

Hacia un Modelo de Madurez para la Gestión de Calidad de Datos en Inteligencia de Negocios

Alejandro Daza

Departamento de Ingeniería de Sistemas y Computación
Universidad Católica del Norte
Antofagasta, Chile
adu001@ucn.cl

Pablo de La Torre

Departamento de Ingeniería de Sistemas y Computación
Universidad Católica del Norte
Antofagasta, Chile
pdo002@ucn.cl

Vianca Vega Zepeda

Departamento de Ingeniería de Sistemas y Computación
Universidad Católica del Norte
Antofagasta, Chile
vvega@ucn.cl

Claudio Meneses Villegas

Departamento de Ingeniería de Sistemas y Computación
Universidad Católica del Norte
Antofagasta, Chile
cmeneses@ucn.cl

Abstract— Se presenta un Modelo de Madurez para la Gestión de la Calidad de Datos en BI (GCDBI). Esta propuesta fue desarrollada en base a los modelos IQM3 y TDQM que son utilizados para la evaluación y mejora de los datos. Además se usa como referencia el ciclo de vida de un proyecto de Inteligencia de negocios desarrollado por Kimball. El modelo propuesto consta de cinco niveles, donde cada nivel representa una etapa evolutiva en el proceso y la capacidad de gestión de la calidad de la información. Para mejorar la comprensión y facilitar la aplicación del modelo, se creó una guía de implementación, la cual comienza con la aplicación de un cuestionario que permite conocer cuál es el nivel de madurez actual de una organización. Una vez identificado el nivel de madurez, se proponen las actividades y procesos que se deben realizar para aplicar el modelo GCDBI y conseguir el avance de la organización a un nivel de madurez mayor.

Keywords— Inteligencia de Negocios; Gestión de Calidad; Modelos de Madurez; Mejora de Procesos; Calidad de Datos

I. INTRODUCCIÓN

Las empresas han experimentado cambios significativos, la mayoría de los cuales han sido impulsados por la tecnología. El resultado de estos cambios ha sido un exceso de información. Sin embargo, este exceso de información no ha resultado necesariamente en una organización más informada ni en una toma de decisiones más eficiente. En la actualidad las empresas están gestionando la información más que nunca y se están dando cuenta de los problemas de calidad que existen en sus empresas [1]. En consecuencia, abundan los problemas de Calidad de datos e Información (en adelante CDI) en Business Intelligence (BI), por lo cual la gestión CDI se hace muy importante.

Por esta razón se deben identificar los procesos de Producción de Información y el rol de los sistemas de información en el proceso (en adelante SI). Se puede conceptualizar un SI como la producción de información (como una empresa de manufactura), donde los datos, haciendo el papel de materias primas, son utilizados para producir un “producto de datos” siguiendo un proceso de

fabricación pre-establecido, normalmente implantado en el SI. De este producto de datos, se extrae un “producto de información” (en adelante PI) que se procesa por un usuario para el desarrollo de un proyecto en un escenario específico [1]. Se hace referencia a que se utiliza un producto de información para desarrollar un proyecto BI, en donde un usuario tiene que tomar una decisión táctica, estratégica y operativa. Su decisión, afectará de alguna manera, algunos de los procesos de negocio de la organización. Independiente de las capacidades del usuario, cuanto mayor es el nivel de CDI del producto, mayor será la calidad de BI.

En este artículo se presenta una propuesta de mejora para la calidad de los datos e información de las organizaciones, a través de la adaptación y complemento de los modelos IQM3 [2] y TDQM [2]. Mediante la aplicación de una encuesta, se define el nivel de madurez en BI que posee una organización. Una vez identificado el estado actual, la propuesta identifica un conjunto de actividades a desarrollar para alcanzar el siguiente nivel de madurez.

El artículo comienza con la descripción de los modelos utilizados como base para el desarrollo de la propuesta. A continuación se describen las técnicas de gestión de calidad de los datos estudiadas. En la cuarta sección se muestra el modelo de madurez adaptado para la propuesta. La quinta sección describe la guía para mejorar la gestión de calidad de los datos e información. La sección seis presenta el proceso de mejora desarrollado. En la séptima sección se describen los casos prácticos con los cuales se validó la encuesta para medir el nivel de madurez de las organizaciones. El artículo termina con las conclusiones y trabajo futuro.

II. PRELIMINARES

Los modelos de madurez han sido utilizados con éxito en muchas disciplinas, incluyendo el Modelo de Madurez de Gestión de Calidad de Información [2] (Information Quality Management Maturity Model en inglés, IQM3 en adelante) y el Modelo de Madurez Gestión de Calidad Total de Datos [2] (Total Data Quality Management en inglés, TDQM en adelante), que son utilizados en la evaluación y mejora de los datos. Basándose en la metodología de IQM3, se propone un

modelo de madurez similar que constará de cinco niveles, donde cada nivel representa una etapa evolutiva de la capacidad de Gestión de CDI, este modelo se llamará Modelo de Madurez para la Gestión de Calidad de Datos en BI (GCDBI en adelante). Al separar los objetivos de CDI sobre BI en varios niveles como en IQM3, será más fácil lograr obtener objetivos parciales de una manera incremental. El modelo de madurez además adapta la metodología de TDQM a CDI, alineando las etapas de TDQM con los niveles de madurez.

A continuación se explicarán los conceptos que influyen directamente en el modelo de madurez propuesto.

A. Calidad de datos e información

Los problemas de CDI cuestan a las organizaciones cientos de millones de pesos cada año. El impacto de un manejo pobre de CDI con inconvenientes operacionales y decisiones mal informadas provoca una total detención de la operatividad del negocio. Además, el manejo de CDI de forma inadecuada tiene serias implicaciones para la satisfacción del cliente, así como los costos operativos y financieros de la presentación de informes. Por lo tanto, hay una necesidad creciente en organizaciones por herramientas y métodos que puedan utilizarse para evaluar y mejorar CDI.

La definición común de CDI es “adecuado para el uso” [3] y es un problema crítico en la organización. Sin embargo, esto implica que el concepto de CDI es subjetivo, lo que significa que la información y datos con calidad considerada apropiada para un uso no poseen la suficiente calidad para otros usos. Por lo tanto, la evaluación de CDI puede ser una tarea difícil. Además, aunque la visión convencional de CDI ha significado “Precisión”, una serie de dimensiones u objetivos han sido identificados por Richard Wang [4], como se muestra en la Tabla I.

TABLA I. CATEGORÍAS Y DIMENSIONES DE CDI

Categoría CDI	Dimensión CDI
Específico	Precisión, Objetividad, Credibilidad, Reputación
Representativo	Interpretable, Fácil de entender, Representación Breve, Representación Consistente
Contextual	Relevancia, Valor adherido, Menos Tiempo, Integridad, Cantidad de Información
Accesible	Accesibilidad, Seguridad de Acceso

Como resultado, un problema de CDI se puede definir como cualquier dificultad que se encuentra a lo largo de una o más dimensiones de calidad que será arrastrado en la gran parte de los usos que se le dé a esa información.

B. IQM3

IQM3 es un modelo estructurado en niveles de madurez como CMMI [5]. Son cinco niveles de madurez: (1) Inicial, (2) Definido (3), Integrado, (4) Gestionado Cuantitativamente y (5) Optimizado. Cada uno se dirige a un objetivo específico de gestión de CDI [4]. Con el fin de obtener una eficiente gestión de CDI, varias Áreas Claves de Procesos (Key Process Areas en inglés, KPA en adelante) han sido identificadas. Cada KPA se centra en un objetivo técnico o gestión de CDI aunque algunos de ellos pueden

colocar el límite entre ambos aspectos. El conjunto de KPA técnicas está relacionada con subprocesos de fabricación o ingeniería y el conjunto de KPA de gestión está relacionado con el subproceso de Gestión CDI. Cabe destacar que se identifican actividades, que definen objetivos de CDI, por lo cual para cada actividad se identifican productos de entrada y salida, se proponen técnicas y herramientas, y se especifica quién y cuándo debe llevar a cabo dicha actividad.

La Figura 1 muestra la estructura de IQM3.

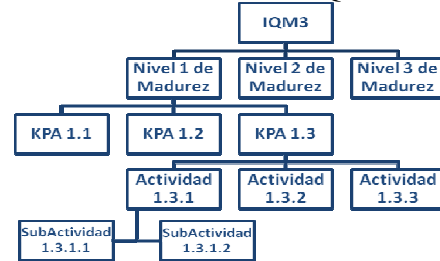


Figura 1. Estructura de IQM3

C. TDQM

TDQM ha adaptado el ciclo de mejora de la calidad de la empresa manufacturera para mejorar CDI. TDQM aplica el ciclo “Planificar, Hacer, Revisar, Actuar”, enfatizando que el mejoramiento de CDI depende de la continua retroalimentación de los procesos produciendo la información [4] como se aprecia en la Figura 2.

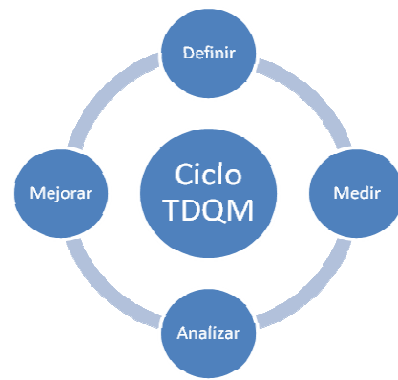


Figura 2. Ciclo de vida de TDQM

A continuación se define la adaptación del modelo TDQM para la CDI

1) *Definición de PI*: Se identifican los principales PI cuya CDI es básica para las necesidades de los usuarios. También se identifican aquellos requisitos de CDI según los cuales los usuarios deben determinar la calidad de los citados PI.

2) *Medición de CDI*: En esta fase el objetivo es identificar las métricas de CDI más eficientes para medir dicha calidad.

3) *Análisis de CDI*: Consiste en el desarrollo de un análisis dirigido por datos y por procesos para descubrir las causas de la pobre CDI en los PI seleccionados.

4) *Mejora de CDI*: El principal objetivo es diseñar e implementar estrategias para mejorar CDI de los PI sobre las causas encontradas en el paso anterior.

III. TÉCNICAS DE GESTIÓN DE CALIDAD DE DATOS

Existen 3 herramientas para mejorar la calidad de datos [2] que se han adaptado al modelo de madurez. Estas son:

A. *Data Profiling o Perfilado de datos*

El perfilado de datos es el proceso de examinar los datos que existen en las fuentes de origen de una organización y recopilar estadísticas e información sobre los mismos. El propósito de dichas estadísticas es:

- Determinar qué datos pueden ser usados para otros propósitos.
- Conseguir métricas de calidad de datos que incluyen si los datos cumplen los estándares de la organización.
- Reducir el riesgo de integrar información a nuevas aplicaciones dado que se conoce su estado.
- Realizar seguimiento de la calidad de datos.
- Entender problemas derivados de los datos en proyectos que hagan uso intensivo de los mismos.

B. *Data Cleansing o Limpieza de Datos*

La limpieza de datos es el proceso de detectar y corregir datos corruptos, incoherentes o erróneos. Después del proceso, la información será consistente con otros conjuntos similares de datos. Este proceso permite detectar entradas duplicadas, incompletas u otros problemas y establecer reglas para corregirlas. El objetivo no es borrar información sino mejorar la calidad de los datos construyendo un proceso de mejora continua.

C. *Data Auditing o Auditoría de Datos*

La auditoría de datos es el proceso de gestionar cómo los datos se ajustan a los propósitos definidos por la organización. Se establecen políticas para gestionar los criterios de datos para la empresa.

IV. MODELO DE MADUREZ DE GESTIÓN DE CALIDAD DE DATOS PARA INTELIGENCIA DE NEGOCIOS

El modelo de madurez pretende identificar los factores respectivos para cada nivel en relación a la Gestión de Calidad de Datos en BI. Además el modelo proporcionará una base para evaluar la Gestión de CDI en BI. Al aumentar los procesos de Gestión de CDI significará un mayor CDI y que al separar los objetivos de GCDBI como en IQM3, se logrará alcanzar los objetivos de CDI [4]. El modelo además define procesos específicos para mejoras en GCDBI, esto guiando a las organizaciones a niveles de madurez más alto. El modelo conceptual se valida a través de un cuestionario, el cual será explicado posteriormente.

En la figura 3 se ilustra el modelo de madurez en forma resumida. En esta figura, se muestran las KPA definidas para cada nivel de madurez.

La Tabla II muestra la descripción de los niveles de madurez. La Tabla III presenta los objetivos de cada KPA definida para los distintos niveles de madurez.



Figura 3. Modelo de Madurez de Gestión de Calidad de Datos para Inteligencia de Negocios

TABLA II. NIVELES DE MADUREZ

Nivel de Madurez	Descripción
Inicial	No hay conocimiento de problemas de CDI, además no se evalúa ni mejora CDI
Definido	Procesos básicos de un Proyecto BI se han establecido. Se tiene algún grado de conocimiento en problemas de CDI, pero no es mejorada
Integrado	La Organización ha desarrollado su propia GCDBI, las cuales son documentadas y utilizadas. Tiene un mejor conocimiento de los problemas de CDI, pero sólo reacciona a los problemas de CDI una vez ocurridos
Gestionado	La Organización ha desarrollado procesos propios del Proyecto BI, los cuales son documentados y utilizados, CDI es evaluado continuamente y mejorado
Optimizado	El Proyecto BI es cuantitativamente gestionado y las métricas son usadas para mejorar su rendimiento. Por lo cual el proceso sigue mejorando continuamente

V. GUÍA DE IMPLEMENTACIÓN DE GCDBI

La propuesta considera el modelo de Kimball [6] para los proyectos de BI. Se aplican mejoras en la calidad de los datos e información por medio de la utilización del modelo de madurez para obtener un producto de BI que realmente ayude a la toma de decisiones y no fracase en su intento de ser una ventaja estratégica dentro de la organización. Se plantea agregar una etapa antes de la planificación del proyecto de BI en sí, en donde se evalúe la calidad de los datos y la información que la organización ya tiene, para disminuir los riesgos y costos asociados a los datos de mala

calidad que pudieran ser utilizados dentro del proyecto y que son una de las principales causas de fracaso en el ámbito de BI.

TABLA III. OBJETIVOS DE LAS KPA

Nivel	KPA	Objetivo
Inicial	Dirección del Equipo de Trabajo a cargo de Gestión de CDI	Manejar un Equipo de Trabajo de Gestión de CDI a cargo de evaluar y mejorar las iniciativas.
Definido	Gestión de los requerimientos del usuario	Gestionar, administrar y rastrear los requerimientos de CDI del usuario para el Proyecto BI, PI y SI.
	Gestión del Proyecto BI	Administrar un proyecto definiendo todos los problemas del Proyecto BI
	Gestión de CDI	Definir cualitativa y cuantitativamente los problemas para las medidas de CDI de los componentes del Proyecto BI.
	Gestión de las fuentes de datos.	Identificar y caracterizar tanto las fuentes de datos como los objetivos para el Producto de Información.
	Desarrollo y Mantención del DW	Proveer un DW que pueda dar un soporte adecuado a todos los requerimientos de CDI de los usuarios.
Integrado	Verificación y Validación del Proyecto BI	Desarrollar un plan para validar y verificar el Proyecto BI y sus resultados.
	Gestión de Riesgos e Impacto de la baja CDI	Delimitar el alcance y todos los documentos relacionados con los riesgos que afectan a BI debido a la baja CDI.
	Gestión de Políticas de CDI a Nivel Organizacional	Desde un conocimiento de CDI Organizacional y de las necesidades con respecto a las necesidades de PI, el fin de este proceso es establecer y documentar políticas organizacionales.
	Gestión de Estandarización de CDI	Crear una cultura organizacional sobre CDI a través de lecciones aprendidas.
Gestionado	Planificación de Gestión de las mediciones del Proyecto BI.	Desarrollar planes para conseguir las medidas obtenidas en Gestión de CDI.
	Planificación de Automatización de Gestión de mediciones del Proyecto BI	Automatizar los planes en el modelo de madurez.
Optimizado	Análisis casual para la Gestión de la prevención de errores.	Identificar la raíz de las causas de la baja CDI.
	Desarrollo de Gestión Organizacional y de Información.	Mejorar el rendimiento del Proyecto BI.

El ciclo de vida del proyecto de BI con la nueva etapa se muestra en la Figura 4.

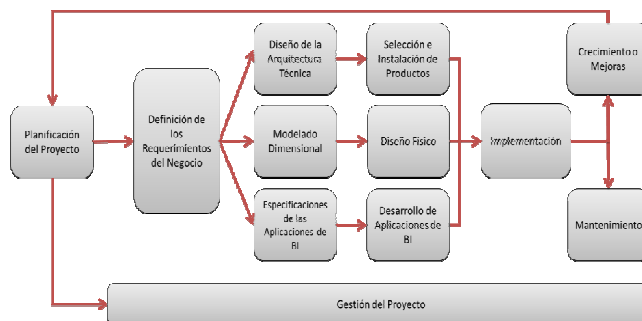


Figura 4: Ciclo de Vida de un proyecto de BI.

A. ¿Qué nivel de madurez en la calidad de datos tiene la empresa?

Lo primero que se debe realizar para mejorar la calidad de los datos y la información dentro de la organización es saber en qué nivel o grado de gestión se encuentra la organización. Lo anterior con el objetivo de tener el punto de partida y saber qué acciones tomar. Para conseguirlo, se ha generado un cuestionario que ayudará a los gerentes y dueños de la empresa a saber realmente cómo es la gestión de su información antes de gastar tiempo y recursos en la implantación de un proyecto de BI en la organización.

B. Tipo de Estudio

El tipo de estudio en que se enfocó el cuestionario es en el estudio descriptivo, ya que se busca definir las propiedades, características y calidad de datos e información, así como los procedimientos, tecnologías utilizadas para los procesos más importantes relacionados con los datos. De igual manera, el estudio tiene un enfoque cuantitativo, para poder analizar los resultados de los cuestionarios que se aplicarán a las organizaciones. Así se permite establecer con exactitud el nivel de madurez en la calidad de datos e información en que la empresa se encuentra.

C. Diseño de cuestionario

El cuestionario fue realizado de acuerdo a las necesidades de la investigación. Cabe señalar que se buscó con énfasis personas con conocimientos tanto en CDI como en BI, por ejemplo, que las personas del departamento de informática fuesen las encargadas de responder los cuestionarios, debido a que ellos cuentan con la información necesaria para poder responderlos. La Tabla IV muestra el tipo de preguntas consideradas.

TABLA IV. TIPOS DE PREGUNTAS INCLUIDAS EN LA ENCUESTA

Tipo pregunta	Respuestas
Dicotómica	Sí
	No
Escala Likert	Sí, implementada, documentada y regularmente revisada
	Sí, implementada y documentada
	Sí, implementada
	Sí, documentada
	No

La Escala Likert mide la actitud de la empresa frente a la Gestión de Calidad de datos e Información, a través de una serie de afirmaciones por parte de la organización [1].

La encuesta se forma de 40 preguntas, con una puntuación máxima de 124 puntos, los cuales son divididos por nivel de madurez. La Tabla V muestra la relación entre el puntaje obtenido y el nivel de madurez identificado.

TABLA V. RELACIÓN PUNTAJE – NIVEL DE MADUREZ

Rango de puntaje	Nivel de Madurez
0 – 48	Nivel 1
49 – 78	Nivel 2
79 – 103	Nivel 3
104 – 118	Nivel 4
119 - 124	Nivel 5

D. *¿Qué hacer ahora que se sabe el nivel en que se encuentra la organización?*

Después de realizar y evaluar el cuestionario, se conoce el nivel de madurez en la calidad de datos e información. Por lo tanto, lo que se debe realizar dependerá del grado de madurez. En la sección VI se presentan las actividades a realizar para mejorar la calidad, divididas por niveles.

VI. PROPUESTA PARA LA MEJORA DE LA CALIDAD DE LOS DATOS E INFORMACIÓN

A. Nivel Inicial

La GCDBI se encuentra en el nivel inicial cuando no se ha realizado ningún esfuerzo por mejorar CDI. El siguiente KPA debería ser implementado:

1) *Dirección del Equipo a cargo de la Gestión de CDI.* Esta KPA está dirigida a constituir y organizar un equipo a cargo de la Gestión de CDI que tiene la suficiente responsabilidad y experiencia con Proyectos BI para apoyar y mejorar las iniciativas. Este equipo debe guiar las iniciativas acorde a las ideas de la organización, políticas y predicciones acerca de CDI. También debe identificar roles y responsabilidades. El equipo debe ser multidisciplinario y tener habilidades gerenciales, las cuales deben ser usadas para estimular a toda la organización a alinearse bajo las políticas de CDI. Técnicas o herramientas para gestionar recursos humanos, trabajo en equipo y entrenamiento deben ser usados en este KPA. La Tabla VI muestra las actividades y tareas a desarrollar para la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA VI. ACTIVIDADES PARA IMPLEMENTAR KPA DIRECCIÓN DEL EQUIPO A CARGO DE LA GESTIÓN DE CDI

Actividad	Tareas
Determinar equipo de gestión de calidad de datos	Identificar los roles necesarios para gestionar la calidad de datos e información
	Identificar y asignar funciones al personal más apto
Determinar el ambiente de trabajo	Establecer un código o reglas de trabajo
	Definir los canales efectivos para comunicar las actividades
	Seleccionar las herramientas y técnicas más aptas para cada KPA del Modelo de Madurez

B. Nivel Definido

La GCDBI se encuentra en el nivel 2 cuando éste ha sido definido y planeado. Esto implica identificar todos sus componentes y las relaciones entre ambos. Una vez que el Proyecto BI ha logrado este nivel, se puede asegurar que el Proyecto BI es repetible, y los mismos resultados pueden ser obtenidos cada vez que el Proyecto BI es ejecutado. Para mejorar este objetivo, los siguientes KPA necesitan ser implementados:

1) *Gestión de Requerimientos del Usuario.* Tres tipos de requerimientos deben ser identificados, documentados e informados para cualquier escenario de CDI: aquellos relacionados con el PI, aquellos relacionados al Proyecto BI y aquellos relacionados a CDI.

Todos los documentos obtenidos son el punto de inicio para adaptar el actual modelo de datos y modelo de negocios para el Proyecto BI.

La Tabla IX muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA VII. ACTIVIDADES PARA IMPLEMENTAR KPA GESTIÓN DE REQUERIMIENTOS DEL USUARIO

Actividad	Tareas
Gestionar los requerimientos de los usuarios para el proyecto BI	Recolectar y documentar los requerimientos de los usuarios para GCDBI
	Gestionar los cambios en el Proyecto BI en cuanto a la especificación de los requerimientos de los usuarios
	Validar el Proyecto BI en cuanto a la especificación de los requerimientos de los usuarios
Gestionar los requerimientos de Usuario para PI	Recolectar y documentar los requerimientos de los usuarios para PI
	Gestionar los cambios del PI en cuanto a la especificación de los requerimientos de los usuarios
	Validar el PI en cuanto a la especificación de los requerimientos de los usuarios
Gestionar los requerimientos de los usuarios para CDI	Recolectar y documentar los requerimientos de los usuarios para CDI
	Gestionar los cambios de CDI en cuanto a la especificación de los requerimientos de los usuarios
	Validar la CDI en cuanto a la especificación de los requerimientos de los usuarios

2) *Gestión de BI.* El principal objetivo de este KPA es crear e implementar un plan acorde a los esfuerzos y recursos cuando se diseña, programa e implementa un Proyecto de BI. El resultado de este KPA es un documento que describe una agenda de actividades y un presupuesto para la implementación del Proyecto BI. Este documento también debe incluir una descripción del alcance del Proyecto BI, tanto en sus fases técnicas y gerenciales como en sus actividades. Además, el plan puede incluir una estimación de los costos y el recurso humano necesario para alcanzar los objetivos.

La Tabla XIII muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA VIII. ACTIVIDADES PARA IMPLEMENTAR KPA GESTIÓN DE BI

Actividad	Tareas
Definir los componentes del Proyecto BI y sus relaciones	Definir el alcance del Proyecto BI
	Definir el ciclo del Proyecto BI
	Definir el desarrollo técnico del proyecto
	Gestionar la especificación de requerimientos de los usuarios para el modelo de negocios y el modelo de datos
	Estimar los recursos necesarios y la frecuencia de uso
	Estimar el tiempo necesario para cada actividad
	Gestionar los riesgos del proyecto
Definir e implementar el proyecto de implantación del Proyecto BI	Crear un calendario para las actividades necesarias
	Establecer el presupuesto del proyecto de BI
	Definir el proyecto de implantación del Proyecto BI
	Determinar la viabilidad del proyecto
	Diseñar el modelo de procesos
	Diseñar el modelo de datos
	Ejecutar la implementación del proyecto

3) *Gestión de CDI*. Hay que juzgar cuantitativamente CI en orden a cuantificar qué tan bueno es un escenario y cuánto debe ser mejorado. Para ello hay que identificar las dimensiones de CDI y seleccionar sus medidas. Es importante que las medidas se determinen en los diferentes casos de uso.

La Tabla XVI muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA IX. ACTIVIDADES PARA IMPLEMENTAR KPA GESTIÓN DE CDI

Actividad	Tareas
Identificar las dimensiones de CDI	Identificar las dimensiones de CDI para las fuentes de datos y objetivos de los productos de información
	Identificar los CDI para los trabajadores
	Identificar los CDI para cada componente del Proyecto BI
Identificar las métricas para cada dimensión de CDI	Identificar las métricas de las dimensiones de CDI para las fuentes de datos y objetivos de los productos de información
	Identificar las métricas de CDI para los trabajadores
	Identificar las métricas para cada componente del Proyecto BI

4) *Fuente de Datos y Objetivos de la Gestión de Datos*. En la actualidad, debido a las nuevas tecnologías, es más fácil obtener datos de cualquier fuente. Esto puede dar paso a nuevos problemas como tener diferentes valores para el mismo ítem viniendo de diferentes fuentes. Por lo cual todo procedimiento debe ser identificado, documentado e informado a quien corresponda, en orden a establecer mecanismos para evadir futuros problemas. La Tabla XIX muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA X. ACTIVIDADES PARA IMPLEMENTAR KPA FUENTE DE DATOS Y OBJETIVOS DE LA GESTIÓN DE DATOS

Actividad	Tareas
Describir y caracterizar la fuente de datos	Identificar las fuentes de datos candidatas
	Seleccionar las fuentes candidatas más idóneas
	Definir el proceso de adquisición de datos
Describir y caracterizar los objetivos de los productos de información	Identificar los PI candidatos
	Seleccionar los PI más adecuados
	Definir el proceso de entrega de datos

5) *Desarrollo y Mantenimiento del DW*. Este KPA está enfocado en establecer el modelo de datos en ayuda de las actividades del Proyecto BI. Teniendo CDI como el objetivo, es necesario que el DW pueda soportar todos los tipos de requerimientos de CDI de los usuarios. Es por ello que es fundamental definir un proyecto para adquirir, desarrollar o mantener un DW, destacando que el DW resultante debe ser válido para todo, o al menos el SI del Proyecto BI. La Tabla XXII muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA XI. ACTIVIDADES PARA IMPLEMENTAR KPA DESARROLLO Y MANTENCIÓN DEL DW

Actividad	Tareas
Definir adquisición, desarrollo y/o mantenimiento de proyectos de DW	Definir la viabilidad de alternativas
	Elegir la mejor alternativa
Ejecutar Proyectos	Ejecutar el mejor plan

C. Nivel de integración

La GCDBI se encuentra en el nivel 3, cuando el nivel 2 se ha mejorado y también se han alineado los requerimientos de CI a la organización a través de políticas y estándares de BI. Esto contempla el esfuerzo de estandarizar las diferentes lecciones aprendidas a través de diferentes iniciativas de CI, reuniendo y documentando políticas y estándares de CI. Asegurando que ellas estarán disponibles en nuevos proyectos de BI. El fin de este nivel es proveer un mecanismo para implementar soluciones organizacionales específicas a problemas de CI para evadir errores anteriores y permitir mejores proyectos de BI para el futuro. Los siguientes KPA deben ser implementados:

1) *Componentes de Validación y Verificación de PI y del Proyecto BI*. Los componentes del Proyecto BI como modelos de datos o modelo de negocios, deben ser verificados y validados para corregir errores y/o discordancias con las especificaciones de requerimiento de usuario para BI y para CSI. La Tabla XXV muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA XII. ACTIVIDADES PARA IMPLEMENTAR KPA COMPONENTES DE VALIDACIÓN Y VERIFICACIÓN DE PI Y DEL PROYECTO BI

Actividad	Tareas
Diseñar el plan de verificación y validación	Diseñar el plan de verificación y validación para los componentes del Proyecto BI y PI
Ejecutar el plan de verificación y validación	Ejecutar el plan de verificación y validación del Proyecto BI

2) *Riesgos y Gestión de la baja CDI.* En este punto el equipo de Gestión de CDI cuantifica la extensión del impacto de los riesgos, causados por los niveles bajos de CDI sobre el rendimiento del Proyecto BI. Se debe elaborar un plan de contingencia para cada uno de los riesgos identificados, para minimizar sus efectos sobre el Proyecto BI. Los planes de contingencia deben ser evaluados para determinar si son factibles. Si no, es necesario asumir y estimar qué consecuencias tendrán los riesgos. Si es posible, las acciones deben ser ejecutadas para modificar el Proyecto BI, evadiendo esos riesgos o soportar su impacto.

La Tabla XXVIII muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA XIII. ACTIVIDADES PARA IMPLEMENTAR KPA RIESGOS Y GESTIÓN DE LA BAJA CDI

Actividad	Tareas
Estimar el impacto de la baja CI	Buscar e identificar los riesgos y sus fuentes
	Estimar el impacto del riesgo en el Proyecto BI
Definir planes de contingencia para minimizar los efectos de los riesgos	Definir las soluciones para evitar los riesgos
	Definir los planes de acción para mitigar los riesgos
	Definir planes de contingencia para evitar o mitigar los riesgos
	Estimar la viabilidad de los planes de contingencia
	Implementar los planes de contingencia

3) *Gestión de Políticas Organizacionales de CDI.* Todas las lecciones aprendidas a través de experiencias anteriores en CDI deberían ser reunidas, informadas y transmitidas a la base de conocimientos organizacional. Una política organizacional es un conjunto de reglas que puede ser aplicada a cualquier acción de la organización para trabajar bajo ciertos criterios. De igual manera, las políticas organizacionales de CDI son una forma de universalizar varios problemas, reuniéndolos para manejar dimensiones de CDI, riesgos de CDI y cómo modificar modelos de datos y procesos para ayudar a las mejores prácticas de CDI.

Las organizaciones pueden decir que tienen una cultura de CDI cuando todos sus procesos, relacionados o no con la información y la Gestión de CDI, toma en cuenta problemas de CDI. Estas políticas deben también observar las mejores prácticas relacionadas con la gestión de procesos de datos para los principales elementos de este KPA.

La Tabla XXXI muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA XIV. ACTIVIDADES PARA IMPLEMENTAR KPA GESTIÓN DE POLÍTICAS ORGANIZACIONALES DE CDI

Actividad	Tareas
Diseñar las políticas organizacionales de CDI	Diseñar políticas para las dimensiones y métricas de CDI
	Diseñar políticas para los modelos de datos y de procesos
	Diseñar políticas para el riesgo en CDI
	Diseñar políticas para las técnicas de análisis de CDI

4) *Gestión de Estandarización de CDI.* Solamente por la incorporación de la experiencia de Gestión de CDI más reciente, el Proyecto de BI mejoraría mucho más. Si en los anteriores KPA, las lecciones aprendidas de experiencias fueron trasladadas dentro de las políticas organizacionales, el principal objetivo de este KPA es incorporar las mejores prácticas coleccionadas por la organización dentro del Proyecto de BI. Es importante darse cuenta cómo los modelos de datos y de procesos pueden ser modificados a través de la incorporación de requerimientos específicos de los usuarios. Después de esto, se deben propagar a SI los nuevos cambios.

La Tabla XXXIII muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA XV. ACTIVIDADES PARA IMPLEMENTAR KPA GESTIÓN DE ESTANDARIZACIÓN DE CDI

Actividad	Tareas
Elegir los estándares para CDI	Elegir estándares de CDI
Elegir las políticas organizacionales para CDI	Elegir las políticas organizacionales para CDI
Revisar y finalizar los requerimientos de los usuarios de acuerdo a los estándares y políticas creados	Revisar y actualizar los requerimientos de los usuarios para satisfacer los nuevos estándares de CDI
	Revisar y actualizar requerimientos de los usuarios para satisfacer las políticas organizacionales para CDI
	Gestionar los cambios y actualizaciones de requerimientos de los usuarios

D. Nivel de Gestión

El Proyecto BI se encuentra en el nivel de Gestión, cuando además de integrar algunos planes de medición también éstos han sido desarrollados, implementados y automatizados para obtener valores para el diseño de medidas. La principal ventaja es la capacidad de tener un cumplimiento cuantitativo para que el rendimiento del Proyecto BI se prolongue en un período que permanezca constante en cuanto a términos de variación y estabilidad a través de un conjunto de medidas confiable. En la medida que es posible, el proceso de medición debe ser

automatizado para obtener más medidas confiables. Por ello, es necesario automatizar los planes de medida.

1) *Plan de Gestión de Medidas CDI*. El objetivo de esta KPA es elaborar planes de medida, tomando en cuenta los factores que afectan los procesos de medición, para obtener medidas que serán usadas para revisar la conformidad del Proyecto BI y PI en las especificaciones. Finalmente, es importante delimitar cómo y a quiénes deben ser presentados los resultados de las medidas.

La Tabla XXXVII muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA XVI. ACTIVIDADES PARA IMPLEMENTAR KPA PLAN DE GESTIÓN DE MEDIDAS CDI

Actividad	Tareas
Definir los planes de medición para cada métrica	Identificar el contexto de cada métrica
	Definir el plan específico de medición para cada métrica
Ejecutar los planes de medición	Ejecutar los planes de medición

2) *Plan de Gestión de Automatización de CDI*. Su objetivo es estudiar todos los errores relativos a la automatización de los procesos de medida, determinando si esto es posible. En el caso que lo sea, se deben tomar en cuenta otros problemas relacionados, dónde y cómo almacenar los resultados de las medidas. Las opciones son: en el mismo DW (que implica modificar el esquema del modelo de datos para soportar nuevos datos) o en otro DW (lo que implica definir el DW y los procesos para intercambiar los nuevos datos).

La Tabla XL muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA XVII. ACTIVIDADES PARA IMPLEMENTAR KPA PLAN DE GESTIÓN DE AUTOMATIZACIÓN DE CDI

Actividad	Tareas
Definir las situaciones en que los planes de medición pueden ser automatizados	Identificar las métricas que pueden ser automatizadas
Aplicar mediciones automatizadas	Modificar los modelos de proceso y de datos con los planes de automatización de métricas

E. Nivel optimizado

El Proyecto BI se encuentra en el nivel Optimizado cuando las medidas obtenidas son usadas para implementar continuamente mejoras al Proyecto BI. Esto puede ser para eliminar errores, o proponer e implementar varias mejoras para el proyecto. Es importante destacar que este KPA está orientado hacia el mejoramiento del Proyecto BI por sí solo.

1) *Análisis Casual para la Gestión de Prevención de Errores*. Desde el estudio de los resultados de las medidas, algunos diagramas típicos de calidad, técnicas y herramientas pueden ser aplicadas para detectar errores de

CDI y así, identificar sus causas. Las conclusiones obtenidas deben proveer una base para permitir un correspondiente proceso de mantención para el Proyecto BI, cuyo propósito es eliminar los errores encontrados en los recursos afectados.

La Tabla XLIII muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA XVIII. ACTIVIDADES PARA IMPLEMENTAR KPA ANÁLISIS CASUAL PARA LA GESTIÓN DE PREVENCIÓN DE ERRORES

Actividad	Tareas
Crear informes con errores encontrados	Encontrar e identificar defectos en el Proyecto BI
Analizar para identificar fuentes defectuosas	Realizar un estudio de las causas de los errores identificados
Diseñar soluciones para evitar y eliminar fuentes defectuosas	Corregir los defectos del modelo de procesos
Implementar soluciones para evitar y eliminar fuentes defectuosas	Implementar y validar la solución propuesta

2) *Gestión de Innovación y Desarrollo Organizacional*. Esta KPA está destinada a la implementación del concepto de mejoramiento continuo del Proyecto BI. Los resultados obtenidos de la medición serán usados para mejorar el proyecto en cuanto a alcanzar rendimientos más altos, mayor tiempo planeado de ejecución o costos más bajos en términos computacionales o de recurso humano. Para realizar los correspondientes análisis, se puede identificar cómo y dónde debe ser re-estructurado el Proyecto BI para obtener mejores valores para las mediciones.

La Tabla XLVIII muestra las actividades y tareas a desarrollar para conseguir la implementación de la KPA. Para mayores detalles respecto a estas actividades ver [7].

TABLA XIX. ACTIVIDADES PARA IMPLEMENTAR KPA GESTIÓN DE INNOVACIÓN Y DESARROLLO ORGANIZACIONAL

Actividad	Tareas
Crear informes destacando posibles mejoras	Identificar posibles mejoras en el Proyecto BI
Analizar datos para mejorar la organización	Realizar un estudio para identificar posibles mejoras
Diseñar propuesta para la mejora de CI	Reingeniería del modelo de negocios para mejorar el rendimiento
Implementar soluciones para la mejora de temas seleccionados dentro de GCDBI	Implementar y validar las soluciones propuestas

VII. VALIDANDO LA ENCUESTA

Con el afán de saber si la encuesta, al ser aplicada, realmente entrega el estado de la organización en el tema de Gestión de Calidad de Datos e Información y si los resultados se ajustan a los diferentes niveles del modelo de madurez, se aplicó en varias organizaciones que están implementando o han implementado Inteligencia de Negocios. Si bien la muestra no es demasiado extensa, y sólo se aplicó a una persona por organización, sirve para - en

cierta forma - validar que los resultados que se obtienen al aplicar la encuesta muestran la realidad de la organización.

A. Resultados obtenidos

La encuesta se aplicó en tres organizaciones, una de ellas en el sector gran minería (empresa A), con aproximadamente 10000 trabajadores entre personal propio y contratistas. La segunda corresponde a la mediana minería, empresa con aproximadamente 1500 trabajadores entre personal contratado directamente por la empresa y empresas contratistas (empresa B). La última pertenece al sector de Educación Superior (empresa C), esta empresa cuenta con alrededor de 800 trabajadores.

Aplicación en el área de Educación Superior (empresa C)

El proyecto que se lleva a cabo dentro de la universidad busca poder Asesorar y apoyar técnicamente los procesos internos de planificación, acreditación, entre otras funciones académicas. Para facilitar el cumplimiento y la toma de decisiones relacionadas con los objetivos del Plan estratégico, el proyecto es liderado por el Director de Análisis y Gestión de la institución, quien fue la persona que respondió la encuesta.

El panorama al que se pudo llegar en las reuniones previas a la aplicación de la encuesta, fue que si bien la organización gestiona correctamente los procesos relacionados con la parte operacional y administrativa, el problema se genera al tratar de unificar toda esta información para tomar decisiones estratégicas y de gestión, ya que tienen muchas fuentes de datos que no están debidamente reguladas, esto quiere decir que no se conoce bien quiénes son los encargados de ciertos datos, no se tienen protocolos o estándares de trabajo, tampoco existe la documentación para los procesos encargados de vincular las diferentes fuentes de datos (planillas Excel, bases de datos, sistemas transaccionales). Lo que ha llevado a tener muchos problemas con los procesos ETL dentro del proyecto, porque los datos son de poca calidad, y los resultados tienden a ser poco confiables y se deben hacer muchas revisiones. Esto a priori indica que la organización está en el nivel inicial del modelo de madurez de la organización. Luego de realizada la encuesta, se obtuvo un puntaje de 32 puntos de un total de 124, con lo cual se confirma el nivel de madurez de la calidad de datos de la organización, el cual corresponde al nivel inicial.

Aplicación en la Gran Minería (empresa A)

En este caso la empresa es responsable de explorar nuevos terrenos mineros, incluyendo la identificación y adquisición con el propósito de emprender nuevos proyectos de exploración. Todas estas actividades están soportadas a través de una gestión de seguridad y riesgos, gestión de medio ambiente, gestión en proyectos, gestión de la información, finanzas, recursos humanos y otros.

Como se observa la gestión de la información es un factor crítico al momento de implementar un proyecto, en especial un proyecto orientado a la toma de decisiones como lo fue la adquisición e implementación de SAP en su organización. La encuesta en este caso fue respondida por el

Information Management Service Delivery Coordinator MinExAmerica. De un total de 124 puntos, la empresa obtuvo un puntaje de 111, lo que sitúa a la organización en el nivel 4. Esto significa que existen planes de medición de calidad de datos e información, los cuales son valorizados para analizar el rendimiento del Proyecto BI. Cabe destacar que este nivel es el más difícil de alcanzar por las organizaciones y es el adecuado para asegurar la calidad de datos e información en Proyectos BI, ya que en este nivel CDI es continuamente evaluado y mejorado.

Aplicación en Mediana Minería (empresa B)

El proyecto que se lleva a cabo dentro de la minera es la Implementación de un sistema web administrativo que permite la inscripción del personal a diferentes capacitaciones impartidas tanto por la empresa como por externos. Este proyecto está a cargo de un Analista de Sistemas de la organización, quien respondió la encuesta.

Según la encuesta, de un total de 124 puntos, la empresa obtuvo un puntaje de 101, por lo cual la organización se encuentra en el nivel 3. Esto implica que la organización ya ha desarrollado su propia GCDBI, las cuales son utilizadas y documentadas, además ya se relaciona con los problemas de CDI. Por lo cual si quiere mejorar su nivel de madurez, se tienen que verificar y validar los componentes del Proyecto BI, además de estandarizar las políticas de CDI a nivel organizacional, ya que sólo reacciona a los problemas de CDI una vez ocurridos.

B. Comparación de los resultados de las organizaciones

Los resultados muestran que la encuesta desarrollada en el presente estudio es un instrumento con suficiente validez y confiabilidad para medir el grado de gestión de calidad de datos e información en Proyectos BI, ya que los resultados obtenidos concuerdan con la realidad de las organizaciones.

Del universo total de encuestados, el 66% obtuvo la máxima puntuación, demostrando que al menos 2 organizaciones coinciden en la importancia de la valoración de la gestión de calidad de datos e información en Proyectos BI. Como se puede apreciar en la Figura 5, existe una significativa diferencia, entre las organizaciones respecto a la implementación de un Proyecto BI. Si bien entre la empresa A y B la diferencia es mínima, al comparar éstas con la empresa C, se aprecia la precaria GCDBI que tiene esta última.

Algunas razones que pueden explicar esto son:

- Las dos organizaciones con mayor puntaje disponen de un equipo de gestión de calidad de datos e información.
- Como se dispone de un equipo de gestión de calidad de datos, se han estandarizado los procesos para la obtención de datos críticos para la gestión y análisis.
- Las organizaciones que alcanzaron mejores puntajes llevan más de 3 años trabajando en el mejoramiento de la calidad de la información y con proyectos de Inteligencia de Negocios ya implementados.
- La de menor puntaje lleva menos años trabajando y es su primer proyecto de inteligencia de negocios, por lo cual recién se están detectando los problemas asociados a la falta de madurez en la gestión de la información.

La organización con menor puntaje tiene problemas de fiabilidad con los datos que manejan, ya que algunos datos son erróneos y las fuentes de datos no son confiables. Esto se debe principalmente a la falta de gestión y estandarización de los procesos para la generación y obtención de datos.

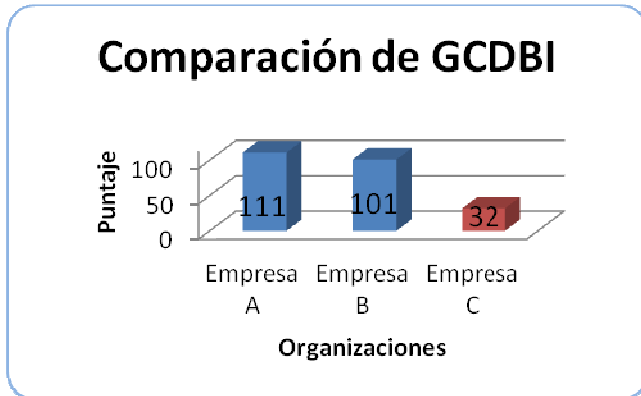


Figura 5. Comparación de la GCDBI en 3 organizaciones.

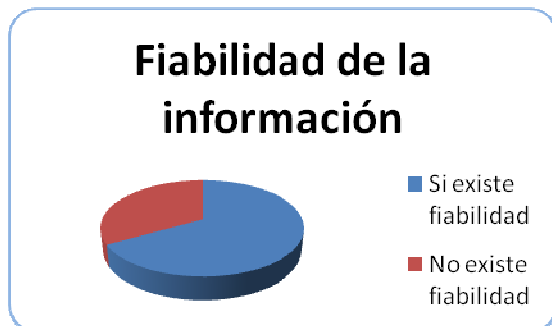


Figura 6. Gráfico de la fiabilidad de la información

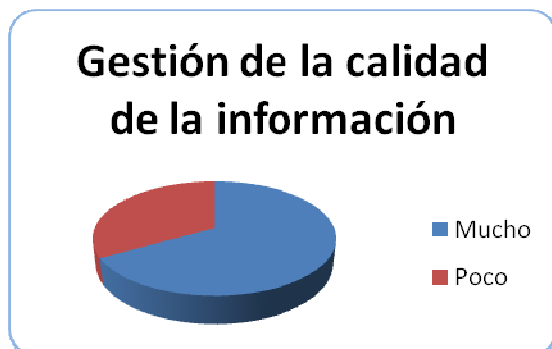


Figura 7. Calidad de la información de los SI

Un punto importante que se puede apreciar es que las tres organizaciones tienen procedimientos estandarizados, documentados de cómo se maneja la información dentro de los SI y cómo se realizan las vinculaciones entre las diferentes fuentes de datos. Pero sólo las con más alto puntaje monitorean y controlan la calidad de la información que se genera y se ingresa a los SI, teniendo claramente definidos los propietarios de los datos, y realizan auditorías e inspecciones con el fin de mantener y mejorar la calidad de la información que se utiliza.

Según los datos obtenidos, se observa que las tres organizaciones valoran los medios de información como importantes e incluso críticos para la gestión de su negocio, pero sólo las con mayor puntaje han visto la importancia de tener estándares y políticas que ayuden a mejorar la Calidad de la Información.

Aunque la encuesta no niega, ni afirma inequívocamente la presencia de calidad de datos e información, permite detectar a quienes hacen uso de una eficiente gestión de calidad de datos e información para asegurar que las decisiones tomadas, basadas en la información que tienen, les permite asegurar la obtención de ventajas competitivas.

VIII. CONCLUSIONES

Se puede concluir que BI permite analizar los planes y resultados de una empresa y contribuye a un mejor conocimiento de ésta. Ofreciendo una visión de los aspectos que funcionan de forma satisfactoria e identifica las áreas conflictivas sin tener un costo en tiempo y dinero. Puede utilizarse para reconocer tanto oportunidades como problemas potenciales y alertar a la organización cuando se encuentre en situaciones problemáticas.

Los datos precisos y oportunos utilizados en las aplicaciones de BI son fundamentales para el correcto funcionamiento de muchas organizaciones. Sin ellos, los sistemas de BI pueden suponer una amenaza que debilite la capacidad de las organizaciones para evaluar el estado real de la organización y al tomar las medidas adecuadas para dirigir sus empresas y competir de forma eficaz.

Una implantación satisfactoria de la calidad de los datos a nivel empresarial permite a las organizaciones maximizar el retorno de las inversiones en BI consiguiendo una ventaja competitiva y el liderazgo en el mercado.

La principal contribución de esta guía es la estructura de los niveles de madurez de calidad de información para el Proyecto BI, haciendo más fácil alcanzar los objetivos en forma gradual, mejorando la calidad de los datos e información a través de su gestión de una manera efectiva. Si bien, en el presente documento se habla de cómo aumentar la calidad de datos e información a la hora de implementar un Proyecto BI, con el fin de disminuir riesgos, existen otros factores críticos de errores, como los usuarios o los proveedores, es por ello que se puede adoptar esta metodología como el complemento de una guía que abarque por completo los errores críticos a la hora de implementar un Proyecto BI.

En relación a la encuesta, los resultados obtenidos han sido considerados válidos tanto por los propios encuestados, como por los autores. Dicha afirmación se basa en observaciones objetivas y características de cada empresa encuestada. La empresa A, cuyo resultado fue el nivel 4, cuenta con un equipo bien establecido dedicado a la implementación de aplicaciones de BI. Ha desarrollado numerosas aplicaciones de este tipo, y considera un presupuesto establecido para las actividades de BI. En cuanto a la empresa B, no existe tanta experiencia en proyectos BI como en la empresa A. El recurso humano dedicado a los proyectos de BI también es menor. Finalmente la organización C que fue caracterizada como en nivel inicial,

recién se está comenzado a realizar esfuerzos aislados para generar aplicaciones de BI.

Como trabajo futuro se puede repetir la aplicación de la encuesta, ampliando el universo de encuestados para incorporar tanto a usuarios de la información como a quienes están encargados de generar las aplicaciones BI.

IX. NOMENCLATURA

BI	Business Intelligence
CDI	Calidad de Datos e Información
CMMI	Capability Maturity Model Integration
DW	Data warehouse
ETL	Extract, transform and load
GCDBI	Gestión de la Calidad de Datos en BI
IQM3	Information Quality Management Maturity Model
KPA	Key Processes Area
PI	Producto de Información
SI	Sistemas de Información
TDQM	Total Data Quality Management

REFERENCIAS

- [1] Al-Hakim, L: "Information Quality Function Deployment"; Proc. of the Ninth International Conference on Information Quality, Cambridge, MA, USA, 2004.
- [2] Batini, C. and Scannapieco, M.: "Data Quality: Concepts, Methodologies and Techniques"; Heidelberg, Berlin, 2006
- [3] Caballero, I., Verbo, E. M. Calero, C: "A Data Quality Measurement Information Model based on ISO/IEC 15939"; Cambridge, MA, 2007
- [4] Richard Wang, Craig Fisher, Introduction to Information Quality, Information Systems Marist College, 2006
- [5] Mary Beth Chrissis, CMMI: Guía para la integración de procesos y la mejora de productos. Editorial Pearson, 2009
- [6] R. Kimball y M. Ross, The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence, Wiley Publishing, Inc., 2010.
- [7] Daza A., De La Torre P., Gestión de calidad aplicada a Inteligencia de Negocios. Departamento de Ingeniería de Sistemas y Computación, Universidad Católica del Norte, 2011.