y especificación de requerimientos funcionales y no funcionales en BPs. Estos métodos generalmente se definen como extensiones de BPMN por ser el estándar de facto [4]. En 2007 se presenta una extensión para expresar requerimientos de seguridad en BPs [9]. En 2008 se propone un método para modelar requerimientos no funcionales [10]. En 2010 se presenta una extensión para especificar necesidades de los clientes [19]. En 2011 se proponen un método para representar explícitamente restricciones legales [20], una extensión para definir propiedades no funcionales de BPs [11] y un método para analizar el rendimiento de BPs [21].

Múltiples trabajos han abordado la DQ en distintos contextos [22, 23]. Dentro del contexto de BP, en 2004 se presenta un patrón para asegurar DQ reutilizable en una organización [24]. En 2006 se propone un marco conceptual para modelar BPs; esto considerando la estimación y gestión de DQ en los sistemas de información [25]. Dos marcos conceptuales son presentados en 2009: el objetivo del primero es el mejoramiento continuo de BPs con reglas de negocio que representen requerimientos de DQ [26]; el segundo identifica atributos de calidad en BPs basándose en cuatro dimensiones [27]. Finalmente, un trabajo en 2010 provee la base para evitar problemas de inexactitud de datos en BPs [28].

Los trabajos anteriores no abordan directamente el modelado de DQ en BPs. Uno que si lo hace es [12]. En éste, el método BPiDQ es presentado como una extensión de BPMN que permite la especificación de requerimientos de DQ en modelos de BP. BPiDQ tiene cuatro etapas. La primera se dedica a especificar los requerimientos de DQ de un proceso de negocio. Esta tarea es realizada por expertos de negocio, quienes son responsables de definir las necesidades del negocio. En la segunda etapa, sin embargo, son los expertos de TI quienes expresan estas especificaciones de manera más concreta y técnica. Esto es realizado en base a, por ejemplo, la terminología definida por el estándar ISO/IEC 25012 [14].

El método BPiDQ es un buen ejemplo donde el uso de terminologías distintas para expresar los mismos conceptos podría causar malentendidos e interpretaciones erróneas [5, 6]. Este problema es denotado en [7], donde se indica la falta de entendimiento del negocio por parte de la gente de TI, así como el problema de la comunicación entre éstos y los expertos de negocio. Entonces, si se dispone de una terminología que facilite la participación de la gente de negocios, que luego pueda ser mapeada a la terminología usada por la gente de TI (idealmente de forma automática), se podría ayudar considerablemente a evitar estas situaciones.

III. IDENTIFICACIÓN DE LA TERMINOLOGÍA DE CALIDAD DE DATOS

Se utilizó una estrategia similar a la utilizada en [18] para identificar la terminología de DQ usada por la gente de negocios. Por un lado, se realizó una Revisión Sistemática de

la Literatura (RSL) y, por otro, una Revisión Sistemática de la Web (RSW) sobre sitios especializados de BP. Primero se presentan el objetivo y la pregunta de investigación de estas revisiones, para luego describir la realización de éstas. Finalmente, se presenta la terminología de DQ preliminar en base a los resultados de las revisiones.

A. Objetivo y Pregunta de Investigación

El objetivo de la investigación fue de *analizar estudios y trabajos en el área de negocios que contengan terminología para la definición de requerimientos y problemas de DQ usando un lenguaje cercano al ámbito de negocios*.

En base a éste, se definió la siguiente pregunta de investigación: *¿Qué términos son usados en el ámbito de negocios para expresar características y problemas de DQ?*

B. Revisión Sistemática de la Literatura

Cada etapa de la RSL se realizó siguiendo la metodología propuesta en [16]. El protocolo de búsqueda se definió en base al objetivo y la pregunta de investigación. Las fuentes de búsqueda fueron Google Scholar, Scopus, Springer Link, SciELO y Science Direct. Los términos y combinaciones de búsqueda, en inglés, son presentados en la Tabla 1.

TABLA I
TÉRMINOS Y COMBINACIONES DE BÚSQUEDA

<i>Términos</i>	"Data Quality", "Requirements", "Information Quality", "Business", "Problem"
<i>Combinaciones</i>	"Problem" + "Data Quality" + "Business" "Requirements" + "Data Quality" + "Business" "Problem" + "Information Quality" + "Business" "Requirements" + "Information Quality" + "Business"

La búsqueda fue realizada tanto en inglés como en español: en inglés por ser la lengua franca para las publicaciones académicas [29], mientras que el español fue seleccionado como un complemento al ser el lenguaje nativo de los autores. De este modo, ocho búsquedas fueron realizadas por cada fuente. El protocolo de revisión consideró los siguientes criterios de inclusión: (i) que el título debe incluir al menos uno de los términos de búsqueda, y (ii) que al menos uno de los términos de búsqueda debe estar en las palabras clave. Por otro lado, los criterios de exclusión considerados fueron: (i) que ningún término de búsqueda aparezca en la publicación, y (ii) que la publicación no asocie DQ con la gente de negocios.

El desarrollo de la revisión se realizó en base a la planificación anterior. Un resumen de los resultados de la revisión es presentado en la Tabla 2, la cual presenta, por cada fuente, el número de resultados al buscar las combinaciones de términos definidas, el número de publicaciones analizadas y el número de publicaciones seleccionadas. En base a la cantidad de resultados de cada búsqueda, se tomó la decisión de revisar las primeras 100 publicaciones por cada búsqueda.

C. Revisión Sistemática de la Web

La planificación de la RSW consideró un protocolo de búsqueda en base al objetivo y pregunta de investigación propuestos. Las fuentes de búsqueda fueron BPTrends [30], Information Management [31], Global Excellence [32], Global

Data Excellence [33] y Google [34]. Los mismos términos y combinaciones aplicados en la RSL (Tabla 1) fueron usados para la RSW, excepto por el término “Business”, el cual solo fue utilizado para la búsqueda en Google.

TABLA II
RESULTADOS DE LA REVISIÓN SISTEMÁTICA DE LA LITERATURA

Fuente	Resultados Combinaciones	Analizados	Seleccionados
Google Scholar	213.547	800	22
Scopus	89	89	1
Springer Link	57	57	1
SciELO	675	430	6
Science Direct	924	678	1
Total	215.292	2.054	31

Para el protocolo de revisión se definieron los siguientes criterios de inclusión: (i) que el sitio o documento se refiera a problemas y/o requerimientos de DQ, y (ii) que el sitio o documento muestre características de DQ expresadas por gente de negocios en el ámbito de negocios. Se consideró como criterio de exclusión: el que la información presentada no tuviera relación alguna a DQ y su definición en el contexto de la gente de negocios.

La revisión se realizó en base a la planificación anterior. La Tabla 3 resume los resultados, mostrando, por cada fuente, el número de resultados al buscar las combinaciones de términos definidas, el número de sitios y documentos analizados y el número de sitios y documentos seleccionados. Para los sitios especializados y cuando los resultados eran suficientes, se revisaron los primeros 50 resultados de cada búsqueda. En el caso de Google, también se revisaron hasta 50 resultados por búsqueda, pues se observó que resultados posteriores a los primeros 50 no eran relevantes para la investigación.

TABLA III
RESULTADOS DE LA REVISIÓN SISTEMÁTICA DE LA WEB

Fuente	Resultados Combinaciones	Analizados	Seleccionados
BPTrends	46	46	5
Information Management	1,793	177	12
Global Excellence	12	12	0
Global Data Excellence	59	59	3
Google	11,820,000	400	39
Total	11,821,910	694	59

D. Obtención de Terminología de DQ Preliminar

Se han identificado 97 problemas de DQ a través de la revisión de los resultados seleccionados en la RSL y la RSW. Éstos fueron analizados y clasificados en 18 tipos según sus similitudes. Entre ellos, imprecisión, falta de completitud e inconsistencia fueron los más recurrentes. Adicionalmente se obtuvo una lista con 457 características de DQ. Luego de

analizarlas, se eliminaron las repetidas y el resto fue agrupado para así obtener un conjunto de 36 características de DQ.

Los 18 tipos de problemas de DQ definidos por medio de la revisión y las referencias a sus fuentes son los siguientes:

- (i) *Ambigüedad en los Datos* [34, 35]
- (ii) *Asimetría en los Datos Registrados* [36, 37]
- (iii) *Datos Duplicados* [31, 34, 38]
- (iv) *Datos Imprecisos* [34, 35, 39]
- (v) *Datos Mal Escritos* [34]
- (vi) *Datos no Accesibles* [34, 35, 40]
- (vii) *Datos no Actualizados* [31, 34]
- (viii) *Datos no Adecuados* [31, 34, 41]
- (ix) *Dato no Válidos, Erróneos* [31, 33-35, 42]
- (x) *Datos no Veraces* [42]
- (xi) *Datos Poco Confiables* [34, 35]
- (xii) *Falta de Completitud de los Datos* [30, 34, 35, 38]
- (xiii) *Falta de Compromiso al Registrar Datos* [34]
- (xiv) *Falta de Estandarización de los Datos* [34]
- (xv) *Falta de Integridad en los Datos* [31, 40]
- (xvi) *Falta de Procedimientos de Validación de los Datos* [34]
- (xvii) *Inconsistencia en los Datos* [31, 34, 43]
- (xviii) *Pérdida de Datos* [35, 40]

Las 36 características de DQ definidas por medio de la revisión y las referencias a sus fuentes son las siguientes:

- (i) *Accesibilidad** [31, 34], 44-50]
- (ii) *Actualidad** [33, 34, 51, 52]
- (iii) *Adecuación* [34, 35, 48-51]
- (iv) *Amplitud* [34, 35, 53]
- (v) *Aplicabilidad* [34]
- (vi) *Completitud** [30, 31, 33-35, 44, 53-56]
- (vii) *Comprensibilidad** [34, 35, 49]
- (viii) *Confiabilidad* [34, 35, 49, 55, 57]
- (ix) *Confidencialidad** [34]
- (x) *Conformidad** [31, 34]
- (xi) *Consistencia** [31, 33, 34, 50, 58]
- (xii) *Credibilidad** [34, 42, 44, 50]
- (xiii) *Disponibilidad** [34, 44, 53]
- (xiv) *Duplicidad* [34, 53]
- (xv) *Eficacia* [34]
- (xvi) *Eficiencia** [34, 35]
- (xvii) *Estandarización* [31, 34, 35, 59]
- (xviii) *Estructurada* [31, 34]
- (xix) *Exactitud** [25, 33-35, 45, 46, 48, 53-55, 60]
- (xx) *Integridad* [30, 31, 33, 34, 46, 48, 50]
- (xxi) *Interpretabilidad* [34, 35, 44, 48, 50]
- (xxii) *Libre de Errores* [34, 44]
- (xxiii) *Mantenibilidad* [44, 53]
- (xxiv) *Objetividad* [34, 44, 48, 50]
- (xxv) *Operatividad* [49]
- (xxvi) *Oportunidad* [34, 43, 48, 49, 53-55]
- (xxvii) *Relevancia* [31, 34, 35, 43, 47, 48, 50, 53, 55]
- (xxviii) *Representación* [44, 50]
- (xxix) *Reputación* [34, 47, 48]
- (xxx) *Seguridad* [34, 47, 48]

- (xxxi) *Sincronización* [34, 53]
- (xxxii) *Trazabilidad** [46]
- (xxxiii) *Usabilidad* [31, 34, 44, 46, 48, 53]
- (xxxiv) *Utilidad* [34, 35, 61]
- (xxxv) *Valor Añadido* [34, 48, 50]
- (xxxvi) *Veracidad* [49, 51-53]

En esta lista, un asterisco (*) destaca aquellas características (12) que también han sido definidas en el estándar ISO/IEC 25012 [13], que representa la terminología usada por la gente de TI, la cual es más técnica. Éste es usado de referencia por ser el estándar de calidad de datos para administradores de TI. El estándar considera 15 características. Aquellas no incluidas en la lista obtenida en base a la revisión son Precisión, Portabilidad y Recuperabilidad. La falta de estas características sería consistente con el enfoque de investigación seleccionado, pues éste identifica la terminología de calidad de datos usada por gente de negocios y no por gente de TI. Portabilidad y Recuperabilidad son conceptos técnicos, mientras que Precisión es usualmente usada como sinónimo de Exactitud.

Una vez identificados los problemas y características de DQ, se analizaron las relaciones entre ambos. Por ejemplo, se determinó que el problema de “Ambigüedad en los Datos” está relacionado con la característica de “Interpretabilidad”, pues, por su descripción, este problema se vincula con temas de datos que tienen significados diferentes para distintas personas.

Tablas conteniendo la lista de los 97 problemas y las 457 características de DQ, agrupados en los 18 problemas y 36 características presentados en este artículo, además de las relaciones identificadas entre problemas y características, pueden ser encontradas en los materiales complementarios (<http://colvin.chillan.ubiobio.cl/mcaro/ortega/>).

IV. VALIDACIÓN DE RESULTADOS

Se aplicó una encuesta para validar la terminología de calidad de datos. Las encuestas son un método empírico para compilar información sobre un grupo de personas, para así describir, comparar o explicar sus conocimientos, actitudes o comportamiento [62]. La encuesta se enfocó sobre expertos del área de negocios y su uso de la terminología obtenida. Una encuesta consta de las siguientes actividades [62]: (i) definición de objetivos, (ii) diseño de la encuesta, (iii) creación del instrumento, (iv) validación del instrumento, (v) aplicación de la encuesta, (vi) recopilación y análisis de los datos, y (vii) reporte de resultados. El desarrollo de la encuesta en base a estas actividades es descrito a continuación, seguido del análisis de los resultados del estudio.

A. Desarrollo de la Encuesta

Los objetivos de la encuesta fueron: (i) validar el uso de características de DQ en el área de negocios, y (ii) validar los tipos de problemas de DQ en el área de negocios. Ambos objetivos a través de la consulta a expertos en esta área.

Se consideró el estilo observacional al diseñar la encuesta. Éste es utilizado cuando se recolecta la información sin cambiar el ambiente y con una interacción transversal, lo que implica intervenir con las personas solo una vez [62]. Para este caso, se consultó la opinión de un grupo de expertos de negocios sobre

el uso de la terminología de DQ.

Respecto a la construcción del instrumento, se ha logrado identificar 18 tipos de problemas de DQ y 36 características de DQ a través de la RSL y la RSW. Con el objetivo de evitar el tedio en los encuestados, se decidió crear dos cuestionarios, basados en preguntas cerradas, conteniendo 9 tipos de problemas de DQ y 18 características de DQ cada uno. Además, se incluyeron cinco preguntas demográficas respecto al ambiente laboral y experiencia del encuestado y una pregunta abierta en caso de que éste utilizara otros términos de DQ no considerados por el instrumento. Las preguntas cerradas fueron medidas utilizando una Escala Likert (EL) de 5 puntos. Una versión del cuestionario, incluyendo los 18 tipos de problemas y 36 características de DQ, puede ser encontrada en los materiales complementarios (<http://colvin.chillan.ubiobio.cl/mcaro/ortega/>).

Una vez definido el instrumento, se realizó un estudio piloto con dos participantes antes de aplicar la encuesta. Éste permitió asegurar la adecuación de las preguntas y mejorar su comprensión.

Los participantes fueron seleccionados utilizando el método no probabilístico y el método de bola de nieve por ser adecuados para este tipo de estudio y por su facilidad de aplicación [60].

Se envió un correo electrónico a 25 contactos, describiendo el estudio y un enlace para contestar la encuesta en línea. En el cuestionario se incluían las instrucciones, mientras que no había un límite sobre el tiempo de respuesta.

Se recolectó un total de 38 cuestionarios válidos, 19 de cada tipo. Debido a la utilización del método de bola de nieve, no fue posible conocer la cantidad exacta de participantes que recibieron la encuesta.

B. Análisis de los Resultados del Estudio

Un análisis extendido de los resultados de la encuesta, con toda la información estadística, puede ser encontrado como material complementario en la siguiente dirección: <http://colvin.chillan.ubiobio.cl/mcaro/ortega/>. Dado los límites de espacio considerados en este artículo solo es posible mostrar un resumen de las respuestas de los encuestados el que se presenta a continuación.

El análisis demográfico mostró que los encuestados tenían un promedio de 21,3 años de experiencia laboral, en distintos tipos de industria, lo que les da el nivel de conocimiento necesario para identificar problemas y características de DQ.

Si bien la terminología se obtuvo de la literatura y de sitios web especializados, lo cual se podría argumentar que le da validez para el área de negocios, para seleccionar el conjunto final de términos, se usó como criterio que 55% o más de los participantes valoraran con 3, 4 o 5 (ocasionalmente, casi siempre y siempre) su uso del término correspondiente.

En la Tabla 4 se presentan algunos resultados del análisis estadístico sobre los tipos de problemas de DQ. Por cada tipo de problema se muestra el porcentaje de participantes que indicaron utilizar el término con 3, 4 o 5, su media y su moda.

En base a la media de cada término, los tipos de problemas de DQ más usados son “Datos no Actualizados”, “Datos

Duplicados” y “Pérdida de Datos”, mientras que los menos usados son “Falta de Integridad en los Datos”, “Asimetría en los Datos Registrados” y “Falta de Estandarización de los Datos”. Considerando el criterio de validación, todos los tipos de problemas se consideraron válidos.

TABLA IV
ANÁLISIS DE RESULTADOS DE TIPOS DE PROBLEMAS DE CALIDAD DE DATOS

Tipo de Problema	$\sum 3 a 5 pts. (%)$	Media	Moda
Ambigüedad en los Datos	73,68	3,21	3
Asimetría en los Datos Registrados	63,20	3,00	4
Datos Duplicados	73,70	3,47	5
Datos Imprecisos	63,16	3,26	4
Datos Mal Escritos	63,20	3,26	2
Datos no Accesibles	63,20	3,11	4
Datos no Actualizados	68,42	3,53	5
Datos no Adecuados	57,89	3,11	2
Datos no Válidos, Erróneos	68,40	3,37	4
Datos no Veraces	63,20	3,11	4
Datos Poco Confiables	68,42	3,16	3
Falta de Completitud de los Datos	78,95	3,37	4
Falta de Compromiso al Registrar Datos	57,89	3,26	2
Falta de Estandarización de los Datos	57,90	3,11	4
Falta de Integridad en los Datos	57,90	2,84	2
Falta de Procedimientos de Validación de los Datos	68,42	3,21	5
Inconsistencia en los Datos	73,68	3,42	5
Pérdida de Datos	68,40	3,47	5

En la Tabla 5 se presentan algunos resultados de la validación de las características de DQ. Por cada característica se muestra el porcentaje de participantes que indicaron utilizar el término con 3, 4 o 5, su media y su moda.

En base a la media de cada término, las características más usadas son “Aplicabilidad”, “Exactitud” y “Actualidad”, mientras que las menos usadas son “Representación”, “Accesibilidad”, “Consistencia” y “Reputación”. Usando el criterio de validación, se determinó como válidas a la mayoría de las características. La excepción fue “Representación”.

V. CONCLUSIONES Y TRABAJO FUTURO

La gente de negocios y la de TI en las organizaciones tienen distintos puntos de vista para resolver problemas, normalmente usando distintas terminologías para expresarse sobre los mismos conceptos. Esto se vuelve serio cuando una mala interpretación de los requerimientos del negocio podría llevar a la gente de TI a implementar una infraestructura de TI insatisfactoria para las necesidades de la organización.

En este trabajo, el cual se basa en revisiones sistemáticas de la literatura y de la web, se ha logrado establecer una lista con 18 tipos de problemas de DQ y 36 características de DQ, representando así la forma en que las personas del área de negocios se expresan respecto a la calidad de datos.

TABLA V
ANÁLISIS DE RESULTADOS DE CARACTERÍSTICAS DE CALIDAD DE DATOS

Característica	$\sum 3 a 5 pts. (%)$	Media	Moda
Accesibilidad	63,16	3,16	4
Actualidad	94,70	4,21	5
Adecuación	94,70	3,79	4
Amplitud	84,20	3,74	4
Aplicabilidad	94,70	4,26	5
Completitud	100,00	4,05	4
Comprensibilidad	94,70	3,79	4
Confiabilidad	84,20	3,79	4
Confidencialidad	73,70	3,26	4
Conformidad	84,21	3,58	4
Consistencia	68,40	3,16	4
Credibilidad	94,70	4,11	4
Disponibilidad	73,70	3,42	4
Duplicidad	73,70	3,37	3
Eficacia	84,20	3,47	4
Eficiencia	78,95	3,53	5
Estandarización	100,00	4,21	4
Estructurada	94,70	3,84	4
Exactitud	94,70	4,21	5
Integridad	100,00	3,79	4
Interpretabilidad	100,00	4,00	4
Libre de Errores	84,21	3,42	3
Mantenibilidad	89,47	3,63	4
Objetividad	94,74	3,84	4
Operatividad	100,00	4,21	4
Oportunidad	78,95	3,58	4
Relevancia	84,21	3,63	4
Representación	52,63	2,89	2
Reputación	84,21	3,16	3
Seguridad	73,68	3,63	5
Sincronización	84,21	3,63	5
Trazabilidad	89,47	3,74	5
Usabilidad	84,21	3,63	4
Utilidad	89,47	3,84	4
Valor Añadido	73,68	3,26	4
Veracidad	78,95	3,58	4

Una encuesta fue aplicada a 38 expertos del área de negocios, los cuales tenían experiencia gestionando los datos utilizados por sus organizaciones para la toma de decisiones. Esta encuesta permitió validar la terminología de problemas y características de DQ identificada. Solo una característica de DQ no pudo ser validada, la cual es “Representación”, obteniendo un conjunto final de 35 características de DQ. Además, se pudo observar que, si bien la gente de negocios le da importancia a la calidad de datos, el vocabulario utilizado por estos para ello es variado. Del mismo modo, se pudo confirmar que la gente de negocios utiliza la mayoría de las

características definidas en el estándar ISO/IEC 25012, pero que, además, agregan varias características más específicas.

Al analizar las respuestas de la encuesta, se observa que la experiencia laboral de los encuestados exhibe un alto valor, lo que indica un buen nivel de experiencia de los encuestados.

Como trabajo futuro se planea el desarrollo de una ontología de calidad de datos en el contexto de procesos de negocio, la cual permita capturar el conocimiento adquirido con relación a la perspectiva de negocios de la calidad de datos. Dentro de las actividades a realizar para el desarrollo de esta ontología, se tiene en consideración la realización de una jerarquización de los términos identificados en este trabajo, a modo de definir distintas categorías y niveles de términos o conceptos. Adicionalmente, es de interés el integrar esta ontología con alguna extensión de Calidad de Datos para el modelado de Procesos de Negocio existente, como lo es el método BPiDQ.

REFERENCIAS

- [1] M. Weske, "Business Process Management—Concepts, Languages, Architectures, Verlag," ed: Berlin, 2007.
- [2] O. OMG, "Unified modeling language (OMG UML)," Language, 2007.
- [3] B. P. Model, "Notation (BPMN) version 2.0," OMG Specification, Object Management Group, pp. 22-31, 2011.
- [4] P. Harmon and C. Wolf, "Business process modeling survey," Business process trends, p. 36, 2011.
- [5] Cravero Leal, J. N. Mazón, and J. Trujillo, "A business-oriented approach to data warehouse development," Ingeniería e Investigación, vol. 33, no. 1, pp. 59-65, 2013.
- [6] D. E. Jenz, "Business process ontologies: Speeding up business process implementation," Jenz & Partner GmbH2003, 2003.
- [7] M. A. T. Cespedes, "Towards a Methodology of Business Process Modeling," *IEEE Latin America Transactions*, vol. 16, no. 3, pp. 996-1003, 2018.
- [8] J. Coughlan, M. Lycett, and R. D. Macredie, "Understanding the business-IT relationship," International Journal of Information Management, vol. 25, no. 4, pp. 303-319, 2005.
- [9] Rodríguez, E. Fernández-Medina, and M. Piattini, "A BPMN extension for the modeling of security requirements in business processes," IEICE transactions on information and systems, vol. 90, no. 4, pp. 745-752, 2007.
- [10] J. Pavlovski and J. Zou, "Non-functional requirements in business process modeling," in Proceedings of the fifth Asia-Pacific conference on Conceptual Modelling-Volume 79, 2008, pp. 103-112: Australian Computer Society, Inc.
- [11] P. Bocciarelli and A. D'Ambrogio, "A BPMN extension for modeling non functional properties of business processes," in Proceedings of the 2011 Symposium on Theory of Modeling & Simulation: DEVS Integrative M&S Symposium, 2011, pp. 160-168: Society for Computer Simulation International.
- [12] Rodríguez and A. Caro, "Obteniendo Casos de Uso centrados en la Calidad de los Datos desde Procesos de Negocio descritos con BPMN," RISTI-Revista Ibérica de Sistemas e Tecnologías de Informação, no. 10, pp. 65-80, 2012.
- [13] M. Bobrowski, M. Marré, and D. Yankelevich, "A software engineering view of data quality," Proceedings of Second International Software Quality in Europe, 1998.
- [14] ISO/IEC. Software engineering-Software product Quality Requirements and Evaluation (SQuaRe) Data quality model, ISO/IEC 25012, pp. 1-13, 2008.
- [15] E. Verbo, I. Caballero, R. Perez, C. Calero, and M. Piattini, "MEPLAMECAL: A Methodology Based on ISO/IEC 15939 to Elaborate Data Quality Measurement Plans," *IEEE Latin America Transactions*, vol. 7, no. 3, pp. 361-368, 2009.
- [16] Kitchenham, "Procedures for performing systematic reviews," Keele, UK, Keele University, vol. 33, no. 2004, pp. 1-26, 2004.
- [17] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering—a systematic literature review," *Information and software technology*, vol. 51, no. 1, pp. 7-15, 2009.
- [18] G. Márquez, F. Osses, and H. Astudillo, "Review of architectural patterns and tactics for microservices in academic and industrial literature," *IEEE Latin America Transactions*, 2018.
- [19] K. Saeedi, L. Zhao, and P. R. F. Sampaio, "Extending BPMN for supporting customer-facing service quality requirements," in 2010 IEEE International Conference on Web Services, 2010, pp. 616-623: IEEE.
- [20] S. Goldner and A. Papproth, "Extending the BPMN Syntax for Requirements Management," in International Workshop on Business Process Modeling Notation, 2011, pp. 142-147: Springer.
- [21] Lodhi, V. Küppen, and G. Saake, "An extension of bpmn meta-model for evaluation of business processes," Scientific Journal of Riga Technical University. Computer Sciences, vol. 43, no. 1, pp. 27-34, 2011.
- [22] A. Urrutia, E. Chavez, R. Motz, and R. Gajardo, "An Ontology to Assess Data Quality Domains. A Case Study Applied to a Health Care Entity," *IEEE Latin America Transactions*, vol. 15, no. 8, pp. 1506-1512, 2017.
- [23] A. K. Veiga, A. M. Saraiva, and E. A. Cartolano, "Data quality control in biodiversity informatics: the case of species occurrence data," *IEEE Latin America Transactions*, vol. 12, no. 4, pp. 683-693, 2014.
- [24] H. Bringel, A. Caetano, and J. M. Tribolet, "Business Process Modeling Towards Data Quality: A Organizational Engineering Approach," in ICEIS (3), 2004, pp. 565-568.
- [25] S. Bagchi, X. Bai, and J. Kalagnanam, "Data quality management using business process modeling," in Services Computing, 2006. SCC'06. IEEE International Conference on, 2006, pp. 398-405: IEEE.
- [26] W. el Abed, "Data Governance: A Business Value-Driven Approach," in La Gouvernance Des Données: Une Approche De Valeur Conduite Par Les Métiers, 2009.
- [27] M. Heravizadeh, J. Mendling, and M. Rosemann, "Dimensions of Business Processes Quality (QoBP)," in Business Process Management Riga Workshops, 2009, pp. 80-91: Springer.
- [28] P. Soffer, "Mirror, mirror on the wall, can i count on you at all? Exploring data inaccuracy in business processes," in Enterprise, business-process and information systems modeling: Springer, 2010, pp. 14-25.
- [29] Duszak and J. Lewkowicz, "Publishing academic texts in English: A Polish perspective," Journal of English for Academic Purposes, vol. 7, no. 2, pp. 108-120, 2008.
- [30] BPTrends. (2018). BPTrends BPM Analysis, Opinion and Insight. Available: <http://www.bptrends.com>
- [31] SourceMedia. (2018). Information Management. Available: <http://www.information-management.com/>
- [32] Global Excellence. (2018). Global Excellence. Available: <http://www.global-excellence.com/>
- [33] Global Data Excellence. (2018). GlobalDataExcellence. Available: <https://www.globaldataexcellence.com>
- [34] Google. (2018). Google. Available: <https://www.google.com/>
- [35] T. C. Redman, "The impact of poor data quality on the typical enterprise," Communications of the ACM, vol. 41, no. 2, pp. 79-82, 1998.
- [36] Hernández Ortega, J. Martínez Jiménez, and J. DeHoyos, "Calidad de la información Web en la banca electrónica," in International Congress "Marketing Trends". Venice, 2008.
- [37] F. M. Guardia and N. R. Fuentes, "La naturaleza competitiva de la empresa banco-aseguradora," Estudios de economía aplicada, vol. 22, no. 2, pp. 377-378, 2004.
- [38] J. G. Walsh and J. M. Walsh, "Method and system for enterprise business process management," ed: Google Patents, 2005.
- [39] T. C. Redman, Data driven: profiting from your most important business asset. Harvard Business Press, 2008.

- [40] R. Clarke, "Data risks in the cloud," *Journal of theoretical and applied electronic commerce research*, vol. 8, no. 3, pp. 59-73, 2013.
- [41] W. Howard, "Data Quality Isn't Just a Data Management Problem," *Information Management*, vol. 17, no. 10, p. 16, 2007.
- [42] B. S. Black, "Requisitos legales e institucionales para el establecimiento de un mercado de valores sólido," *revista del mercado de valores*, p. 11, 2011.
- [43] M. Talalweh, "La divulgación de información financiera en Internet de las principales empresas españolas: variables económicas determinantes," *Gestión Joven*, no. 5, p. 2, 2010.
- [44] N. Berente, B. Vandenbosch, and B. Aubert, "Information flows and business process integration," *Business Process Management Journal*, vol. 15, no. 1, pp. 119-141, 2009.
- [45] Borek, A. K. Parlikad, P. Woodall, and M. Tomasella, "A risk based model for quantifying the impact of information quality," *Computers in Industry*, vol. 65, no. 2, pp. 354-366, 2014.
- [46] M. J. Eppler and P. Muenzenmayer, "Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology," in *IQ*, 2002, pp. 187-196. CiteSeer.
- [47] Falge, B. Otto, and H. Österle, "Data quality requirements of collaborative business processes," in *System Science (HICSS)*, 2012 45th Hawaii International Conference on, 2012, pp. 4316-4325: IEEE.
- [48] P. Lillrank, "The quality of information," *International Journal of Quality & Reliability Management*, vol. 20, no. 6, pp. 691-703, 2003.
- [49] C. R. d. A. Lima, J. M. d. A. Schramm, C. M. Coeli, and M. E. M. d. Silva, "Review of data quality dimensions and applied methods in the evaluation of health information systems," *Cadernos de saude publica*, vol. 25, no. 10, pp. 2095-2109, 2009.
- [50] R. Y. Wang, "A product perspective on total data quality management," *Communications of the ACM*, vol. 41, no. 2, pp. 58-65, 1998.
- [51] R. Liy, R. Montes, and R. Nakashima, "Propuesta de mejora para la cadena de suministro en el negocio de lubricantes," 2013.
- [52] L. Padilla, "Planeacion De Los Recursos De La Empresa: ERP," ed, 2006.
- [53] V. Évora Torres, "Procedimiento para la Gestión de la Calidad de Datos del Sistema Informativo Bancario," *Universidad Central "Marta Abreu" de Las Villas*, 2012.
- [54] L. Calderón Amaya and C. Rodríguez Monroy, "Un enfoque gerencial de factores críticos para el éxito de los sistemas de información en la Pyme metalmeccánica venezolana," *Contaduría y administración*, vol. 57, no. 1, pp. 79-102, 2012.
- [55] S. W. Tee, P. L. Bowen, P. Doyle, and F. H. Rohde, "Factors influencing organizations to improve data quality in their information systems," *Accounting & Finance*, vol. 47, no. 2, pp. 335-355, 2007.
- [56] Vázquez, "Past, present and future of public and social dimensions in Marketing conceptual development," *International Review of Public and Non Profit Marketing*, vol. 1, no. 1, pp. 9-34, 2004.
- [57] N. Gorla, T. M. Somers, and B. Wong, "Organizational impact of system quality, information quality, and service quality," *The Journal of Strategic Information Systems*, vol. 19, no. 3, pp. 207-228, 2010.
- [58] C. Guerra-García, I. Caballero, and M. Piattini Velthius, "A Survey on How to Manage Specific Data Quality Requirements during Information System Development," *Evaluation of Novel Approaches to Software Engineering*, pp. 16-30, 2011.
- [59] P. R. Correa and R. G. Cruz, "Una investigación empírica sobre los factores que afectan el éxito de los sistemas ERP en Chile," *Revista Ingeniería Informática*, Edición, no. 11, 2005.
- [60] H. D. Lee et al., "Prototype system to manage data on coloproctology surgery," *Journal of Coloproctology (Rio de Janeiro)*, vol. 31, no. 4, pp. 351-361, 2011.
- [61] Mosquera, D. Rincón, and M. G. Romero, "La organización basada en los sistemas de información," *Opción: Revista de Ciencias Humanas y Sociales*, no. 34, pp. 68-85, 2001.
- [62] M. Genero, J. Cruz-Lemus, and M. Piattini, "Métodos de investigación en ingeniería del software," *Editorial RA-MA: Madrid, Spain*, pp. 171-199, 2014.



Luis Ortega es Ingeniero Civil en Informática de la Universidad del Bío-Bío, Chile, y obtuvo su Maestría en Ciencias de la Computación de la misma universidad en 2017. Actualmente trabaja en la industria de software.



Angélica Caro es profesora jornada completa en el depto. de Ciencias de la Computación y Tecnologías de la Información de la Universidad del Bío-Bío, Chile. Recibió su PhD en Informática en la Universidad de Castilla-La Mancha, España. Sus intereses de investigación son ingeniería de software, arquitecturas empresariales y calidad de datos.



Alfonso Rodríguez es profesor jornada completa en el depto. de Cs. de la Computación y Tecnologías de la Información de la Universidad del Bío-Bío. Recibió su PhD en Cs. de la Computación de la Universidad de Castilla-La Mancha, España. Sus áreas de investigación son ingeniería de software, arquitecturas empresariales y procesos de negocio.



Ignacio Velásquez es Ingeniero Civil en Informática de la Universidad del Bío-Bío, Chile. Recibió su Maestría en Ciencias de la Computación de la misma universidad en 2017, donde actualmente es ayudante de investigación en proyectos de investigación sobre arquitecturas empresariales, tecnologías de información y gobierno electrónico.