



FACULTAD DE CIENCIAS EMPRESARIALES

CARRERA PROFESIONAL DE INGENIERÍA DE SISTEMAS
EMPRESARIALES

**DASHBOARD PARA LA MEDICIÓN DE LA CONFIABILIDAD DE
LOS DATOS DEL DATAMART DE UNA ENTIDAD FINANCIERA
APLICANDO PRINCIPIOS DE CALIDAD**

TRABAJO DE INVESTIGACIÓN PARA OPTAR EL GRADO ACADÉMICO DE
BACHILLER EN INGENIERÍA DE SISTEMAS DE INFORMACIÓN Y GESTIÓN

PRESENTADO POR:
César Balaguer García
Sandro Arce Villanueva

DIRECCIÓN DE:
Dra. Sandra Meza Balbín

Lima, Perú
2019



ANEXO 6
RESULTADO DEL INFORME FINAL DEL TRABAJO DE INVESTIGACIÓN
PARA OPTAR EL GRADO DE BACHILLER

Fecha: __27__/_03__/_2019

Facultad	Ciencias Empresariales		
Carreras de pre grado			
Ingeniería de Sistemas Empresariales	<input checked="" type="checkbox"/>	Ingeniería de Sistemas de Información y Gestión - CPE	<input type="checkbox"/>
Ingeniería Económica y de Negocios	<input type="checkbox"/>	Administración de Empresas - CPE	<input type="checkbox"/>
Administración de Negocios Internacionales	<input type="checkbox"/>		
Administración de Empresas	<input type="checkbox"/>		
Marketing y Administración	<input type="checkbox"/>		
Código	Alumno	Correo electrónico	Teléfono
100022767	Balaguer García César Fernando	cesarbg01@gmail.com	997864830
100023944	Arce Villanueva, Sandro	sandroarce@hotmail.com	992118038

Título del trabajo de investigación:

DASHBOARD PARA LA MEDICIÓN DE LA CONFIABILIDAD DE LOS DATOS DEL DATAMART DE UNA ENTIDAD FINANCIERA APLICANDO PRINCIPIOS DE CALIDAD

El Jurado revisor concluye que el Trabajo de Investigación para optar el Grado de Bachiller es:

- Aprobado (13 – 15)
- Aprobado – Muy Buena (16 – 18)
- Aprobado – Sobresaliente (19 – 20)
- Desaprobado (0 – 12)

Calificándola con nota de: __15__ en letras (__QUINCE__)

Comentario/sugerencias/recomendaciones/observaciones:

- En este caso se plantean los objetivos de manera correcta a lo que se busca probar en el trabajo de investigación, sin embargo, podría hacer un trabajo de mayor profundidad que concluya en las medidas concretas y tecnológicas necesarias para mejorar el nivel de confiabilidad de los datos. Sin embargo, cumple con lo necesario para aprobar.

Atentamente,

Profesor del curso
Mg. Horacio Chacón Cursack

*Este documento puede ser enviado por correo electrónico

CONTENIDO

CONTENIDO.....	3
ÍNDICE DE TABLAS	5
ÍNDICE DE ILUSTRACIONES	6
RESUMEN.....	7
1. PROBLEMA DE INVESTIGACIÓN.....	9
1.1. Planteamiento del problema	9
1.2. Formulación del problema.....	10
1.2.1. Problema general	10
1.2.2. Problemas específicos.....	10
1.3. Objetivos de la investigación	10
1.3.1. Objetivo general.....	10
1.3.2. Objetivos específicos	10
1.4. Justificación	11
1.4.1. Práctica	11
1.4.2. Teórica	11
1.4.3. Metodológica	11
1.5. Delimitaciones.....	12
1.5.1. Tiempo	12
1.5.2. Espacio.....	12
1.5.3. Recursos.....	12
1.6. Hipótesis	13
1.6.1. Hipótesis general.....	13
1.6.2. Hipótesis específica	13
1.7. Variables y Unidades de Análisis.....	13
1.7.1. Variables	13
1.7.2. Unidad de análisis	14
1.8. Operacionalización de las variables	15
2. MARCO REFERENCIAL	17
2.1. Antecedentes Teóricos.....	17
2.1.1. Tesis Internacionales	17
2.1.2. Tesis Nacionales	18
2.2. Marco Teórico	19
2.3. Marco Conceptual	33

3. METODOLOGÍA	35
3.1. Tipo de estudio	35
3.2. Diseño de estudio.....	35
3.3. Población, muestra y muestreo.....	36
3.3.1. Población	36
3.3.2. Muestra	36
3.3.3. Muestreo	37
3.4. Métodos y técnicas e instrumentos de recolección de la información	37
3.4.1. Métodos	37
3.4.2. Técnicas.....	38
3.4.3. Instrumentos.....	38
3.4.4. Técnicas de análisis de datos.....	39
4. RESULTADOS	41
CONCLUSIONES.....	49
ANEXOS	51
Anexo I - Matriz de consistencia	51
Anexo II – Cronograma de actividades.....	52
Anexo III – Cuestionario de medición de la confiabilidad de los datos del Datamart	53
Anexo IV – Tabulación de las respuestas de la encuesta	55
BIBLIOGRAFÍA	55

ÍNDICE DE TABLAS

Tabla 1 - Matriz de Operacionalización de las variables	16
Tabla 2 - Niveles de aceptabilidad del dato según sus umbrales.....	33
Tabla 3 - Matriz de consistencia.....	51

ÍNDICE DE ILUSTRACIONES

Ilustración 1 - Patrón de problemas surgidos según el valor intrínseco en los datos.....	23
Ilustración 2 - Dificultad en la utilización de los datos.....	25
Ilustración 3 - Relación entre los datos, su calidad y la sociedad	27
Ilustración 4 - Semaforización de los umbrales de medición de la calidad	33
Ilustración 5 - Resultado de encuesta - pregunta 1	41
Ilustración 6 - Resultado de encuesta - pregunta 2	42
Ilustración 7 - Resultado de encuesta - pregunta 3	42
Ilustración 8 - Resultado de encuesta - pregunta 4	43
Ilustración 9 - Resultado de encuesta - pregunta 5	43
Ilustración 10 - Resultado de encuesta - pregunta 6	44
Ilustración 11 - Resultado de encuesta - pregunta 7	45
Ilustración 12 - Resultado de encuesta - pregunta 8	46
Ilustración 13 - Resultado de encuesta - pregunta 9	46
Ilustración 14 - Resultado de encuesta - pregunta 10	47
Ilustración 15 - Resultado de encuesta - pregunta 11	48

RESUMEN

La aplicación de los conceptos y los instrumentos de evaluación de la calidad de los datos que nos permite entender las causas básicas de los problemas y nos permite buscar soluciones para poder disponer de datos de calidad.

La aplicación de herramientas de análisis de calidad de los datos debe ser una actividad regular y sistemática de los programas actualmente usados en el banco, y debe de insertarse como un componente integral de la supervisión y la evaluación.

Los esfuerzos del banco para garantizar la calidad del dato deben enfocarse en todos los niveles de manejo de los datos, con un especial énfasis en el nivel local, pues es el punto principal de entrada de datos.

En el presente proyecto de tesis proponemos el desarrollo de un modelo para la calidad de los datos que se recibe del datamart, y que permite al banco la evaluación de la información que guarda en su Base de Datos, así como de sus protocolos relacionados.

Este modelo nos permitirá la realización de investigaciones de calidad sobre la base de los parámetros especificados en la norma ISO/IEC 25000, enfocándose especialmente en la ISO/IEC 2501, 2502, 2503 y 2504. Asimismo, planificamos su aplicación práctica en los datos relacionados al Bussiness Inteligent o Inteligencia de Negocios de una entidad bancaria.

INTRODUCCIÓN

Hoy en día, hablar de información es parte del vocabulario cotidiano y más aún cuando se trata de una organización, muchos términos se han puesto de moda, sin embargo su aplicación se ha vuelto un problema complicado de hacer, muchos conceptos y terminologías que entender, y al final todo vuelve a lo mismo, teorías, estudios, conceptos y confiar en que la información que recibimos es válida, confiamos en lo que nos dicen, lo que vemos en internet o en alguna red social o porque a alguien le fue útil cierta información.

En las organizaciones de hoy, la información es esencial para la toma de decisiones,

1. PROBLEMA DE INVESTIGACIÓN

1.1. Planteamiento del problema

Para poder entender de manera concisa es necesario conocer algunos conceptos que se aplican en el presente estudio.

- Business Inteligent: Una idea universal y resumida que se tiene de este concepto es la habilidad que se tiene para transformar los datos en información y la información en conocimientos para en base a estos poder tomar decisiones.
- Datamart: es una base de datos especializada, focalizada en un área de negocio específica con el fin de que estos puedan ser agrupados, depurados y/o transformados para su exploración para convertir los datos en información y conocimientos que permita realizar una mejor toma de decisiones.

El uso de Datamart como una herramienta segura y confiable para la toma de decisiones se ha convertido en una necesidad en las entidades que buscan la excelencia en sus servicios para con el cliente, tanto internos como externos, las entidades financieras no son ajenas a este fin, por tal motivo la entidad financiera en estudio, desde hace unos años, ha implementado esta base de datos, principalmente para las áreas de finanzas y riesgos que les permita poder explorar los datos que puedan ser útiles para la toma de decisiones estratégicas financieras, sin embargo, uno de los puntos álgidos en la toma de decisiones es la confiabilidad de la información, una información confiable y oportuna se puede convertir en un factor de éxito, así como también puede ocurrir lo contrario, siendo nefasto para los propósitos organizacionales.

Las áreas involucradas en el presente estudio aplican estas este conjunto de datos sobre la cual se aplican la mayor parte de las

decisiones financieras. Durante el año 2017 y parte del 2018, se decidió incluir dentro del Datamart la información concerniente a riesgos, de tal manera que pueda complementar la información ya existente; la inclusión de esta información ha generado cierto nivel de desconfianza en la calidad de los datos con la que se cuenta para la toma de decisiones, es a partir de este punto en donde se plantean los problemas que justifican la presente tesis.

1.2. Formulación del problema

1.2.1. Problema general

¿Es la información del Datamart confiable para la toma de decisiones?

1.2.2. Problemas específicos

¿Cuál es la calidad de la información del Datamart en términos cuantitativos?

¿La información del Datamart se encuentra dentro de umbrales aceptables de calidad basados en los principios de calidad de la información?

1.3. Objetivos de la investigación

1.3.1. Objetivo general

Establecer la confiabilidad de los datos del Datamart aplicando indicadores de calidad

1.3.2. Objetivos específicos

Determinar la calidad de los datos del Datamart estableciendo un índice de aceptación deseado de la información.

Generar una herramienta que permita establecer la aceptabilidad de los datos del Datamart mediante umbrales específicos y dinámicos y que sea sostenible en el tiempo

1.4. Justificación

1.4.1. Práctica

La presente investigación se desarrolla por la necesidad que presenta la entidad financiera en estudio, de medir la calidad y confiabilidad los datos del Datamart para la toma de decisiones, pudiendo a su vez generar líneas de trabajo en las diferentes áreas involucradas con la información, que permitan dar mayor precisión y consistencia a la información que es transferida a esta.

Esta necesidad tiene su fuente en la sede central Holding España, la cual mediante una regulación financiera, establece como norma y procedimiento dentro del proceso de toma de decisiones, la certificación de la calidad de la información que se utilizará con estos fines.

1.4.2. Teórica

La presente investigación se ha desarrollado con el fin de definir un esquema que permita trazar lineamientos basados en la calidad de los datos sustentada en las *Normas Internacionales ISO/IEC 25000 en calidad de productos de software* y de la información propiamente dicha, de tal manera que, su aplicación de paso a una serie de iniciativas que permitan introducir los conceptos de calidad de datos en todos los procesos generadores de información, para así poder garantizar la confiabilidad de los datos sobre los cuales se basan las decisiones de la entidad financiera en estudio.

1.4.3. Metodológica

La presente investigación nace por la necesidad de la entidad financiera de definir un marco metodológico sobre los datos en

donde se pueda medir la calidad de la información, aplicando métodos y estándares científicos respaldados por normas y regulaciones internacionales y que de esta manera puedan certificar la validez de los datos utilizados para la toma de decisiones.

1.5. Delimitaciones

1.5.1. Tiempo

La información existente en el Datamart comprende una historia de 3 años por lo que la evaluación y medición de la calidad de la información comprenderá únicamente a los datos comprendidos en ese lapso de tiempo.

Adicionalmente, la investigación realizada se encuentra delimitada en el tiempo otorgado por la entidad financiera para la obtención de los resultados, la cual se encuentra delimitada en un lapso de 6,000 horas las cuales iniciaron en noviembre del 2017. (El cronograma se muestra en el Anexo II – Cronograma de actividades)

1.5.2. Espacio

Las áreas impactadas en la presente investigación se encuentran en la sede central de la entidad financiera, lugar donde se desarrollará íntegramente el proyecto, los accesos a estas oficinas es libre y no presenta limitaciones, exceptuando las que corresponden al tiempo establecido para el desarrollo de este.

1.5.3. Recursos

El desarrollo de la presente investigación ha sido priorizado por la dirección Holding (sede central España) de la entidad financiera, para lo cual se ha asignado un equipo Scrum que se encuentra compuesto por jefaturas de las áreas usuarias

involucradas y su asignación se establece a tiempo completo, así como por recursos de las áreas de Engineering y System; los accesos a la red, información, recursos humanos e información tanto que alimenta el Datamart productivo como la información que esta genera ha sido proporcionada de manera ilimitada, durante el tiempo de duración de la investigación.

1.6. Hipótesis

1.6.1. Hipótesis general

Los datos del Datamart son confiables para la toma de decisiones en relación a las normas y estándares universales de calidad de datos

1.6.2. Hipótesis específica

Los datos del Datamart son confiables y presenta índices de usabilidad deseados para la toma de decisiones.

Mediante un esquema de medición de la calidad de datos del Datamart se puede establecer la confiabilidad de la información y su sostenibilidad en el tiempo.

1.7. Variables y Unidades de Análisis

1.7.1. Variables

Las variables del presente trabajo se establecieron tomando en consideración lo siguiente: en primer lugar, el contexto en el cual nace el presente proyecto, el aseguramiento de la calidad de datos como fin y objetivo fundamental, no solo en lo que el contexto pueda justificar, sino que también como una filosofía dentro de la organización, por ello, la **validación de datos**, se vuelve la razón de ser de nuestro proyecto, en segundo lugar debido a que este concepto es y será parte de

lo que la entidad financiera desea garantizar a sus clientes, que las decisiones que se toman para brindar una mejor atención a quienes son su razón, son realizadas con datos e información que tiene un valor y son confiables.

La segunda variable del presente trabajo es una consecuencia de la primera, y corresponde a la medición propiamente dicha de la calidad de los datos mediante la **aplicación de indicadores de calidad**, que puedan asegurarnos la confiabilidad de la información, para lo cual se utilizarán metodologías estadísticas que puedan darnos umbrales aceptables sobre los cuales puedan medirse de manera cualitativa la información del Datamart.

1.7.2. Unidad de análisis

El presente estudio nace como una medida regulatoria por Holding de la entidad financiera para todas sus filiales a nivel mundial, basados en el aseguramiento de la calidad de la información del Datamart, en el caso de Perú, este se basa fundamentalmente en los datos almacenados concernientes a las áreas de Finanzas y Riesgos.

El Datamart de Finanzas y Riesgos es alimentado por todos los sistemas y aplicaciones del BBVA Banco Continental y recibe información de diversas fuentes, la cual está clasificada en información de clientes, contratos, financiera y de riesgos, esta información es utilizada en todas las instancias gerenciales y de MIS para la toma de decisiones en dos de los tres niveles: táctico y estratégico, es por ello la importancia del aseguramiento de la calidad de los datos.

1.8. Operacionalización de las variables

Tipo	Variable	Definición conceptual	Dimensiones	Indicadores
Independiente	Validación de datos	“Los datos son la base principal en las actividades de toma de decisiones operacionales, tácticas y estratégicas” (Payam Hassany Shariat Panahy; Fatimah Sidi; Lilly Suriani; Marzanah A. Jabar 2013).	- Intrínsecos - Accesibilidad - Contextual	- Exactitud - Credibilidad - Reputación - Accesibilidad - Relevancia - Puntualidad - Integridad
		"Validar datos es el proceso de confirmar que los valores que se especifican en los objetos de datos son compatibles con las restricciones dentro de un esquema del conjunto de datos, al igual que las reglas establecidas para su aplicación" (Microsoft 2016)	- Representacional	- Concisa - Consistente
		Aplicación de	"El marco de evaluación de la calidad	- Completitud

<p>indicadores de calidad de datos</p>	<p>de los datos le muestra cómo medir y controlar la calidad de los datos, asegurando la calidad en el tiempo. Comenzará con conceptos generales de medición y se abrirá paso a través de un marco detallado de más de tres docenas de tipos de medición relacionados con cinco dimensiones objetivas de calidad: perfección, oportunidad, consistencia, validez e integridad" (Laura Sebastian-Coleman's 2013)</p>	<p>- Integridad</p> <hr/> <p>- Precisión</p> <hr/> <p>- Validez</p> <hr/> <p>- Consistencia</p>	<p>encontrarse informado</p> <hr/> <p>- Coherencia funcional</p> <p>- Réplica de cálculos</p> <p>- Reglas de negocio</p> <hr/> <p>- Anclaje con fuente externa</p> <hr/> <p>- Naturaleza del saldo</p> <p>- Rango de valores</p> <p>- Valores en catálogo</p> <p>- Formato</p> <hr/> <p>- Coherencia funcional</p> <p>- Anclaje entre repositorios</p> <p>- Conciliación contable</p> <p>- Integridad referencial</p> <p>- Tendencias</p>
--	---	---	---

Tabla 1 - Matriz de Operacionalización de las variables

2. MARCO REFERENCIAL

2.1. Antecedentes Teóricos

En el presente punto se van a considerar algunos antecedentes teóricos provenientes de tesis, tanto internacionales como nacionales, escritas previamente que presente variables que hayan sido planteadas bajo el esquema de calidad de la información o la importancia de estas, se presentarán así

2.1.1. Tesis Internacionales

Moraga (2013), En su tesis doctoral titulada: *Evaluación de la Calidad de Datos en Portales Web*, de la Universidad de Castilla de la Mancha , se plantea como objetivo principal la definición de un modelo de calidad para portales Web de tal manera que pueda ser utilizado por los desarrolladores, para ello utiliza una metodología probabilística, basada en encuestas de satisfacción. La tesis se enmarcó dentro del proyecto *ORIGIN*. Como conclusión de la presente tesis, Moraga establece dentro del análisis de consecución de objetivos la alineación de un modelo de calidad de DQ con el estándar ISO/IEC 25012 agrupadas en función de puntos de vista: Inherente y Dependiente del sistema y categorías: Intrínseca, Contextual, Representacional y Operacional. Presenta como aportaciones principales un estudio donde se han obtenido de usuarios las preferencias de los usuarios y han sido clasificados en función a diferentes aspectos demográficos, definiendo perfiles de estos usuarios, dependiendo también de las características del DQ en los estándares ISO/IEC 25012.

Ramírez Coronel (2014) en su tesis para la obtención del grado de magister: *Calidad de Datos en Contextos Linked*

Data, de la Universidad Politécnica de Madrid, plantea como objetivo general: proponer un conjunto de indicadores de calidad de datos RDF en contextos de Linked Data, para ello se aplicó un modelo metodológico que va desde la selección del tema propuesto hasta las conclusiones de la investigación, definiendo como sus indicadores de medición de la calidad de datos la accesibilidad, interpretabilidad, tiempo de respuesta, apertura, veracidad, consistencia, estructuralidad, relevancia, validez, reputación, actualización y licencia, esto basado en el análisis de diversas propuestas estudiadas donde se incluyen las normas ISO/IEC 25012, así como las propuestas de Furber y Hepp quienes proponen un enfoque para evaluar la calidad de los conjuntos de datos mediante consultas SPARQL. Ramírez Coronel concluye su tesis presentando tres aportaciones que comprenden: el estudio y análisis de literatura relacionada con la calidad en Linked Data en formato RDF, enfocado en la calidad de datos semánticos; propuestas de un conjunto de indicadores que permitan determinar la calidad de los datos y la aplicación a un caso de estudio en base a los indicadores de calidad propuestos.

Oviedo Blanco (2016) en su tesis Doctoral: *Modelo de madurez para portales de datos abiertos e incorporación a la norma técnica nacional de Costa Rica*, de la Universidad de Alicante, presenta en esta un modelo de madurez para portales de datos abiertos los cuales presentan una serie de requisitos, como niveles del modelo, criterios del modelo; para aquellos portales que desean un nivel de madurez mayor al básico.

2.1.2. Tesis Nacionales

San Miguel (2016), Universidad Católica del Perú, Tesis titulada: Desarrollo de un modelo de calidad de datos aplicando una solución de inteligencia de Negocios en una institución educativa.

Refiere en la aplicación de una solución de inteligencia de negocios que permite el procesamiento de datos usados en las operaciones rutinarias, tanto académicas como administrativas, para obtener distintos reportes e indicadores que pueden ser analizados para tomar decisiones que impliquen cambios importantes en la organización.

Lima (2018) Universidad Mayor de San Marcos, Tesis titulada: La calidad de los datos como un valor de activo diferencial. Propone un modelo de calidad de datos que permita evaluar y mejorar la calidad de los datos en aquellas organizaciones en las que se utilizan diferentes fuentes de base de datos con la misma información, pero con contenido diferente, lo cual no les permite tomar decisiones en forma correcta y oportuna. Específicamente se busca desarrollar un enfoque sistemático, disciplinado y estructurado que se apoye en el modelo propuesto, y que provea las pautas y guías para que la aplicación del modelo de calidad de datos pueda ser repetible y generalizable sobre los datos de cualquier organización.

2.2. Marco Teórico

Validación de datos:

La validación de los datos, cualquiera que fuese el ámbito sobre el que se requiere es un proceso fundamental cuando se trabaja con información para la toma de decisiones, de esta dependerá si se toman buenas o malas decisiones y sobre estas el futuro de una persona, sea natural o jurídica. Sin embargo, a pesar de lo delicado

que puede resultar este tema, son pocos los que se preocupan por tener una validación previa a la generación de la información.

La profesora Yadira Corral, en su tesis: Validez y confiabilidad de los instrumentos de validación para la recolección de datos nos menciona lo siguiente con respecto al tema: "*La problemática al momento de la recolección de datos en la realización de trabajos de investigación se centra en la construcción de los instrumentos a emplear con esta finalidad, de manera que permitan recabar información válida y confiable.*", aquí podemos ver un ejemplo claro que la validez de los datos no es un tema informático, sino que es algo inherente y horizontal a todos los aspectos de la vida cotidiana y laboral; en este mismo ensayo se cita a Ebel (1977, citado por Fuentes, 1989) "establece que validez ... designa la coherencia con que un conjunto de puntajes de una prueba miden aquello que deben medir", solo con datos válidos podemos generar u obtener información que resulte, valga la redundancia, válida para la toma de decisiones.

La RAE (Diccionario de la Real Academia Española, edición 2017), define válido como un adjetivo "*Firme, subsistente y que vale o debe valer legalmente.*" y son precisamente atributos que se deben tomar en cuenta cuando se desea obtener un dato para la generación de información.

El portal de la Educación Dominicana "Educando" nos muestra un artículo sobre el cómo garantizar la fiabilidad de la información obtenida en la web, donde resalta la facilidad para encontrar información y la importancia de clasificarla dado a que esta puede contener errores o no ser suficientemente válida y toma en cuenta tres factores para la validación de la información, puntualizando en aquella obtenida en la web: el autor, la exactitud y verificación de los detalles de la información y por último la vigencia de esta.

Estos tres factores que menciona el portal Web son muy importantes porque no solo nos habla de tener información válida sino que a su vez nos da pautas para poder clasificar esta información.

Garantizar la fiabilidad de la información, tal como nos menciona este portal, es en cierta forma la punta del iceberg en todo aquello que se refiere a los datos, la información y la toma de decisiones.

La revista india de Ciencia y Tecnología nos habla sobre la generación de un marco para la construcción de dimensiones de calidad de datos donde recalca a los datos como la base principal en la actividad de toma de decisiones en todos los niveles: operacionales, tácticas y estratégicas; mientras que Microsoft en la actualización de su página web del año 2016 nos define la validación de los datos como "... el proceso de confirmar que los valores que se especifican en los objetos de datos son compatibles con las restricciones dentro de un esquema del conjunto de datos, al igual que las reglas establecidas para su aplicación".

Para Diane M. Strong, Yang W. Lee y Richard Y. Wang, identifican tres roles para la fabricación de datos: los productores de datos, los custodios de los datos y los consumidores de los datos.

- Los productores de datos o Data Producers, los define como personas, grupos o cualquier otra fuente generadora de datos.
- Los custodios de los datos o Data Custodians los define como las personas que administran recursos informáticos para almacenar y procesar datos.
- Los Consumidores de datos o Data Consumers, definidos como personas o grupos de personas que usan los datos.

Este dato, proporcionado por los autores en su artículo Data Quality In Context nos menciona estos roles, que aunque parezca no tener relevancia en el tema de la validez de los datos, nos invita en cierta

forma en pensar en todo el proceso general asociados a la información, desde la obtención hasta el consumo de los datos.

En el artículo, los autores también se mencionan las dimensiones y/o patrones que debe tener todo proyecto de datos con relación a su calidad, así se mencionan los siguientes:

- Intrínseco

Los autores relacionan una dimensión intrínseca de los datos como "*Los desajustes entre las fuentes de los mismos datos son una causa común de preocupaciones intrínsecas de DQ. Inicialmente, los consumidores de datos no conocen la fuente la cual los problemas de calidad debe ser atribuido*" (Diane M. Strong; Yang W. Lee & Richard W. Wang, 1997)

En la siguiente imagen, los autores tratan de graficar el patrón de los problemas surgidos por el poco valor intrínseco en la calidad de datos.

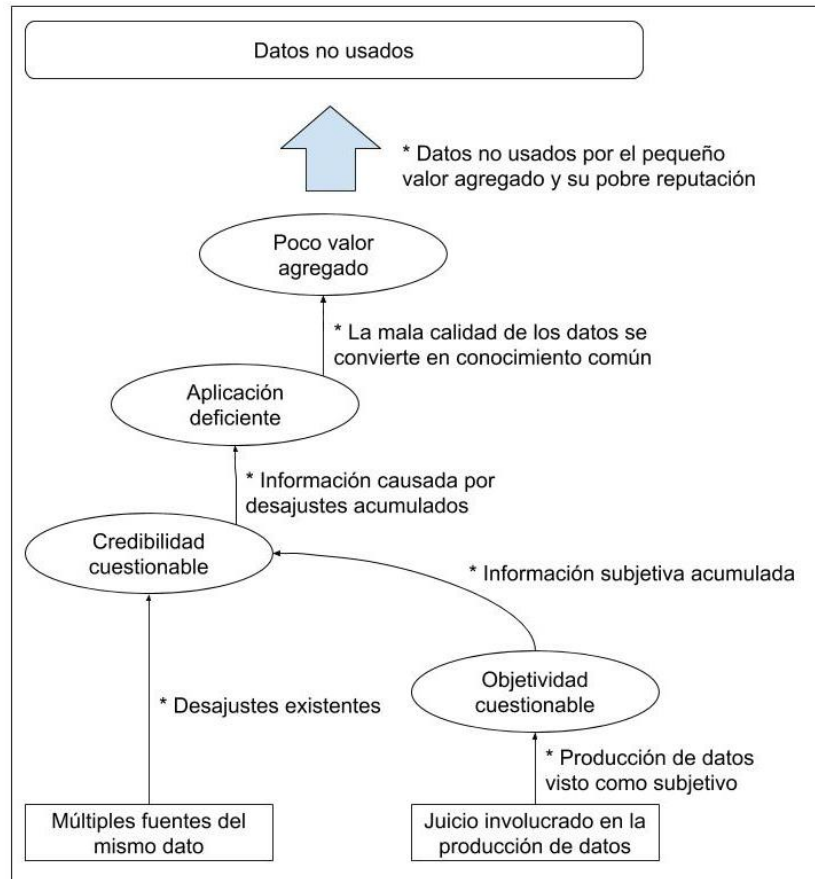


Ilustración 1 - Patrón de problemas surgidos según el valor intrínseco en los datos

Fuente: Communications of the ACM, May 1997/Vol. 40, N° 5

- **Accesibilidad**

La revista virtual byte (2018) menciona que "*Las bases de datos surgen por la necesidad de almacenar la información de una manera estructurada, para asegurar que su accesibilidad y disponibilidad sea inmediata para facilitar la tarea al usuario*", En este artículo se puede entender que la accesibilidad no solo corresponde a un tema técnico de acceso a la información, sino que relaciona la accesibilidad con la disponibilidad del dato, así se entiende que, un dato por más válido que sea, si no está disponible, este no es útil para los fines a los que fue obtenido.

Diane M. Strong; Yang W. Lee & Richard W. Wang en su artículo *Data Quality in Context* de la revista especializada *Communications of the ACM*, publicado en mayo de 1997, refieren a la accesibilidad de datos como: "*Los problemas de accesibilidad DQ se caracterizan por preocupaciones subyacentes sobre accesibilidad técnica, problemas de representación de datos interpretados por consumidores de datos como problemas de accesibilidad y problemas de volumen de datos*", este concepto presenta mucha relación con lo representado en la dimensión intrínseca sobre la interpretación de los datos, sin embargo va un poco más allá, mencionando la representación misma del dato para su correcto entendimiento.

- Contextual

La contextualidad del dato se refleja propiamente en la relevancia de los datos, la puntualidad y la integridad de estos, para ello, Diane M. Strong; Yang W. Lee & Richard W. Wang en su artículo *Data Quality in Context* de la revista especializada *Communications of the ACM*, publicado en mayo de 1997, mencionan los problemas de contextualización de los datos con tres causas subyacentes que conllevan a las principales quejas de los consumidores de datos:

- Los datos disponibles no respaldan sus tareas.
- Datos faltantes o incompletos, lo cual se refiere también a datos mal definidos o mal medidos
- Datos que no fueron agregados de manera apropiada

Para ello, representan gráficamente el problema basado en el ámbito contextual de la información, como se muestra a continuación:

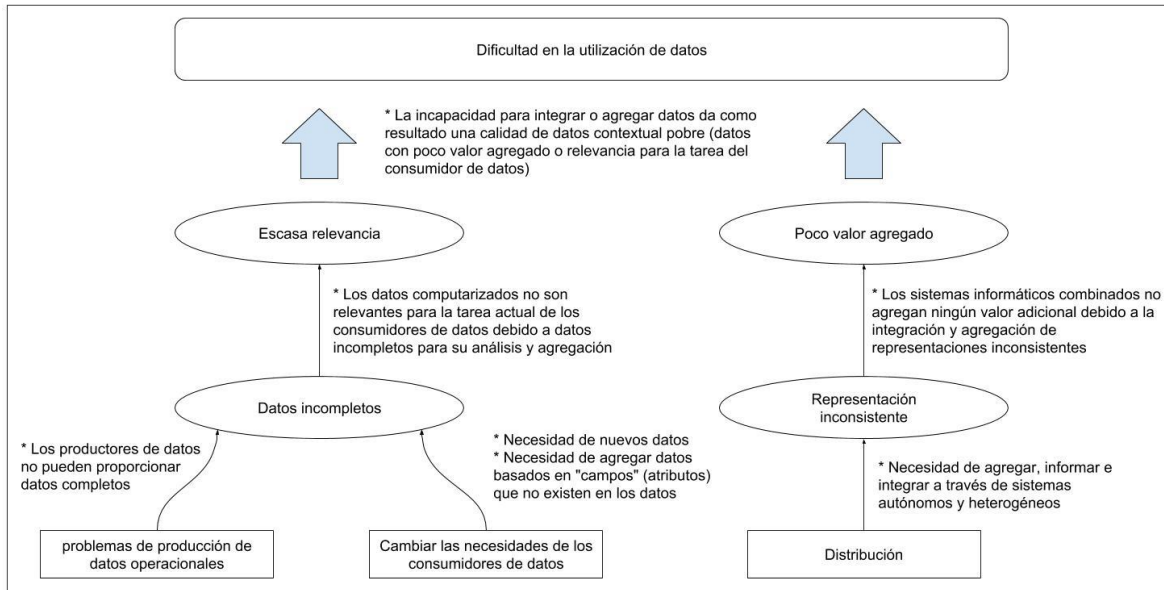


Ilustración 2 - Dificultad en la utilización de los datos

Fuente: Communications of the ACM, May 1997/Vol. 40, N° 5

- Representacional

La dimensión representacional hace referencia principalmente a la consistencia de los datos, sin embargo se encuentra inmersa en las otras dimensiones descritas por los autores, así tenemos que: "Esta investigación descubrió que las dimensiones de DQ representacionales son causas subyacentes de los patrones de problemas de DQ de accesibilidad" (Diane M. Strong; Yang W. Lee & Richard W. Wang 1997), según las definiciones previas, podemos definir a la representacionalidad del dato como un tema conciso y consistente.

La Real Academia Española (RAE, edición 2017) define "conciso" como la brevedad o la economía en el modo de

expresar un concepto con exactitud; en nuestro caso estaría representado con la exactitud del dato, es decir, que este debe representar la realidad tal y como es y esto está relacionado principalmente con la relevancia del dato.

En la misma edición, la RAE define la consistencia como duración, estabilidad, solidez, y una segunda definición como coherencia entre las partes; este concepto también relaciona al dato con su relevancia no solo en representar su valor como tal sino que este debe ser duradero y sólido en el tiempo; así podemos concluir que un dato debe contener un valor sólido y sin ambigüedades.

Aplicación de indicadores de calidad de datos

La aplicación de indicadores de calidad es en sí una forma de medir la validez, exactitud, precisión, coherencia de los valores que este dato pueda tener, esto implica presentar una forma cualitativa de medir y garantizar que el dato es confiable y que puede ser usado para la toma de decisiones.

En este punto, se explicará la forma como se plantea aplicar estos indicadores cualitativos basados en los atributos del dato, tomando en consideración metodologías y teorías ya existentes y que iremos explicando a continuación.

Graeme Shanks y Brian Corbit del departamento de sistemas de información de la universidad de Melbourne, Australia definen dentro de su investigación: "*Entendiendo la calidad de datos: aspectos sociales y culturales*" citan: "Un marco para comprender la calidad de los datos, se basa en una posición filosófica (Burrell y Morgan 1977) y definiciones claras del dato, información y significado.

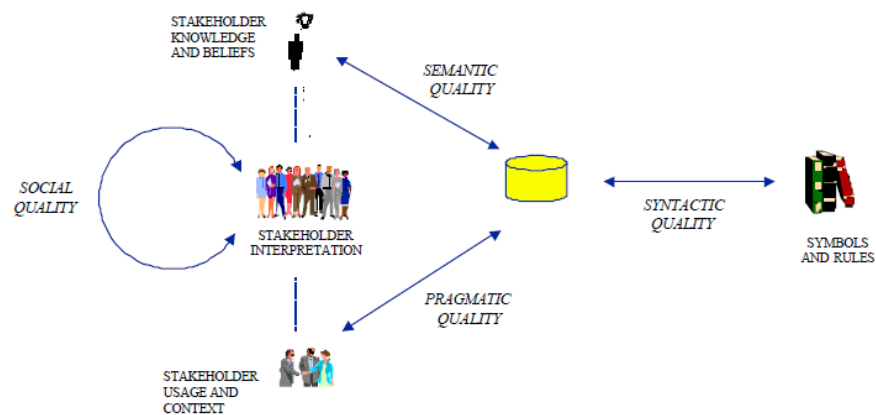


Ilustración 3 - Relación entre los datos, su calidad y la sociedad

Fuente: Entendiendo la calidad de datos: aspectos sociales y culturales;
Departamento de Sistemas de Información de la Universidad de Melbourne

En el gráfico podemos distinguir la relación y la implicancia existente de los datos, su calidad y la sociedad, entre aquellos que la crean, los que la interpretan y los que la usan, todo en un ámbito semántico y pragmático, aplicando símbolos y reglas.

Es por ello la importancia de la aplicación de indicadores de calidad basados en sus atributos que al final determinarán las dimensiones sobre las cuales serán medidos de manera cualitativa.

"El marco de evaluación de la calidad de los datos le muestra cómo medir y controlar la calidad de los datos, asegurando la calidad en el tiempo. Comenzará con conceptos generales de medición y se abrirá paso a través de un marco detallado de más de tres docenas de tipos de medición relacionados con cinco dimensiones objetivas de calidad: perfección, oportunidad, consistencia, validez e integridad" (Laura Sebastian-Coleman's 2013)

La autora nos define 5 dimensiones de la calidad que presentan más de tres docenas de tipos de medición, estas dimensiones de la calidad las enmarca de la siguiente manera:

- Perfección

- Oportunidad
- Consistencia
- Validez
- integridad

La empresa Power Data, especializada en calidad de la información, entre otras, publica: *"Es muy posible que tus datos nunca vayan a ser perfectos. Pero si es posible acercarse a la perfección haciendo que la integridad de datos sea una prioridad."* (Power Data, 4 mejores prácticas para mantener la integridad de los datos, Abril 2017).

Hablar de perfección de datos es quizá algo utópico, sin embargo, si se puede garantizar que el dato es útil, preciso e íntegro, garantizando que su valor refleja la realidad exacta al momento de su obtención.

La enciclopedia electrónica ACE en su versión 2.0 cita: *"Tener disponible la información para los usuarios es uno de los propósitos principales de un sistema de cómputo. Empero, al mismo tiempo es importante para asegurar su integridad."*

La oportunidad del dato se refiere principalmente a la disponibilidad de este, para poder aprovechar el valor agregado que pueda tener un dato cuya exactitud e integridad ha sido validado, este debe poder tenerse en el momento en que se requiere, de esta manera podrá aprovecharse el valor agregado que el dato pueda presentar para una toma de decisiones

Las norma ISO 25000 conocida también como SQuaRE por sus siglas en inglés (System and Software Requirements and Evaluation) tiene por objetivo la creación de un marco metodológico de trabajo común para la evaluación de la calidad del producto de software; esta norma nace como una evolución de normas previas como la ISO/IEC 9126 que describe un modelo de calidad del producto software y la ISO/IEC 14598 que trata sobre los procesos

de evaluación de los productos de software. Esta familia ISO 25000 está compuesta de la siguiente manera:

- ISO/IEC 2501n para el modelo de calidad
- ISO/IEC 2502n para la medición de calidad
- ISO/IEC 2503n para los requisitos de calidad
- ISO/IEC 2504n para la evaluación de la calidad

Para la aplicación de los indicadores de calidad se han definido 5 principios o dimensiones, estas se han basado principalmente en las recomendaciones de los estándares ISO 25000 y de las teorías y conceptos enmarcados en la calidad de datos que se han mostrado previamente, tomando estas adicionalmente como un marco metodológico aplicable al BBVA en su afán de mejoramiento continuo y búsqueda de la excelencia tomando como punto de partida los datos en los que se basan todas sus decisiones.

Así, se definen los siguientes principios o dimensiones para la aplicación de los indicadores de calidad de datos en el Datamart del BBVA.

- **Compleitud**

"Compleitud implica que los datos necesarios se encuentren presentes e informados" (Directiva acerca de la calidad de los datos, BBVA 2017)

Este principio será medido cualitativamente tomando 55 datos considerados de implicancia alta y su valor de medición estará basado en que este se encuentre debidamente informado en el repositorio principal donde debe estar almacenado.

- **Integridad**

"Los datos relacionados son coherentes entre sí y cumplen las reglas de negocio relativas a los mismos." (Directiva acerca de la calidad de los datos, BBVA 2017)

El principio de integridad constará de 10 indicadores que serán medidos cualitativamente mediante tres sub principios:

- Coherencia funcional

Hace referencia al valor del dato con respecto a su naturaleza y a la información que representa, el dato debe tener un valor que sea coherente en relación a lo que este representa.

- Réplica de cálculos

Hace referencia a los cálculos realizados para otorgar el valor de un campo, con respecto al ámbito que representa.

- Reglas de negocio

Los valores de los estados financieros son coherentes con las calificaciones de rating y sus estados.

- **Precisión**

"Los datos deben reflejar la realidad exacta." (Directiva acerca de la calidad de los datos, BBVA 2017)

Este principio se medirá cualitativamente mediante un solo indicador conceptualizado en un sub principio que permita medir el anclaje con fuentes externas que no han sido parte del aprovisionamiento del dato, estas fuentes externas son propias de las áreas de finanzas y de riesgos y sus valores deberán ser consistentes con los que se encuentran almacenados en el Datamart.

- **Validez**

"Los datos tienen unos valores válidos para su uso: formato, longitud, valores permitidos, etc." (Directiva acerca de la calidad de los datos, BBVA 2017)

El principio de validez será medido mediante 16 indicadores distribuidos en 4 sub principios cualitativos:

- Naturaleza de los saldos
- Rango de valores
- Valores en catálogo
- Formatos

- **Consistencia**

"los datos son consistentes a lo largo del proceso en el propio repositorio y respecto al resto de sistemas involucrados." (Directiva acerca de la calidad de los datos, BBVA 2017)

El principio de consistencia será medido mediante 122 indicadores distribuidos en 5 sub principios cualitativos:

- Coherencia funcional
- Anclaje entre repositorios
- Conciliación contable
- Integridad referencial
- Tendencias

Para poder lograr la medición de la calidad de los datos, la entidad financiera proporcionó estos en un cuadro donde se mencionan los principios y sub-principios sobre los cuales deben ser medidos, estos fueron seleccionados por la misma entidad y sus umbrales de aceptación se encuentran también en el mismo documento, el concepto funcional o de significación del dato no fue proporcionado en un documento por ser de índole privado y confidencial, motivo por el cual no se adjunta ni se mencionan en la presente tesis.

Los umbrales de aceptación del dato presentan tres posibles valores:

- **Umbral máximo**

El umbral máximo corresponde al valor porcentual referencial máximo que puede tomar la medición de la calidad de un dato.

- **Umbral permitido o deseado**

El umbral permitido o deseado corresponde al valor porcentual referencial mínimo que puede tomar un dato para ser considerado como aceptable o certificado para la toma de decisiones.

- **Umbral mínimo**

El umbral mínimo corresponde al valor porcentual referencial mínimo de calidad que puede presentar un dato, cualquier dato cuya calidad se encuentre por debajo de este umbral no solo será considerado un dato inválido para cualquier toma de decisiones, sino que a su vez ocasionará la generación de líneas de trabajo que permitan revisar el origen de negocio y funcional del dato, su evaluación y mejora de procesos para su obtención.

En el siguiente gráfico se muestra cómo es que se aplicará la medición de un dato con respecto a sus umbrales, para la aceptabilidad o no del dato.

Umbral máximo	Umbral deseado	Umbral mínimo
Si el porcentaje de calidad supera al umbral máximo será	Si el porcentaje de calidad es mayor o igual al umbral	Si el porcentaje de calidad es mayor al umbral mínimo, se

considerado un dato no válido.	deseado, se considera un dato válido y/o aceptable para la toma de decisiones	considera un dato no válido para la toma de decisiones
Si el porcentaje de calidad es menor o igual al umbral máximo y mayor o igual al umbral deseado el dato será considerado como válido y/o aceptable para la toma de decisiones	Si el porcentaje de calidad es menor que el umbral deseado, se considera un dato no válido para la toma de decisiones	Si el porcentaje de calidad del dato es menor al umbral mínimo el dato no es considerado y debe reportarse para la generación de líneas de trabajo.

Tabla 2 - Niveles de aceptabilidad del dato según sus umbrales

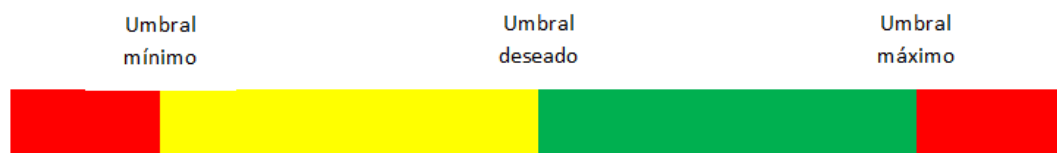


Ilustración 4 - SemafORIZACIÓN de los umbrales de medición de la calidad

2.3. Marco Conceptual

Datos: Según la RAE (Real Academia Española) en su actualización del 2017, define el dato como: Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho, el dato así puede ser definido

como la unidad mínima de la información, así, un conjunto de datos bien estructurados puede representar información útil que puede ser analizada o estudiada.

Información: Es un conjunto de datos estructurados, organizados y ordenados que sirven para construir un mensaje sobre el cual se realizará una toma de decisiones.

Base de datos: Una base de datos es un conjunto de datos estructurados que pertenecen a un mismo contexto, almacenados sistemáticamente para su uso posterior.

Validez: Según la RAE (Real Academia Española, edición actualizada 2017) define validez como la "*Cualidad de válido*"

Business Intelligence: La inteligencia de negocios desde el punto de vista de información, corresponde a un conjunto de metodologías, herramientas y técnicas que permite transformar los datos en información y la información en conocimientos para una mejor toma de decisiones.

Datawarehouse: Bill Inmon creador del término, lo define como almacén de datos caracterizado por ser integrado, temático, histórico y no volátil.

Calidad: La calidad se define como un conjunto de propiedades inherentes a una cosa, que permite caracterizarla y valorarla con respecto al resto de cosas de su propia especie.

Normas ISO: ISO son las siglas en inglés de Organismo Internacional de Normalización, las normas establecidas por ISO son

formas comunes de conseguir el establecimiento de sistemas de calidad que garanticen la satisfacción de las necesidades y expectativas de los consumidores.

3. METODOLOGÍA

3.1. Tipo de estudio

El tipo de estudio realizado es de Investigación Aplicada; José Lozada, 2014 de la revista CienciAmérica de la Universidad IndoAmérica de Guayaquil, Ecuador publica un trabajo de investigación basado en la conceptualización de lo que es una Investigación Aplicada, así en la página 22 y 23 de "*Investigación Aplicada: Definición, Propiedad Intelectual e Industria*" nos indica: "*La investigación aplicada busca la generación de conocimiento con aplicación directa a los problemas de la sociedad o el sector productivo. Esta se basa fundamentalmente en los hallazgos tecnológicos de la investigación básica, ocupándose del proceso de enlace entre la teoría y el producto.*"

Dada la naturaleza del presente trabajo de investigación, su implicancia en la entidad de estudio y su relevancia en la sociedad a través de las decisiones que se puedan tomar sobre los datos y su uso directo, es que se decidió realizar el tipo de estudio mencionado en este punto.

3.2. Diseño de estudio

"*El diseño experimental es una estructura de investigación donde al menos se manipula una variable y las unidades son asignadas aleatoriamente a los distintos niveles o categorías de la variable o variables manipuladas.*" (Roser bono Cabré, Departamento de

Metodología de las Ciencias del Comportamiento de la Universidad de Barcelona)

Debido a la naturaleza de la investigación y a la definición presentada previamente, se aplicará el Diseño Experimental, atendiendo a los objetivos y su consecución que este tipo de diseños presenta.

3.3. Población, muestra y muestreo

3.3.1. Población

El BBVA - Banco Continental es una empresa del sector financiero, cuya finalidad es la de generar valor al dinero confiado de sus clientes mediante instrumentos financieros como inversiones y otros.

La estructura organizativa del BBVA se basa en áreas propias del negocio así como áreas back office cuyos colaboradores se encargan de gestionar la información para la toma de decisiones; oficinas de atención al cliente desde donde estos pueden realizar las transacciones monetarias que requieran

3.3.2. Muestra

La muestra corresponde a las áreas de finanzas y riesgos dado a que son estas áreas los principales usuarios de la información almacenada en el Datamart.

El área de finanzas realiza toma de decisiones con respecto a los clientes y estratégicas según los saldos, la situación financiera de estos y los importes de estanco de los clientes.

El área de riesgos realiza las calificaciones y evaluación crediticia de los clientes según su información financiera, así como las estrategias a seguir para cada uno de los grupos de calificación.

3.3.3. Muestreo

El estudio se realizará mediante un muestreo no probabilístico. Se seleccionó este tipo de muestreo dado a que el proceso de selección no brinda a todos los individuos la misma probabilidad de selección

3.4. Métodos y técnicas e instrumentos de recolección de la información

3.4.1. Métodos

- Analítico sintético

Este método de análisis permite presentar diversos métodos de resolución, primero divide los elementos y luego reúne aquellos que tienen una lógica en común

Método analítico: este método está referido al análisis de los fenómenos, examina y descompone para su estudio, inicia con todo el fenómeno y luego lo revisa parte por parte para comprender su funcionamiento.

Método sintético: Este método sintetiza los fenómenos, todas las partes que fueron divididas durante el análisis son unidas de manera coherente y lógica común para su estudio, este método nos lleva de la causa a los efectos.

- Inferencial

Este método permite inferir a partir de la información proporcionada por la muestra, basada en el comportamiento de la población, este tipo de metodología presenta un riesgo de error el cual puede ser medible probabilísticamente.

La inferencia se puede dividir en dos métodos: la estimación de parámetros y el método de contraste de hipótesis, ambos se basan en los conocimientos teóricos de la distribución probabilística.

- Descriptivo

Este método permitirá describir el comportamiento de las variables para una búsqueda de respuestas, este método implica observar y catalogar la información, permite obtener datos precisos que puedan aplicarse.

Normalmente este tipo de estudios abre paso a otros métodos más complejos.

3.4.2. Técnicas

En esta parte presentamos las técnicas utilizadas que nos sirvió de ayuda para la obtención de los puntos a considerar en el modelo de calidad de los datos.

3.4.3. Instrumentos

Los instrumentos para la evaluación utilizados para el presente trabajo fueron los siguientes, en conjunto con las normas ISO/IEC propuestas:

DATAMART, que es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

En síntesis, se puede decir que los data marts son pequeños data warehouse centrados en un tema o un área de negocio específico dentro de una organiza.

Bussiness Inteligent: También conocida como inteligencia empresarial, es una idea universal y resumida que se tiene de este concepto es la habilidad que se tiene para transformar los datos en información y la información en conocimientos para en base a estos poder tomar decisiones.

Structured Query Language (SQL): Lenguaje de consulta usado mundialmente para realizar operaciones con bases de datos. Esta herramienta permitirá la elaboración de los scripts utilizados para las consultas que permitan el conteo de las métricas de calidad.

3.4.4. Técnicas de análisis de datos

Técnica de analisis de grupo:

Esta técnica de analisis grupal, es un método que permitio realizar una lluvia de ideas entre los dos integrantes del grupo y con el apoyo de informacion recabada a traves de diferentes publicaciones relacionadas al tema.

Para su elaboracion consideramos algunos pasos que consideramos importantes:

1. Primero indicamos el propósito de nuestra evaluación y las preguntas a responder y aclarando las dudas hasta que sea entendido.

2. Posteriormente cada uno de nosotros escribió las ideas que teníamos sobre la pregunta propuesta, en un lapso de tiempo determinado (10 min aprox.).
3. Después cada uno mencionamos nuestras ideas en voz alta.
4. Las ideas son discutidas entre los dos por un lapso de una hora.
5. Priorizamos las ideas mediante votación y reducción de la lista de ideas.

4. RESULTADOS

Se realizó una encuesta a personas que trabajan en diversas entidades, tanto financieras como industriales, debido a que, siendo una investigación basada en una entidad financiera, la calidad de los datos para la toma de decisiones es esencial no solo para empresas de este rubro, sino que es aplicable a cualquier empresa, el proyecto de calidad de datos que se pretende presentar a raíz de este trabajo de investigación debe aplicar a toda empresa que se preocupa en la confiabilidad de su información para toma de decisiones.

- Análisis por pregunta
 - Pregunta 1 (¿Es fácil acceder a la información sobre los diferentes datos de la base principal?)

Con los resultados obtenidos en la primera pregunta se puede observar un nivel de satisfacción alto con respecto a la accesibilidad de los datos de la base principal de la entidad.

En esta primera pregunta se ha tratado de medir la **Accesibilidad** a los datos almacenados.

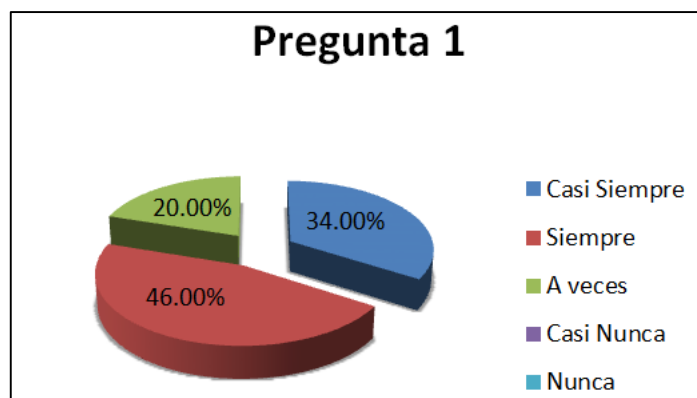


Ilustración 5 - Resultado de encuesta - pregunta 1

- Pregunta 2 (¿Ha recibido la información que necesitaba para tomar una buena decisión?)

Se puede observar al igual que en el cuadro anterior, un nivel de satisfacción considerable, las personas que toman decisiones suelen recibir toda la información necesaria para poder plantear estrategias.

En esta segunda pregunta se ha querido medir el principio de **Compleitud** basada en las normas ISO tratadas en la presente investigación.

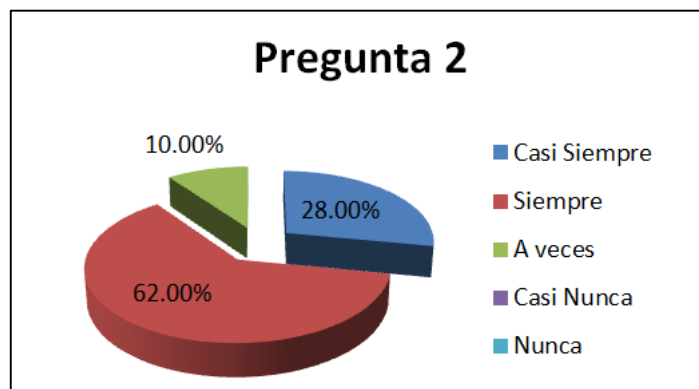


Ilustración 6 - Resultado de encuesta - pregunta 2

- Pregunta 3 (¿La información recibida ha sido clara y confiable?)

Al igual que las preguntas previas, los encuestados consideran que la información recibida ha sido clara y confiable.

En esta pregunta se han medido dos principios: **Integridad** y **Validez**

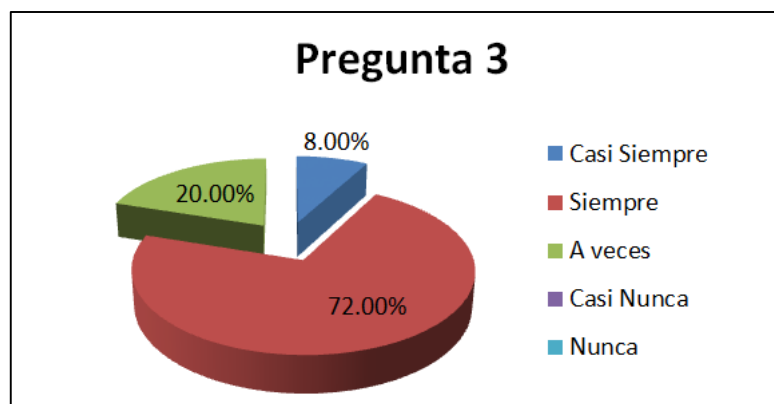


Ilustración 7 - Resultado de encuesta - pregunta 3

- Pregunta 4 (¿EL DATAMART usado actualmente funciona de forma correcta y sin errores?)

En esta pregunta se quiso medir la **Confiabilidad** ya no de los datos puntuales sino de todo el Datamart, nótese que en esta pregunta recién se utiliza este término, la intención fue poder determinar si los encuestados pueden determinar la diferencia entre un Datamart y una base propia de negocios.

El resultado fue, como en los casos anteriores, de conformidad sobre la confiabilidad y usabilidad del Datamart.

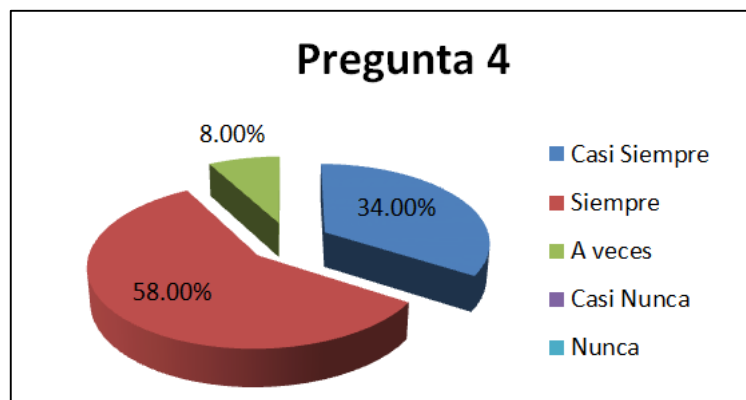


Ilustración 8 - Resultado de encuesta - pregunta 4

- Pregunta 5 (¿La información que buscas en el DATAMART la encuentras fácilmente?)

Con esta pregunta se ha medido el principio **Accesibilidad** al Datamart, a diferencia de la primera pregunta que mencionaba a la información en términos generales.

La respuesta de los encuestados fue mayormente satisfactoria

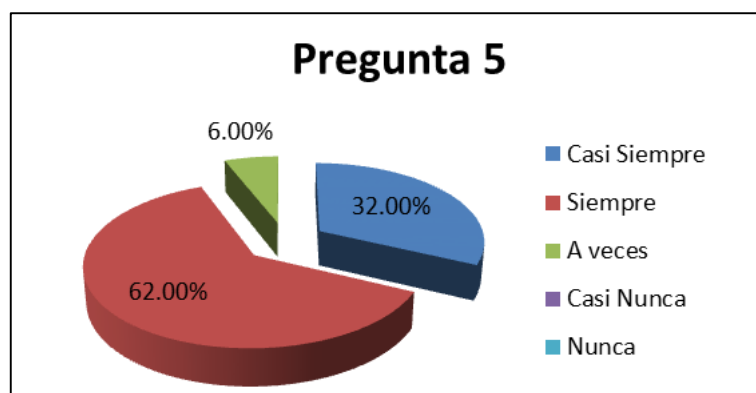


Ilustración 9 - Resultado de encuesta - pregunta 5

- Pregunta 6 (¿La información enviada por el DATAMART es fiable? ¿Es veraz, acertada y muestra la realidad?)

Esta pregunta mide tres de los seis principios de calidad: **Integridad**, **Validez** y **Consistencia**, por lo que se considera como una de las preguntas más importantes de la encuesta.

En este caso, la mayoría de los encuestados considera

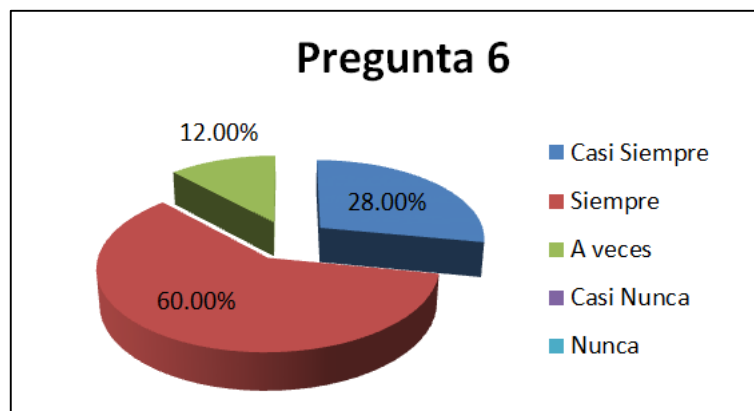


Ilustración 10 - Resultado de encuesta - pregunta 6

- Pregunta 7 (¿Usted recibe el 100% de la información que necesita?)

Esta pregunta trata de medir los principios de **Complejidad** e **Integridad** según las normas de calidad. En este caso se puede observar que la satisfacción de los encuestados se mantiene alta, sin embargo el porcentaje de insatisfacción se incrementa considerablemente, dando un nivel de duda la información recibida.

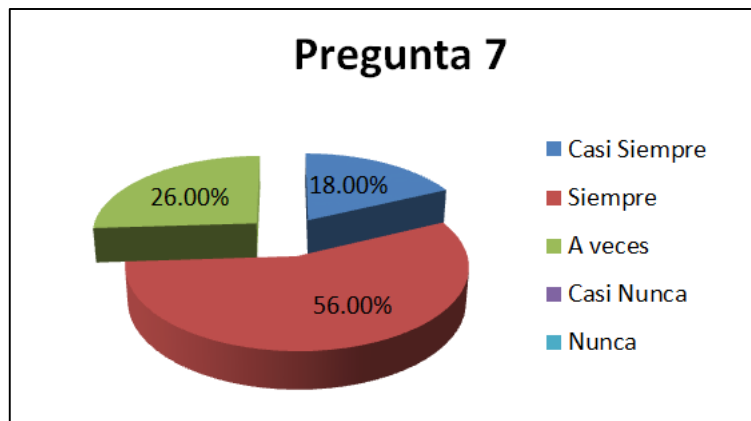


Ilustración 11 - Resultado de encuesta - pregunta 7

- Pregunta 8 (¿En general, cuál es su grado de satisfacción con DATAMART?, siendo 1 el mayor grado de satisfacción y 5 el mayor grado de insatisfacción)

A partir de esta pregunta se buscó medir la satisfacción y la calidad de los datos y del Datamart en General, en este caso la escala de satisfacción se mide desde valores Muy alto hasta Muy bajo.

Aquí se observa un nivel de satisfacción más disperso, sin embargo se observa un alto grado de satisfacción, a pesar de eso, se observa ya un nivel de insatisfacción sobre el Datamart, aunque aún bajo.

Se podría considerar una contradicción con respecto a las respuestas observadas previamente, sin embargo, el porcentaje de insatisfacción alcanzado concuerda con el porcentaje que se mostró en las preguntas previas, que marcaron la diferencia con respecto a la aceptación y/o satisfacción de la calidad de los datos recibidos.

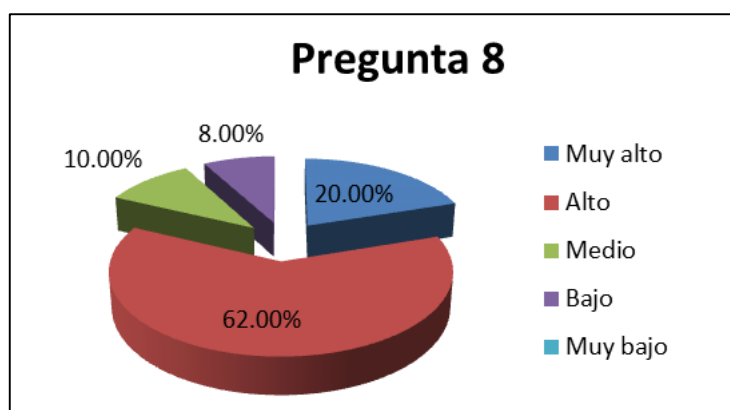


Ilustración 12 - Resultado de encuesta - pregunta 8

- Pregunta 9 (¿Cree tener la necesidad de contar con un nuevo Datamart? Considere 1 como SI y 2 como NO)

Esta pregunta buscó medir, al igual que la anterior, la satisfacción de los encuestados con respecto al Datamart, las respuestas encontradas son consistentes con el nivel de satisfacción encontrado en el transcurso de la encuesta, la mayoría de los encuestados se encuentran satisfechos con el Datamart y no considera la necesidad de cambiar o migrar el que actualmente tienen.

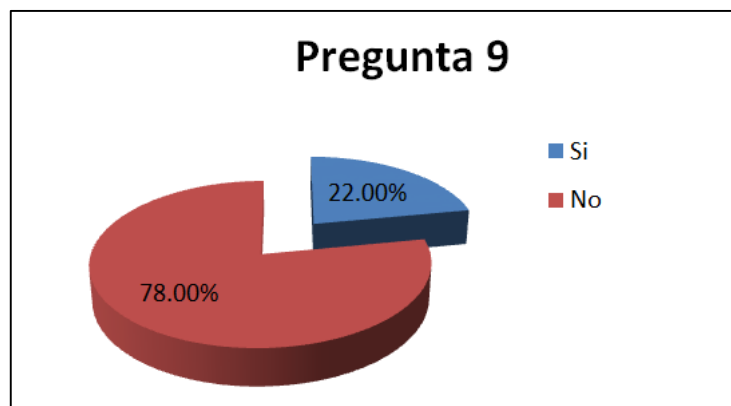


Ilustración 13 - Resultado de encuesta - pregunta 9

- Pregunta 10 (¿Conoce normas y/o estándares de medición de la calidad de la información? Considere 1 como SI y 2 como NO)

Esta pregunta, al igual que la siguiente, trata de medir sobre el conocimiento de normas y estándares que permitan medir la confiabilidad de los datos que utilizan para la toma de decisiones, esto con el fin de poder determinar cuánto “confían a ciegas” los encuestados los datos que reciben para la toma de decisiones.

El resultado en este caso fue el esperado y completamente opuesto a los resultados previos; los encuestados no tienen conocimiento, en su mayoría, de estándares de medición de la calidad de los datos y la

información, **asumiendo** que todos los datos que reciben son correctos y obedecen a la realidad, no siendo esto necesariamente cierto.

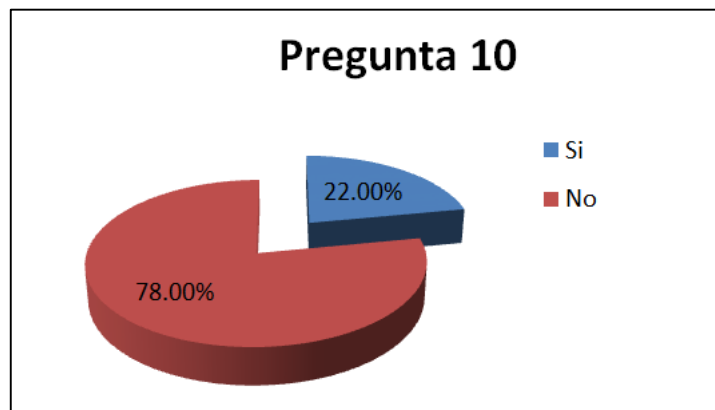


Ilustración 14 - Resultado de encuesta - pregunta 10

- Pregunta 11 (¿Sabe si los datos utilizados para la toma de decisiones han pasado por un proceso de medición de la calidad? Considere 1 como SI y 2 como NO)

Esta pregunta trata de confirmar y/o negar los resultados previos, siendo esencial para poder determinar las conclusiones generales de la investigación realizada.

La mayoría de los encuestados desconoce si los datos recibidos para la toma de decisiones han pasado por un proceso de validación y/o control de calidad aplicando normas o estándares para la determinación de su confiabilidad, esto corrobora en cierta forma que los encuestados, en su mayoría, “confían a ciegas” en la información recibida.

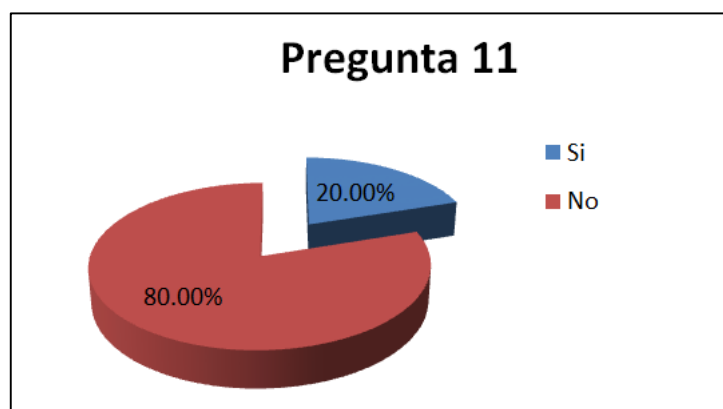


Ilustración 15 - Resultado de encuesta - pregunta 11

CONCLUSIONES

Los resultados obtenidos en la encuesta dan a conocer la necesidad que se presenta, no solo la entidad en estudio sino toda entidad financiera o no, de sentirse seguro y confiado con la información recibida para una correcta toma de decisiones, de la misma manera, se pudo apreciar que se conoce muy poco sobre estándares de calidad de datos, sistemas de medición, normas, principios de calidad de datos y de información, que puedan de alguna manera garantizar que la información utilizada para la toma de decisiones es confiable, es segura y obedece a la realidad.

Muchas entidades no manejan estándares de calidad para su información, confiando ciegamente en muchos casos de los datos que sus sistemas, manuales o computarizados, almacenan o generan, generando esto malas decisiones que llevan al no cumplimiento de metas o en todo caso al sobre costo y/o sobre esfuerzo para el cumplimiento de estas.

Mediante este estudio hemos querido demostrar que los responsables de garantizar la calidad de la información, no están cumpliendo su rol a cabalidad, al igual que aquellas personas encargadas de generar estándares de confiabilidad de la información, la presente investigación es el punto de partida para la generación de un proyecto integral que pueda medir cuan confiables son los datos utilizados para la toma de decisiones, mediante la aplicación de normas estandarizadas y reconocidas, de tal manera que puedan certificar cada uno de los principios que se exigen para poder garantizar la información, estableciendo umbrales de calidad por cada uno de los principios que requiera y/o necesite la entidad y estableciendo semaforizaciones a los datos que serán medidos.

Cabe señalar, que el presente estudio se basa únicamente en datos e información que es utilizada para la toma de decisiones, mas no para información de trabajo o información propia de los procesos de negocio; el presente estudio genera así una línea de trabajo para la aplicación de un Dashboard donde los usuarios encargados de las tomas de decisiones y generación de estrategias (nivel táctico y nivel estratégico) puedan

tener la confianza y la garantía que los datos que utilizan para generar información útil para sus organizaciones se encuentra dentro de los umbrales de aceptabilidad y, de esta manera sus objetivos, metas y estrategias son ajustadas a una realidad vigente.

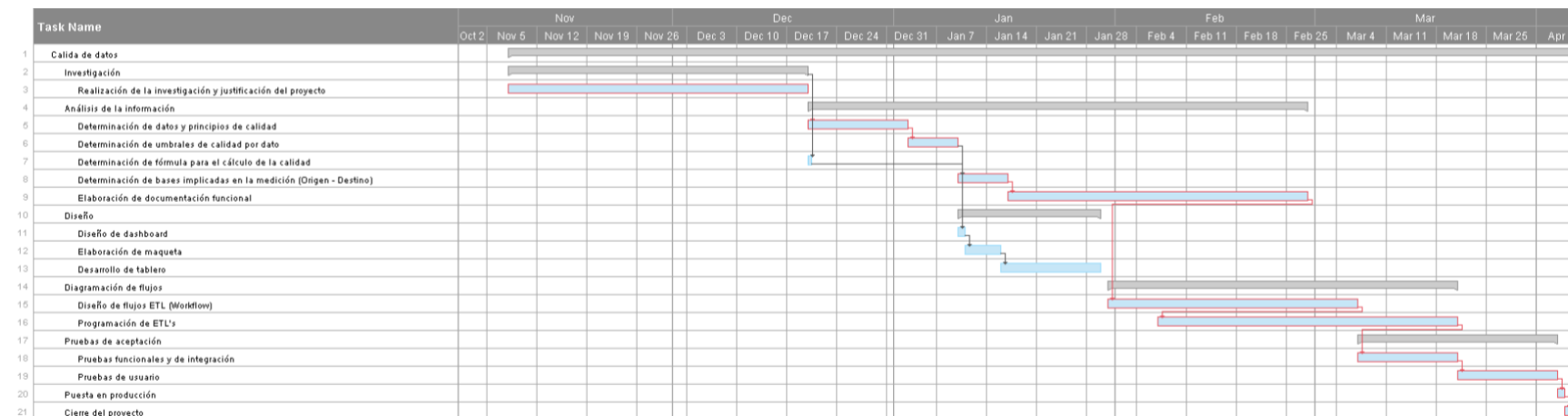
ANEXOS

Anexo I - Matriz de consistencia

Problema	Objetivos	Hipótesis	Variables
Problema general	Objetivo general	Hipótesis general	Variable independiente
¿Es la información del Datamart confiable para la toma de decisiones?	Establecer la confiabilidad de los datos del Datamart aplicando indicadores de calidad	Los datos del Datamart son confiables para la toma de decisiones en relación a las normas y estándares universales de calidad de datos	Validación de datos
Problemas específicos	Objetivos específicos	Hipótesis específicas	Variable dependiente
¿Cuál es la calidad de la información del Datamart en términos cuantitativos?	Determinar la calidad de los datos del Datamart, estableciendo un índice de aceptación deseado de la información	Los datos del Datamart son confiables y presentan índices de usabilidad deseados para la toma de decisiones	Aplicación de indicadores de calidad de datos
¿La información del Datamart se encuentra dentro de umbrales aceptables de calidad basados en los principios de calidad de la información?	Generar una herramienta que permita establecer la aceptabilidad de los datos del Datamart mediante umbrales específicos y dinámicos y que sea sostenible en el tiempo	Mediante un esquema de medición de la calidad de los datos del Datamart se puede establecer la confiabilidad de la información y su sostenibilidad en el tiempo	

Tabla 3 - Matriz de consistencia

Anexo II – Cronograma de actividades



Anexo III – Cuestionario de medición de la confiabilidad de los datos del Datamart

INTRODUCCIÓN

El presente cuestionario tiene la finalidad de conocer la confiabilidad de los datos que envía el Datamart; es anónima, sin embargo con fines estadísticos se requiere algunos datos sociodemográficos. Gracias por su valiosa colaboración.

EDAD	GÉNERO	CARGO

INDICACIONES:

Lee atentamente y luego marca con una (X) la alternativa del recuadro que sea compatible con tu respuesta. Identifica los números con los que se relacionan y colócalos según creas conveniente. El tiempo de aplicación es de 10 minutos. Gracias.

SIEMPRE	CASI SIEMPRE	A VECES	CASI NUNCA	NUNCA
1	2	3	4	5

DIMENSIÓN:		SIEMPRE	CASI SIEMPRE	A VECES	CASI NUNCA	NUNCA
1	¿Es fácil acceder a la información sobre los diferentes datos de la base principal?					
2	¿Ha recibido la información que necesitaba para tomar una buena decisión?					
3	¿La información recibida ha sido clara y confiable?					
4	¿EL DATAMART usado actualmente funciona de forma correcta y sin errores?					
5	¿La información que buscas en el DATAMART la encuentras fácilmente?					

6	¿La información enviada por el DATAMART es fiable? ¿Es veraz, acertada y muestra la realidad?					
7	¿Usted recibe el 100% de la información que necesita?					
8	¿En general, cuál es su grado de satisfacción con DATAMART?, siendo 1 el mayor grado de satisfacción y 5 el mayor grado de insatisfacción					
9	¿Cree tener la necesidad de contar con un nuevo data warehouse? Considere 1 como SI y 2 como NO					
10	¿Conoce normas y/o estándares de medición de la calidad de la información? Considere 1 como SI y 2 como NO					
11	¿Sabe si los datos utilizados para la toma de decisiones han pasado por un proceso de medición de la calidad? Considere 1 como SI y 2 como NO					
12	¿Qué adicionarías al informe de datos?					

Anexo IV – Tabulación de las respuestas de la encuesta

Encuestados	Preguntas										
	1	2	3	4	5	6	7	8	9	10	11
1	2	2	2	1	2	2	2	2	2	2	2
2	2	2	3	1	2	2	2	2	2	2	2
3	1	2	2	2	2	2	2	2	2	2	2
4	2	2	2	1	2	2	2	2	2	2	2
5	3	2	2	1	1	2	2	2	2	2	2
6	2	1	2	1	2	2	2	2	2	2	2
7	2	1	1	2	1	1	1	1	2	2	2
8	3	2	2	3	1	2	2	2	2	2	2
9	1	2	2	2	2	2	2	2	2	2	2
10	2	2	3	2	2	2	3	2	2	2	2
11	1	1	2	2	1	1	2	1	2	2	2
12	2	2	2	2	2	2	2	2	2	2	2
13	2	2	2	1	2	1	2	2	2	2	2
14	1	2	2	2	2	1	2	2	2	2	2
15	3	3	2	2	3	2	3	2	2	2	2
16	1	2	2	2	2	2	2	2	2	2	2
17	2	3	3	1	3	2	3	3	1	1	1
18	3	2	3	1	2	3	3	2	2	2	1
19	1	2	2	2	2	2	2	2	2	2	2
20	2	1	1	2	2	1	1	1	2	1	1
21	2	2	2	2	2	2	2	2	2	1	1
22	2	1	2	2	2	1	2	2	2	2	1
23	1	1	2	2	1	1	1	1	2	2	2
24	1	1	1	2	1	2	2	2	2	2	1
25	2	2	3	2	2	2	3	3	1	1	1
26	3	2	3	2	2	3	3	4	1	2	2
27	1	1	2	2	1	2	2	1	2	1	2
28	1	2	1	1	2	1	1	1	2	1	2
29	2	1	2	1	2	1	2	2	2	1	2
30	1	2	2	1	2	1	2	2	2	1	2
31	2	1	2	2	2	2	2	2	2	2	2
32	3	2	2	3	2	3	3	3	1	2	2
33	2	2	2	2	1	1	2	2	2	2	2
34	3	2	3	2	2	2	3	3	1	2	2
35	1	2	2	2	2	2	2	2	1	2	2
36	1	2	2	1	1	1	1	1	2	2	2
37	1	1	2	1	1	1	1	1	2	2	2
38	2	2	2	1	1	2	2	1	2	2	2
39	2	1	2	2	1	2	2	2	2	2	2
40	2	1	2	2	1	2	2	2	2	2	1
41	1	2	2	1	2	1	2	2	2	2	1
42	2	2	2	2	2	2	3	2	1	2	2
43	3	3	3	2	3	3	3	4	1	2	2
44	1	2	2	1	1	2	1	1	2	1	2
45	2	3	2	1	2	2	2	2	2	1	2
46	2	3	3	2	2	3	3	3	1	2	2
47	1	2	2	2	1	2	1	2	2	2	2
48	3	1	2	2	1	2	1	2	2	1	1
49	3	2	2	3	2	2	3	4	1	2	2
50	2	2	3	3	2	3	3	4	1	2	2

BIBLIOGRAFÍA

Hassany Shariat Panahy, Payam & Sidi, Fatimah & Affendey, Lilly & A. Jabar, Marzanah & Ibrahim, Hamidah & Mustapha, Aida. (2013). A Framework to Construct Data Quality Dimensions Relationships. Indian Journal of Science and Technology. 6. 4422-4431. 10.3923/jas.2013.4422.4431.

Microsoft

Validación de datos en conjuntos de datos

<https://docs.microsoft.com/es-es/visualstudio/data-tools/validate-data-in-datasets?view=vs-2019>

Sebastian-Coleman, L. (2013). Measuring Data Quality for Ongoing Improvement. 10.1016/C2011-0-07321-0.

Universidad de Castilla de la Mancha

Departamento de Sistemas y Tecnología de la información

Tesis Doctoral: Evaluación de la Calidad de los Datos en Portales WEB

Doctora: Carmen Moraga de la Rubia

Ciudad Real (España), Julio 2013

<https://ruidera.uclm.es/xmlui/bitstream/handle/10578/3401/TESIS%20Moraga%20de%20la%20Rubia.pdf?sequence=1>

ISO/IEC 25000

Calidad del producto software

Portal ISO 25000

<https://iso25000.com/>

Ramírez Coronel, Ramiro Leonardo (2014). Calidad de datos en contextos de linked data. Thesis (Master thesis), E.T.S.I. de Sistemas Informáticos (UPM) <<http://oa.upm.es/view/institution/ETSI=5FSistemas=5FInfor/>>.

Fürber, Christian & Hepp, Martin. (2011). Towards a Vocabulary for Data Quality Management in Semantic Web Architectures. 10.1145/1966901.1966903.

Universidad de Alicante. Departamento de Lenguajes y Sistemas Informáticos

Modelo de madurez para portales de datos abiertos e incorporación a la norma técnica nacional de Costa Rica

Oviedo Blanco, Edgar

2016

https://rua.ua.es/dspace/bitstream/10045/54316/1/tesis_edgar_oviedo_blanco.pdf

Corral, Yadira. (2009). Validez y confiabilidad de los instrumentos de investigación para la recolección de datos. Revista Ciencias de la Educación. 19. 228-247.

Real Academia Española
Diccionario de la Real Academia Española, edición 2017
<http://www.rae.es/>

Educando
Portal de la Educación Dominicana
<http://www.educando.edu.do/portal/>

M. Strong, Diane & Lee, Yang & Y. Wang, Richard. (2002). Data Quality in Context. Communications of the ACM. 40. 10.1145/253769.253804.

Revista Byte
Actualidad TI
<https://www.revistabyte.es/>

Communications ACM
<https://cacm.acm.org/>

Shanks, Graeme & Corbitt, Brian. (1999). Understanding data quality: Social and cultural aspects. Proceedings of the 10th Australasian Conference on Information Systems.

El Valor de la Gestión de Datos
4 Mejores prácticas para mantener la integridad de datos, Posted on Mon, Apr 10, 2017
<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/4-mejores-practicas-para-mantener-la-integridad-de-datos>

ACE Enciclopedia
<http://aceproject.org/ace-es>

Sinnexus

Business Intelligence
Informática Estratégica

https://www.sinnexus.com/business_intelligence/datawarehouse.aspx

Revista Cienciamerica

<http://cienciamerica.uti.edu.ec/openjournal/index.php/uti>