

**DESARROLLO RPA PARA MONITOREO DE CALIDAD DE DATOS Y
GENERACIÓN DE ALERTAS**

Autor:

Diego Rodríguez García

drodrigueg@eafit.edu.co

Proyecto de Grado para Optar al Título de Maestría en Ingeniería

Asesor:

Sonia Cardona Ríos

DEPARTAMENTO DE INFORMÁTICA Y SISTEMAS
ESCUELA DE INGENIERÍA
UNIVERSIDAD EAFIT
MEDELLÍN
2020

Dedicatoria

El presente trabajo lo dedico a Dios, quien me ayudo a llegar hasta aquí y darme fuerzas para continuar con en este proceso de obtener uno de los anhelos más deseados dentro de mi formación profesional.

A mi esposa y a mi hija por haberme apoyado a lo largo de este proceso y todas las personas que me animaron siempre para no desfallecer.

Agradecimientos

Le agradezco a Sonia Cardona Ríos por su apoyo, interés y dirección en la elaboración de este trabajo de titulación, a la Universidad EAFIT que me permitió afianzar mis conocimientos cada día más y me dio la oportunidad de superarme como persona y como profesional.

A mi familia quien siempre estuvo acompañándome y dándome ánimo para salir adelante y a Dios por haberme permitido llegar hasta aquí y cumplir con los objetivos deseados.

Tabla de contenido

Introducción	10
Objetivos	14
Objetivo general	14
Objetivos específicos.....	14
1 Marco teórico.....	16
1.1 Gobierno de datos.....	16
1.1.1 Qué es y para qué sirve	16
1.1.2 <i>Frameworks</i>	23
1.1.3 Principales componentes de gobierno de datos	25
1.1.3.1 <i>Organización</i>	26
1.1.3.2 <i>Políticas del gobierno de datos</i>	26
1.1.3.3 <i>Definiciones de datos y análisis</i>	28
1.1.3.4 <i>Fuente de datos</i>	28
1.1.3.5 <i>Calidad de datos y datos maestros</i>	28
1.1.3.6 <i>Operaciones de datos</i>	29
1.1.3.7 <i>Seguridad de datos</i>	29
1.2 Calidad de datos	29
1.2.1 Principales definiciones	30
1.2.2 Beneficios de la calidad de los datos (DQ).....	32
1.2.3 Gestión de la calidad de datos.....	32
1.2.4 Principales tendencias en calidad de los datos (DQ)	33
1.2.5 Problemas comunes para garantizar la calidad de los datos	36
1.3 Monitoreo de la calidad de los datos	38
1.3.1 Metodologías para medir/monitorear la calidad de datos	40
1.3.1.1 <i>Total Data Quality Management (TDQM)</i>	40
1.3.1.2 <i>Data Quality Assessment (DQA)</i>	42
1.3.1.3 <i>A Data Quality Assessment Framework (DQAF)</i>	43
1.3.1.4 <i>Framework and Methodology for Data Quality Assessment (ORME-DQ)</i> ...	45
1.3.1.5 <i>Data Warehouse Quality Methodology (DWQ)</i>	47

1.3.2	Modelo de procesos para la evaluación en calidad de datos de las organizaciones	48
1.3.3	Herramientas tecnológicas para calidad de datos	55
1.4	Automatización de procesos RPA.....	64
1.4.1	¿Qué es RPA?	64
2	Estado del arte	72
2.1	Caso de estudio 1: Allianz.....	72
2.1.1	Información del caso.....	72
2.1.2	Descripción de la empresa	72
2.1.3	Antecedentes y problemática	73
2.1.4	Descripción de la solución	74
2.1.5	Resultados obtenidos	76
2.1.6	Lecciones aprendidas	77
2.2	Caso de estudio 2: Bayer CropScience	77
2.2.1	Información del caso.....	77
2.2.2	Descripción de la empresa	78
2.2.3	Antecedentes y problemática	79
2.2.4	Descripción de la solución	82
2.2.5	Resultados obtenidos	84
2.2.6	Lecciones aprendidas	85
2.3	Caso de estudio 3: Johnson & Johnson	86
2.3.1	Información del caso.....	86
2.3.2	Descripción de la empresa	86
2.3.3	Antecedentes y problemática	86
2.3.4	Descripción de la solución	89
2.3.5	Resultados obtenidos	91
2.3.6	Lecciones aprendidas	91
2.4	Caso de estudio 4: HILTI.....	93
2.4.1	Información del caso.....	93
2.4.2	Descripción de la empresa	93
2.4.3	Antecedentes y problemática	94

2.4.4	Descripción de la solución	97
2.4.5	Resultados obtenidos	101
2.4.6	Lecciones aprendidas	102
3	Desarrollo de la solución propuesta	106
3.1	Modelo de referencia.....	106
3.2	Arquitectura de la solución	107
3.2.1	Vista lógica	107
3.2.2	Vista funcional.....	110
4	Aplicación al caso de estudio	134
4.1	Descripción del caso de estudio	134
4.2	Aplicación de la metodología.....	135
4.2.1	Definición del alcance.....	135
4.2.2	Componente capa de datos.....	137
4.2.3	Fuentes de datos	142
4.2.4	Ejecución del RPA.....	145
4.2.5	Informe de alertas	146
4.2.6	Visualizador de Excel	149
5	Conclusiones, recomendaciones y trabajos futuros.....	153
5.1	Modelo de procesos para la evaluación en calidad de datos de las organizaciones	153
5.2	Automatización de procesos RPA.....	153
5.3	Desarrollo de la Solución Propuesta	153
5.4	Aplicación al caso de estudio	154
5.5	Principales dificultades encontradas para el desarrollo de esta investigación	154
5.6	Contribución de la aplicación.....	154
5.7	Limitaciones	155
5.8	Trabajos futuros.....	155
6	Referencias bibliográficas	156

Lista de ilustraciones

<i>Ilustración 1.</i> Ventajas del IG.....	21
<i>Ilustración 2.</i> Importancia del IG.....	22
<i>Ilustración 3.</i> Componentes del IG.....	25
<i>Ilustración 4.</i> Políticas de datos.....	27
<i>Ilustración 5.</i> Esquematización de la metodología TDQM.....	41
<i>Ilustración 6.</i> Esquematización de la metodología DQA.....	43
<i>Ilustración 7.</i> Process model for data quality assessment.....	50
<i>Ilustración 8.</i> Subprocess Define Scope.....	51
<i>Ilustración 9.</i> Subprocess Define dimensions and metrics.....	51
<i>Ilustración 10.</i> Subprocess Perform measurement.....	52
<i>Ilustración 11.</i> Capacidades de monitoreo de la calidad de los datos.....	58
<i>Ilustración 12.</i> Cuadrante Mágico de Gartner 2019 sobre herramientas de calidad de datos.....	60
<i>Ilustración 13.</i> Soluciones de automatización de procesos.....	65
<i>Ilustración 14.</i> Flujo de roles.....	71
<i>Ilustración 15.</i> Importancia de la calidad de los datos de la jerarquía de productos para negocios procesados en Bayer CropScience.....	81
<i>Ilustración 16.</i> Causas de problemas en la calidad de los datos en Bayer CropScience.....	81
<i>Ilustración 17.</i> Medidas para la gestión proactiva y reactiva de los datos del cliente.....	96
<i>Ilustración 18.</i> Vista lógica.....	109
<i>Ilustración 19.</i> Vista funcional de los componentes.....	110
<i>Ilustración 20.</i> Modelo base de datos Configuración_DQ.....	112
<i>Ilustración 21.</i> Modelo base de datos Bitácora_DQ.....	114
<i>Ilustración 22.</i> Modelo base de datos BRM_DQ.....	116
<i>Ilustración 23.</i> Capa de servicios web.....	122
<i>Ilustración 24.</i> Flujo de trabajo del componente RPA.....	130
<i>Ilustración 25.</i> Resumen dimensiones de calidad de datos.....	131
<i>Ilustración 26.</i> Componente de visualización en Excel.....	132
<i>Ilustración 27.</i> Modelo de proceso pedido de ventas.....	136
<i>Ilustración 28.</i> Artefacto proceso de negocio.....	138
<i>Ilustración 29.</i> Artefacto de perfilamiento de datos.....	139
<i>Ilustración 30.</i> Regla de negocio dirección de entrega.....	139
<i>Ilustración 31.</i> Regla de negocio dirección de cliente.....	140

<i>Ilustración 32.</i> Regla de negocio precio	140
<i>Ilustración 33.</i> Regla de negocio precio ERP	140
<i>Ilustración 34.</i> Número de registro de cargue de datos pedidos de ventas.....	143
<i>Ilustración 35.</i> Modelo de datos de las entidades de pedidos ventas.....	144
<i>Ilustración 36.</i> Cabecera del pedido de venta.....	145
<i>Ilustración 37.</i> Detalle del pedido de ventas.....	145
<i>Ilustración 38.</i> Ciclos de ejecución de RPA	145
<i>Ilustración 39.</i> Informe de alerta de ejecución RPA monitoreo de calidad de datos ciclo 2.	147
<i>Ilustración 40.</i> Informe de alerta de ejecución RPA monitoreo de calidad de datos ciclo 1.	147
<i>Ilustración 41.</i> Visualizador consulta de Excel de inconsistencias	149
<i>Ilustración 42.</i> Resultado de errores en un rango de tiempo.	150
<i>Ilustración 43.</i> Resultado trazabilidad de error por dimensión precisión en un rango de fechas	151
<i>Ilustración 44.</i> Error en ciclo de tiempo del dato errado	152

Lista de tablas

<i>Tabla 1.</i> Cuadro resumen <i>frameworks</i> gobierno de datos.....	24
<i>Tabla 2.</i> Resumen dimensiones de calidad de datos.....	46
<i>Tabla 3.</i> Business Rules at AGCS	75
<i>Tabla 4.</i> Cuadro de criterios comunes casos de estudio	104

Introducción

Un dato representa una unidad, un elemento o un símbolo con cualidades propias, y su finalidad es suministrar información para generar conocimientos. Por eso se constituye en una pieza fundamental dentro de la gestión de procesos personales o empresariales y adquiere una importancia particular desde cualquier punto de vista, especialmente en aquellos casos donde el manejo de los datos se vincula con los procesos que ejecuta una organización.

Actualmente, todas las organizaciones se encuentran sumidas en un proceso de cambio ocasionado por la globalización económica que, de una u otra manera, las lleva a establecer políticas para poner en práctica estrategias, mecanismos o modelos cuyo objetivo es optimizar el volumen de la información que manejan, y con ello garantizar la calidad de los datos que integran el proceso del negocio. En esa dirección, la calidad de los datos debe comprenderse desde dos vertientes: la primera está relacionada con la confiabilidad del servicio que las organizaciones les brindan a los clientes; y, la segunda, como un factor determinante en la toma de decisiones por parte de estas. Así pues, el valor conferido a la calidad de los datos deja de ser un simple requisito de resguardo interno de la información y se convierte en una prioridad de la organización ante el crecimiento y las demandas del mercado, lo cual les exigen garantías a las organizaciones en el manejo de los datos.

En líneas generales, la importancia que tiene la calidad de los datos, también llamada Data Quality Reside, es el beneficio que le aporta a los “procesos operativos” que ejecuta una empresa; y el valor que ha adquirido al garantizar la complejidad, la precisión, y la actualización de los procesos con base en la toma de decisiones. En consecuencia, la calidad de los datos se torna en un componente central relevante de la implementación de una cultura organizativa empresarial accesible, maleable, cimentada en políticas que generen confianza y credibilidad; pues es un “activo estratégico” que enaltece el valor de cualquier organización.

Por lo anterior, se afirma que la calidad de los datos es un elemento imprescindible en este contexto porque el manejo de ellos es un requerimiento operativo para las empresas, una realidad que ahora forma parte de su presente y, por tanto, no puede ni debe dejarse a un lado. Por el contrario, las organizaciones están llamadas a asegurar la calidad de sus datos y a asumir ese desafío competitivo; dado que, con el control del gobierno de datos y al tomar como base de los procesos de negocio el monitoreo de la calidad de los datos, lo que se persigue es: evitar errores, extender la “completitud”, obtener seguridad, exactitud, integridad o limpieza de los datos y la correspondencia. De ese modo, las organizaciones pueden establecer una relación exacta en la calidad de los datos de los procesos que conforman el negocio.

No obstante, la iniciativa de implementar un modelo de proceso para evaluar la calidad de los datos, detectar errores y generar alertas conlleva el uso de métodos, técnicas y herramientas fundamentadas en tres aspectos clave, a saber: i) una organización con una cultura amplificada, ii) gobernanza de datos, y iii) el uso de las innovaciones tecnológicas como mecanismo de soporte para implementar propuestas de mejores prácticas. Cabe anotar que el gobierno de datos fehaciente depende de la calidad de los datos con los cuales cuente la organización, ello significa que estos procesos no pueden funcionar aislados, sino que deben fusionarse para poder alcanzar buenos resultados.

De allí el interés por realizar este trabajo de investigación orientado a las decisiones de la gobernanza de datos, teniendo en cuenta uno de los elementos constitutivos y más relevantes de este proceso, esto es: la calidad de datos. Este último se estudia en el documento desde una perspectiva generalizada, para lo cual se han considerado aspectos vitales que permiten entender su importancia, analizarlos, sistematizarlos y presentar alternativas de solución enmarcadas en las mejores prácticas como un desafío para construir un modelo de negocio consistente, seguro, confiable y que sea comprensible para el usuario.

Lo anterior se llevó a cabo con el propósito de implementar una automatización mediante una tecnología RPA (*Robotic Process Automation*) para monitorear y generar alertas relacionadas con la calidad de los datos de los procesos de negocio de una organización, lo cual deben tener en cuenta las organizaciones por los grandes volúmenes de datos que manipulan actualmente y que de alguna manera ponen en riesgo la estabilidad de la empresa.

Ahora bien, la tecnología RPA, también conocida como metodología RPA, es un método que se utiliza en la automatización de los procesos debidamente reglamentados en normas específicas. Por ello es considerado un software que se realimenta del usuario encargado del modelo del negocio y es ideal para la ejecución de multitareas, además de ser maleable, adaptable a los requerimientos de los procesos de la organización y funcionar en plataformas heterogéneas. De modo que esta metodología puede emplearse en varias funciones empresariales, aunque su implementación requiere que se hagan ajustes previos tanto tecnológicos como estructurales y a nivel humano, este último hace referencia a la actualización de la asignación de roles conforme con los cambios que gesta este método de trabajo (Deloitte, 2017).

De acuerdo con Deloitte (2017) la metodología RPA genera varios beneficios para las empresas. Entre ellos, el autor destacó los siguientes: la adición de controles, el incremento eficaz de los procedimientos que realizan las organizaciones, establece modelos, mejora el tiempo de transferencia y reduce los costos de funcionalidad.

Sin embargo, es importante acotar que la metodología RPA no se usa exclusivamente para monitorear los datos, sino que también tiene otras aplicaciones dentro del desarrollo de los proyectos de innovación, pero que no se abordan en este documento porque no guardan relación con el objeto de la investigación. En ese sentido, este método de trabajo se tuvo en cuenta por su versatilidad y amigabilidad para articular con un *framework* como Cobit (*Control Objectives for Information and related Technology*), que está fundamentado en los

estándares de calidad que se utilizan a nivel mundial como son “ISO 27000, ITIL, TOGAF, PMBOK, DMBOK, COSO, PRINCE2”.

Otro punto que conforma estos elementos es la definición de datos que, junto con la fuente de datos, la calidad de datos y datos maestros constituye el eje central para gestionar las operaciones relacionadas con el manejo de los datos. Cabe anotar que esta gestión se dará de una manera confiable al estar bajo un gobierno de datos y procesos de negocio que son aptos para garantizar y salvaguardar la seguridad que estos requieren, con base en operaciones de actualización y regulación.

En suma, la gestión de la calidad de los datos a través del gobierno de datos, áreas de Tecnologías de Información (TI) y procesos de negocio contribuye en los procesos de definición y asignación de roles funcionales; establecimiento de normas, responsabilidades, procedimentales para obtener datos y darles mantenimiento, estructurarlos y almacenarlos para ser objeto de análisis. Asimismo, en este documento se aborda el estudio de un Modelo de Procesos de Evaluación de Calidad de Datos, el cual esboza una metodología que sirve de apoyo para monitorear y medir la calidad de los datos, así como también para analizar y generar alertas. Dicho modelo se sustenta en una diversidad de métodos que aportan información valiosa a la investigación, entre los cuales resaltan los siguientes:

- Total Data Quality Management (TDQM): se basa en el análisis de los datos, le suministra información al usuario sobre la alta calidad de los datos, y provee el uso de políticas idóneas para gestionar la calidad de datos integrales en una empresa.
- La metodología Data Quality Assessment (DQA): se fundamenta en la evaluación de la competitividad de los sistemas que usa la organización para capturar los datos, transferirlos, registrarlos y comunicar resultados tanto cualitativos como cuantitativos.

- A Data Quality Assessment Framework (DQAF): su función como método de medición y monitoreo es establecer las capacidades de la organización para llevar a cabo una auditoría.
- Framework and Methodology for Data Quality Assessment (ORME-DQ): se caracteriza por el conjunto de recursos que posee para ejecutar el proceso de medición. Este abarca: la organización, los procesos, el modelado del negocio, la asociación de los datos y los flujos de datos.
- Data Warehouse Quality Methodology (DWQ): es una metodología innovadora que nutre el estudio y el análisis de la DB de metadatos basados en estándares formales de calidad de datos y, por ende, de información con la que contribuye favorablemente.

Objetivos

Objetivo general

Diseñar e implementar un mecanismo automatizado para monitorear y generar alertas, teniendo como base la adaptación de un modelo metodológico, en aspectos relacionados con la calidad de datos contenidos en fuentes de información; con el objetivo de brindar mayor eficiencia, asertividad y oportunidad en la toma de decisiones que forma parte de la gestión de dicho Gobierno de Información (IG) organizativo.

Objetivos específicos

- Identificar marcos y modelos metodológicos propuestos para monitorear la calidad de datos en las organizaciones, con el fin de determinar los componentes, las técnicas, las herramientas y las tecnologías esenciales para habilitar su implementación.
- Diseñar y construir un mecanismo automatizado tipo RPA para monitorear, medir, evaluar y generar alertas en aspectos de calidad de datos organizacional, tomando como

base una adaptación de un modelo metodológico en aspectos relacionados con la calidad de datos.

- Aplicar la solución RPA desarrollada en un caso de negocio en una empresa del sector real.
- Sintetizar los hallazgos generados a partir de la aplicación práctica y las lecciones aprendidas de todo el proceso desarrollado.

La forma como se ha estructurado el documento es la siguiente: en la primera sección se ha abordado la introducción al objeto de estudio y los objetivos que se persiguieron en la investigación. Seguidamente, en el segundo apartado se presenta el desarrollo teórico de las variables y los componentes más importantes que integran la investigación. Luego, en el tercer capítulo se desarrolla el estado del arte y se exponen algunos casos de estudios con temas que guardan una relación con la calidad de los datos, lo que permitió identificar la inconsistencia en los datos en los sistemas de información; además, se encontró que esos estudios se relacionan con el monitoreo y la gestión de la calidad de los datos en los procesos de algunas organizaciones.

Posteriormente, en el cuarto capítulo se plantea una solución, mostrando la arquitectura y describiendo todos sus componentes funcionales, así como su interrelación; y en el quinto apartado se presenta el piloto de la aplicación de esta propuesta de solución al caso de estudio: una empresa del sector farmacéutico. Cabe anotar que la aplicación se realizó en unos procesos de negocio de dicha organización como referente de validación para la propuesta. Finalmente, en el sexto capítulo se presentan las conclusiones de la propuesta y se sugieren mejoras o adiciones a esta con base en los hallazgos obtenidos en la investigación.

1 Marco teórico

1.1 Gobierno de datos

1.1.1 Qué es y para qué sirve

El gobierno de datos es un método de vigilancia de calidad que permite adicionar mayor firmeza y disciplina en el complejo procedimiento de dirigir, aplicar, optimizar y resguardar la información dentro de las empresas u organizaciones. Esta práctica ha ganado espacios en la toma de decisiones porque brinda seguridad en cualquier operación que requiera de un análisis informativo (Aiken et al., 2007), además de configurarse en una herramienta novedosa para el avance empresarial y proponer varios modelos fundamentados en el manejo de las bases de datos. Adicionalmente, algunos autores como Castillo (2015) asumen que el gobierno de datos es una práctica de empoderamiento y vigilancia que desempeña un papel relevante en los procesos de proyección, ejecución y seguimiento a la gestión eficaz de los datos.

Así pues, el gobierno de dato se acepta como una política empresarial que parte estratégicamente desde la estructura más alta de la empresa para involucrar a la parte operativa y conformar un marco articulado entre la “gestión de datos y el modelo del negocio”. En ese contexto se han realizado varias investigaciones que evidencian lo importante que es esta práctica dentro de una organización, como la realizada por Colina (2019), quien estudió “el gobierno de datos como un referencial entre el gobierno TI y la inteligencia de negocios”, y apuntó que los procesos de manejo de datos y su administración se sofistican cada vez más, dando nuevas directrices de manipulación y manejo informativo. Acorde a la investigación hecha por Colina se puede decir que las empresas y las organizaciones incrementan cada día los volúmenes de información, lo que tiene fuertes implicaciones para la toma de decisiones acertadas o erradas.

Por lo tanto, es necesario que —a nivel organizacional o empresarial— se equilibren los datos del negocio y el gobierno TI, con el objetivo de centralizar los procesos de la empresa u organización para darles el control a los usuarios. Pues la ventaja de la gestión TI es que les permite a los usuarios examinar los datos y responder sus inquietudes en tiempo real, sin que ello requiera que la gestión TI restaure o cree aplicaciones aisladas a la ejecución con lo cual se garantiza una administración de datos confiable, sólida, íntegra y firme. Igualmente, Colina estableció que el gobierno de datos se sustenta en elementos teóricos concretos que ayudan a la organización u empresa a trazar metas dinámicas y a proveer al usuario de resultados que incitan intrínsecamente a los grupos de trabajo; así, por medio del gobierno de datos, se alcanza la concepción de formas de conocimiento ilimitado desde un punto de vista analítico, para fortalecer la enunciación de políticas y procesos propios de un gobierno acorde a un negocio perceptible con capacidades para desarrollar herramientas de datos integrales.

En el estudio citado también se encontró que el gobierno de datos proporciona la optimización de los datos agrupados, además de adoptar datos referenciados y estandarizados, y herramientas unificadas para gestionar, integrar, informar y analizar. De ese modo, la investigación de Colina permitió conocer el ambiente bajo el cual se gestiona el gobierno de datos, y la forma como este le brinda a la empresa u organización las herramientas necesarias para planificar y lograr las metas de una manera sistemática y consciente. Por ende, se considera que los aportes teóricos extraídos del estudio de Colina (2019) son fundamentales para entender con mayor precisión el gobierno de datos, así como también su aplicación en las empresas u organizaciones y su incidencia en la toma de decisiones.

Por su parte, Garcés (2016) estudió la implementación del gobierno de datos en empresas y organizaciones que presentaban problemas referentes al manejo de la información, y a los efectos de esta en la toma de decisiones. En su investigación Garcés

comprobó que existe un número significativo de empresas que no cuentan con procesos ni políticas que garanticen la seguridad de los datos y, con miras a solucionar dichas eventualidades, propuso un modelo de gobierno de datos centrado en el prototipo de madurez de gobernabilidad como alternativa para instaurar el ordenamiento y el control de los datos. Aunque, cabe anotar, actualmente se cuenta con una diversidad de modelos que le da libertad al usuario para seleccionar el que mejor se ajuste a las demandas de su organización u empresa.

En ese orden de ideas, la autora sugirió el uso de esta herramienta como estrategia emprendedora del gobierno de datos, dado que les permite a las empresas u organizaciones determinar su estatus teniendo en cuenta el nivel actual en el que se encuentran, y también les muestra a los usuarios el procedimiento. Específicamente, Garcés recomendó los modelos “EIM-Gartner, IBM, DataFlux, MDM Institute, Oracle y Kárido”, los cuales permiten abordar los problemas organizacionales y vinculan tanto a las personas como la tecnología y los procesos. Así pues, se considera que la investigación de Garcés es útil, porque permite visualizar los distintos métodos y seleccionar el más apropiado de acuerdo con las características de las empresas.

Ahora bien, existen otros autores que, al igual que Garcés, han afirmado que el sistema de madurez gubernamental es la mejor opción para implementar el gobierno de datos. Uno de los más destacados en ese sentido es Saltz (2017), quien habló de las dificultades que a menudo presentan las organizaciones empresariales para manipular numerosas cantidades de datos, las cuales son producto de la celeridad, la complejidad, el volumen, la autenticidad y el valor que se incrementan a velocidades vertiginosas en “comparación con los recursos informáticos”. Conforme a la postura de esos autores, el modelo de madurez más adecuado es aquel que le proporciona a la organización datos transparentes en relación con la información que se maneja y la toma de decisiones.

Otros investigadores como Smits y Van Hillegersberg (2015) también comparten el criterio de Garcés, pues consideran que el gobierno de datos es una herramienta de trabajo que le brinda seguridad a la empresa en el manejo de la información y, por ende, en la toma de decisiones. Estos autores también plantearon que la madurez va de la mano del avance de las organizaciones, y propusieron un modelo de madurez ITG (MM) que se amolda tanto a los procesos como la estructura y la cultura organizacional. Frente a lo anterior, Buchwald, Urbach y Ahlemann (2014) estudiaron los factores que afectan al gobierno TI para evaluar la importancia que tiene un gobierno de datos basado en este (gobierno TI), con lo cual demostraron la eficiencia en las empresas del gobierno de datos, y lograron consolidar la teoría de que este modelo genera innumerables bondades que pueden ser utilizadas en beneficio de la empresa u organización.

Por otra parte, Salah (2017) hace referencia a las capacidades y al desempeño del gobierno de datos en su investigación, donde demostró que el modelo no solo proporciona seguridad en la gestión de los datos, sino que además es sustentable y confiable para el manejo de información empresarial. Respecto a este último punto, Mesa y Cotelo (2015) estudiaron cómo se da y se relaciona el manejo de la información con la toma de decisiones empresariales fundamentada en “valores, posiciones y transacciones, datos de mercado y datos de referencia”, y concluyeron que la mejor alternativa para implementar un programa de gobierno de datos son los modelos “Data Management Association (DAMA), Data Governance Institute y Data Management Maturity Model”.

Basado en los aportes generados por los autores referidos se argumenta que el gobierno de datos es un modelo que forma parte del negocio e involucra tanto a los “(stakeholders) de la empresa” como a los usuarios o quienes guardan relación con el modelo de negocio, pero de un modo equilibrado, es decir, engrana el modelo de negocio y las TI para garantizar una gestión de datos eficiente y confiable. Ello significa que el mayor

beneficio que provee un gobierno de datos está asociado con el fortalecimiento funcional de la organización, dado que convierte los datos manipulados por la organización empresarial en una fuente de información flexible y adaptable en pro de una gestión de negocio evidente. Por eso a esta práctica se le asocian funciones claras, precisas y bien orientadas dentro de la organización para gestionar los datos, algunas de las cuales se relacionan con el establecimiento de una estructura organizacional que regula la “integridad de la información” mediante reglamentos y principios donde establecen las obligaciones contraídas.

Asimismo, el gobierno de datos define la administración de políticas garantes del cumplimiento de los objetivos propuestos en el modelo de negocio y en la creación de datos. La instauración de dichas políticas administra los datos, y va desde la alta gerencia hasta la persona que maneja los datos, quien los da de alta, bloquea, modifica, valida, registra cambios y muestra resultados. Sin embargo, para consolidar una administración correcta se requiere de políticas globales que cubran todos los espacios del modelo de negocio junto con el manejo de datos.

En este orden de ideas también es importante resaltar que en la mayoría de las organizaciones existe un órgano central a quien le corresponde darle una dirección a la empresa, asegurar el cumplimiento de las obligaciones contractuales existentes y verificar que las regulaciones de los procesos internos y externos realmente cumplan con los lineamientos de calidad establecidos. De allí que, por lo general, la documentación gire en torno de dicho gobierno, conste y fluya de forma jerárquica partiendo de la “misión, visión y valores” de la empresa hasta los procedimientos, procesos y flujos de información que existen en el negocio.

Respecto a lo anterior, Deloitte (2015) resaltó que es necesario que toda organización cuente con un gobierno de datos, además de señalar los diversos factores que conducen a adoptar esta medida, una de las cuales involucra el abordaje de la gestión de datos dentro de

la empresa desde una perspectiva real y valorar los datos como parte elemental de la operatividad. Ahora bien, aunque la optimización de la gestión de datos determina un fructífero desarrollo empresarial, se debe contar con un marco que reglamente la gestión (diseño, inspección y vigilancia) e incluya el personal administrativo de la empresa y al departamento tecnológico. Ello muestra lo útil que resulta el gobierno de datos en el funcionamiento eficiente de todos los datos de la organización, dado que este: garantiza que los datos satisfacen las demandas, minimiza los costos de gestión y resguardo de los datos con base en normas establecidas, y preserva la privacidad de la información. Así pues, queda claro que una organización que no cuenta con una reglamentación adecuada, que además garantice la calidad de sus datos, pondrá en duda su funcionamiento y, por esa razón y las ventajas que aporta (ver Ilustración 1), el gobierno de datos se convierte en una política empresarial necesaria para las organizaciones.

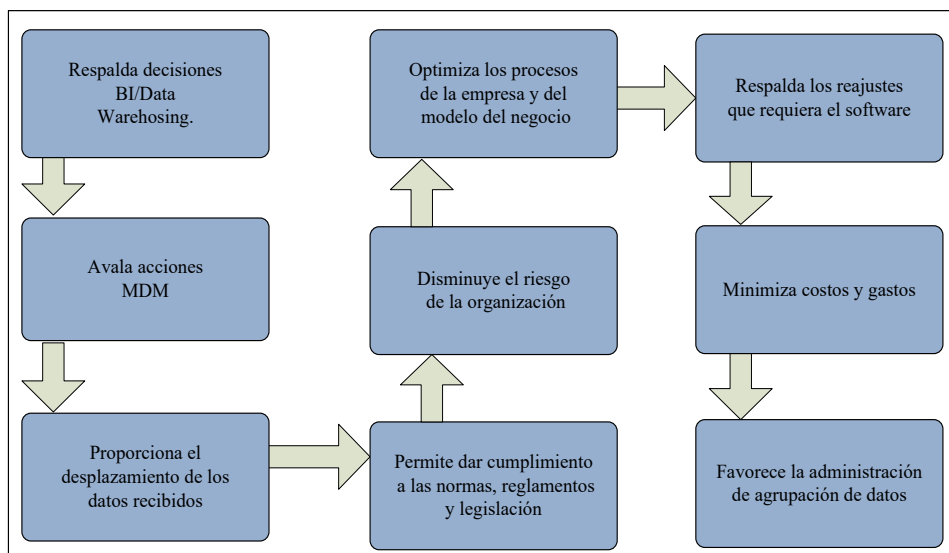


Ilustración 1. Ventajas del IG

Fuente: elaboración propia

Contrario a Deloitte (2015), los autores Marulanda, López y Valencia (2017) argumentaron que el uso del gobierno de datos debe evaluarse desde diversas vertientes, por la facilidad que brinda para acceder a los datos de forma pertinente, su confiabilidad y su

ajuste al establecimiento previo de políticas con sentido generalizado. Igualmente, este garantiza que los datos manejados dentro de la organización realmente cumplan con las demandas de las partes, es decir, tanto de la organización empresarial como del cliente.

Los datos son administrados como un activo que forma parte de la organización y de la funcionalidad de esta, por eso tramitar e impulsar los datos como activo empresarial permite cumplir con los requerimientos de quienes se benefician internamente, así como también de los clientes, lo cual contribuye significativamente a la toma de decisiones acertadas. Otro punto relevante y ventajoso del gobierno de datos es que conserva la integridad de los datos, porque prevé incompatibilidades entre las aplicaciones o los sistemas con características diferentes y evita la pérdida de datos durante la operatividad del software, al valorar la operatividad o funcionalidad, o bien al proveer algún tipo de información. Debido a que sus respuestas están sujetas y dependen de las demandas presentes el gobierno de datos puede establecer directrices que posibilitan el uso, la seguridad y la integridad de los datos.

Además, estos autores sostuvieron que el gobierno de datos complementa el valor de la empresa, dado que determina e implanta modelos de comunicación que abren espacios para entender, organizar y sistematizar la información confiable, la cual sirve de base para tomar decisiones acertadas. Para los investigadores la importancia de utilizar un gobierno de datos se reduce a la información expuesta en la Ilustración 2:

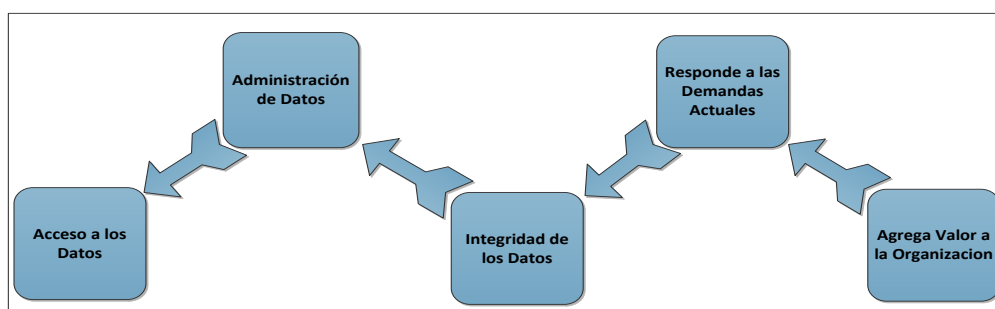


Ilustración 2. Importancia del IG

Fuente: elaboración propia

De lo anterior se concluye que la falta de implementación de un gobierno de datos o las debilidades que presente el modelo implementado incide en la funcionalidad de la organización. Puesto que el gobierno de datos permite que los departamentos de una misma empresa tengan una comunicación fluida, lo cual ayuda a tener conocimiento y control sobre la información que ingresa o sale de la organización. Pero cuando no existe esa interacción, generada por un modelo de gobierno de datos, no se puede responder oportunamente a las demandas de los usuarios (internos o externos), se retrasa el trabajo, se acumulan las tareas, ocasiona un descontrol en la administración de los datos, el intercambio y la coordinación de información entre los departamentos no es acertada, la asignación de roles no es visible, y las estrategias y los procesos no son determinados de manera consensuada.

1.1.2 Frameworks

Governance (s.f.) explicó que las *frameworks* de un gobierno de datos son una estructura que guía la reglamentación las funciones, los roles, la toma de decisiones y todos los procedimientos que implican el manejo de datos y de información. Estas incluyen –como parte de su conformación estructural– el apoyo económico, de administración e integración de los usuarios a todos los procesos, así como la definición de una normativa y el impulso de la organización. Es así como las empresas se han visto en la necesidad de implementar marcos para gestionar sus datos, entre los cuales destacan los siguientes (ver Tabla 1):

Tabla 1. Cuadro resumen *frameworks* gobierno de datos

Fuente: elaboración propia

<i>Framework</i>	<i>Descripción</i>
Cobit©4.1 (Control Objectives for Information and related Technology).	<ul style="list-style-type: none"> • Estudia, desarrolla y promueve un “marco de control” de gobierno de datos. • Está dirigido al modelo del negocio y se basa en los requerimientos, las demandas de información, los procesos del negocio y los datos de la empresa. • Orienta los procesos del negocio a través de un modelo referencial y un lenguaje único para facilitar la administración de los datos por todos los usuarios del sistema. • Se fundamenta en el establecimiento de controles y consta de un proceso de gestión de alto nivel y objetivos para establecer controles precisos. • Orientado a procesos de medición, es decir, mide el desempeño de los procesos del negocio, determina las áreas críticas y propone herramientas o modelos empresariales de monitoreo (Muñoz, 2012).
Cobit 5	<ul style="list-style-type: none"> • Basada en estándares utilizados a nivel mundial, entre los cuales se encuentran: “ISO 27000, ITIL, TOGAF, PMBOK, DMBOK, COSO, PRINCE2”. • Ejecución de procesos de control a través de la comprensión de los requerimientos de la organización, el resguardo de la información, la fundamentación en una guía única de gestión y gobernanza de datos. • Estudia los eventos considerando todas las interacciones que lo producen para establecer diferencias entre la gestión y el gobierno de datos (Gobierno, 2017).
Norma ISO 9001 (2008).	<ul style="list-style-type: none"> • Determina las herramientas requeridas para brindar información confiable a quienes forman parte de la organización y a terceros. • Aseguramiento de la privacidad de la información personal. • Establece procedimientos garantes de la calidad de datos. • Gestiona procedimientos relacionados con la arquitectura de los datos. • Gestiona metadatos como parte de los procesos del negocio. • Admite la asociación inteligente de los procesos del negocio con el almacenamiento de los datos. • Creación de inventarios a partir del flujo de datos. • Cimenta el cumplimiento de “políticas de calidad”, integrando a los procesos del negocio “mejores prácticas métricas de procesos”.
ISO/IEC 38500	<ul style="list-style-type: none"> • Acoplamiento de metadatos agrupada (Portilla, 2018). • Establecimiento de estándares de calidad para la gobernanza de procesos, y toma de decisiones organizacionales respecto al manejo de los datos. • Garantiza la calidad de los datos y la información a partir de la aplicación de normas. • Dirige a los usuarios de los datos en el control y uso de estos. • Provee las bases pertinentes para evaluar la calidad de los datos. • Pone a disposición de las organizaciones mecanismo regulatorios para administrar, monitorizar y evaluar la calidad de los datos (Ballester, 2010).

1.1.3 Principales componentes de gobierno de datos

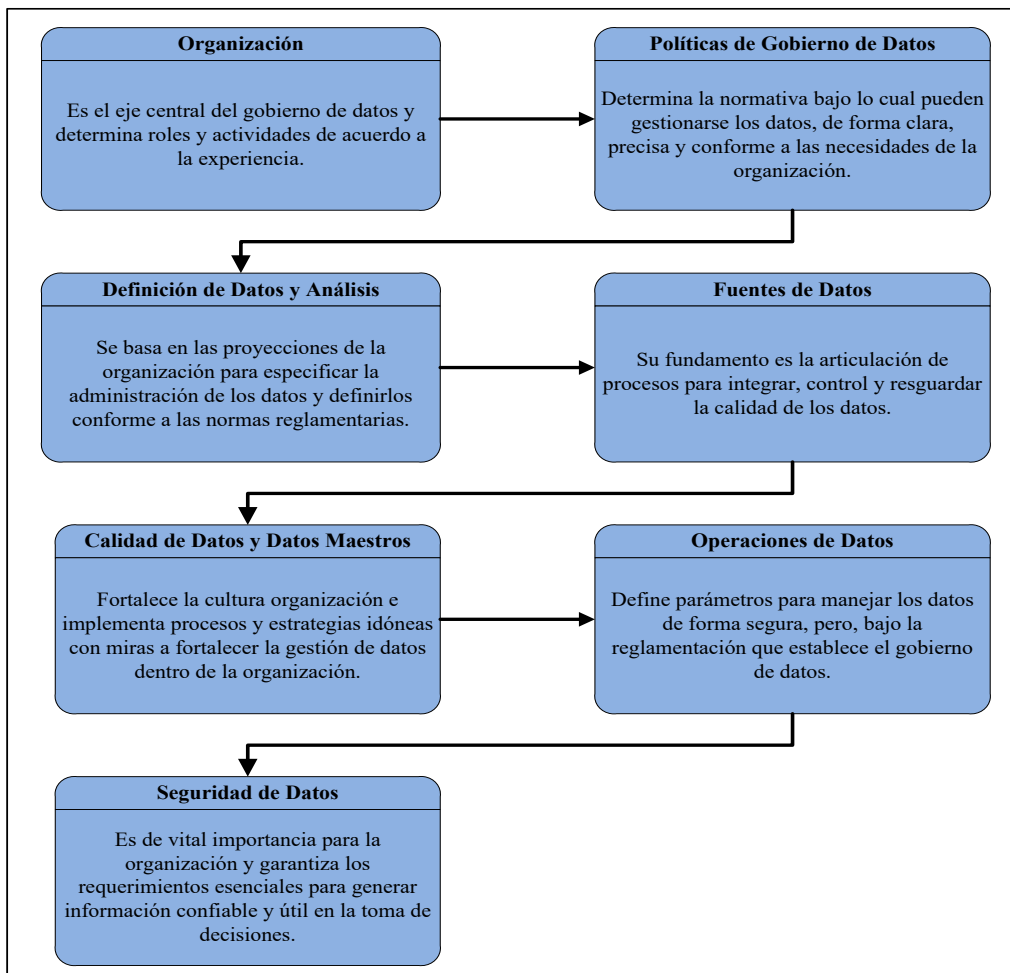


Ilustración 3. Componentes del IG

Fuente: elaboración propia

En la Ilustración 3 se sintetizan los componentes del gobierno de datos, entre los cuales destacan: el gobierno de datos, las políticas bajo las cuales puede operar un gobierno de datos, la definición de los datos y el análisis de estos, la fuente de datos como proceso fundamental para una gestión eficiente, la calidad de datos y los datos maestros, las operaciones de datos y la seguridad de los datos. Seguidamente, se describen los aspectos

más relevantes de cada uno de ellos, con la finalidad de brindar una visión clara y comprensible del papel que estos desempeñan dentro de la organización.

1.1.3.1 Organización

El MinTIC (2016) se refirió a la importancia que tiene la organización como parte del proceso de gobierno de datos, y explicó que gracias a esta se pueden especificar los roles y la función de cada uno de los usuarios que se relacionan con el manejo de los datos. Por lo tanto, sobre el gobierno de datos recae la responsabilidad de asignar funciones, especificar los roles de los usuarios conforme a sus capacidades, definir las acciones ajustadas al plan de seguridad de datos y de la información buscando satisfacer las demandas funcionales de la organización.

1.1.3.2 Políticas del gobierno de datos

En cuanto a las políticas del gobierno de datos, Félix et al. (2018) las definieron como todas las directrices o normas que protegen el establecimiento, la recuperación, la integridad, los atributos, la seguridad y el manejo de los datos. Estas deben comprender los modelos de datos y los procesos involucrados o aquellos que conforman su gestión; de ese modo responden a las necesidades de la organización, es decir, a los requerimientos y a las posibles soluciones, pero también descartan aquello que no requiere la organización a fin de optimizar los procesos con fundamento en el manejo adecuado de los datos.

Por otro lado, la Alcaldía Mayor de Bogotá (2018) señaló, en la Política de Gobierno de Datos, que su efectividad recae sobre las normas y reglas que se apliquen. Sin embargo, se requiere de la colaboración del potencial humano para poder lograr los objetivos, así como también para que se consolide y se sostenga el modelo implantado. En

esa dirección, se propuso una serie de políticas específicas (ver Ilustración 4) en las que se toman en cuenta los siguientes elementos:

- Administración del Diccionario de Datos: consiste en determinar los datos críticos dentro del proceso.
- Administración del gobierno de datos: se fundamenta en la configuración e instauración del gobierno de datos, los roles, las obligaciones, las funciones y las tareas que cada participante debe desarrollar.
- Administración de los procedimientos para establecer la calidad de datos: se basa en la caracterización de los datos críticos, la creación de indicadores, la organización de herramientas para tener dominio sobre los indicadores, la generación de alarmas de advertencias ante fallas en las actividades y la corrección oportuna de estas.
- Administración de secuencia: centra la atención en las tareas vinculadas al uso o puesta en práctica de las políticas del gobierno de datos.
- Creación de escenarios para el desarrollo de políticas de gobierno de datos junto con otros procedimientos. En síntesis, las políticas del gobierno de datos permiten:



Ilustración 4. Políticas de datos

Fuente: elaboración propia

1.1.3.3 *Definiciones de datos y análisis*

Según Hernández, Mayorga y Peña (2012) la definición y análisis de los datos parte de las proyecciones de la empresa y se cristaliza con la gestión, la calidad de los datos, en la administración del modelo de negocio y en la administración de los riesgos. Esta conceptualización engloba los procedimientos realizados por la organización para definir los datos, en ese sentido, el gobierno de datos garantiza que la definición de los datos se ejecute bajo los parámetros establecidos en la reglamentación pertinente.

1.1.3.4 *Fuente de datos*

Por su parte, Caballero, Blanco y Piattini (2014) mencionaron que el modelo de gobierno de datos prioriza la fuente de datos por ser un elemento representativo dentro de la gestión de la calidad de datos. Pues la finalidad de este modelo es garantizar el uso adecuado de los datos, esto es: la integridad, la seguridad y la privacidad deben estar articuladas, lo cual requiere una fuente de datos confiable y eficaz que permita ejecutar procesos transparentes.

1.1.3.5 *Calidad de datos y datos maestros*

Los autores Hernández, Mayorga y Peña (2012) expresaron que el gobierno en la calidad de los datos es vital, porque este representa el punto de partida para implementar los procesos y las estrategias en función de la cultura organizacional y del modelo de negocio, y a través de esta dupla enaltece la eficacia de los datos. Desde esa perspectiva, esta práctica provee las herramientas necesarias para llevar adelante la gestión de datos, detectar errores y hacer las correcciones pertinentes, es decir, su función es optimizar la calidad de los datos por medio de operaciones como el control, la actualización y las consultas.

1.1.3.6 Operaciones de datos

El cuidado y la protección de los datos deben definirse bajo requerimientos estrictos que les permitan a los usuarios recuperar los datos. A su vez, las operaciones de datos deben cumplir con las reglamentaciones necesarias para salvaguardar la información, y establecer los niveles del sistema por departamento (Caballero et al., 2014).

1.1.3.7 Seguridad de datos

La seguridad de los datos involucra la planeación, el desarrollo y la puesta en marcha de políticas y procesos de seguridad que deben garantizar: la veracidad, la libertad, el acceso a los datos y una correcta auditoría de los datos, por ser activos esenciales de la organización para generar información. Ahora bien, el gobierno de datos asienta normas y mecanismos de seguridad de datos idóneos para garantizarle al usuario el uso y la restauración de los datos de una manera adecuada, así como también la confiabilidad y la privacidad de la información.

1.2 Calidad de datos

Los datos representan el eje central en el manejo de la información y su calidad es directamente proporcional a la toma de decisiones acertadas o erradas. Actualmente, la gestión de datos se ha convertido en un elemento clave a nivel empresarial, porque manejar los datos sin un control o sin contar con estándares de calidad pone en riesgo la administración de los procesos de negocio. De allí la relevancia de examinar a profundidad el tema, conocer los pros y los contras de los procedimientos para el resguardo y mantenimiento de los datos, pues, aunque existen técnicas, modelos y herramientas innovadores que permiten evaluar la calidad de los datos, muchas las organizaciones no

tienen consistencia en sus datos, lo cual les resta confiabilidad y competitividad en el mercado.

1.2.1 Principales definiciones

De acuerdo con lo señalado en la Norma ISO (2008), la calidad de los datos es una caracterización propia de la información porque hace referencia a la categoría de los datos, la fuente que los suministra, las necesidades que los originan y las condiciones bajo las cuales se gestionan. En palabras de Serrano (2014), la calidad de los datos se refiere a las cualidades que tiene un dato como unidad única de información, y depende exclusivamente de la definición, el análisis, el tratamiento, el procesamiento y la administración en sí del dato como activo de la organización.

Así, la calidad de los datos (en adelante DQ) cobra un papel fundamental dentro de la implementación y la administración de los sistemas de información, porque no solo garantiza la eficiencia del sistema, sino también las respuestas que este le puede dar al usuario. De la calidad, que es la esencia de todo sistema, se desprende la confiabilidad y la eficiencia de los datos que son manipulados y procesados para convertirlos en información. Al respecto, los autores Corrales, Ledezma y Corrales (2016) explicaron que los avances tecnológicos le han dado apertura a un escenario que le permite al usuario aproximarse a nuevos cambios en el procesamiento y el reprocesamiento centrado en la calidad de los datos. Según los planteamientos de los autores citados, la calidad juega un papel importante dentro del manejo de los datos y, por lo tanto, deben ser garantizados coherentemente con la meta trazada y previamente establecidos en el preproceso.

Por su parte, Muñoz (2012), al igual que Soto (2014) y Corrales, Ledezma y Corrales (2016), afirmó que la calidad de los datos es un desafío al que están expuestas las

empresas, por consiguiente, estas deben ajustarse a las nuevas demandas de procesamiento y reprocesamiento bajo ciertos parámetros que determinen la DQ. Para Batini et al. (2007) este tema es novedoso y surge como respuesta a las necesidades de las organizaciones empresariales, con el fin de ejecutar operaciones efectivas en el manejo y el procesamiento de los datos, en lo cual tiene vital importancia la valoración y la vigilancia de la DQ. La propuesta de los autores para evaluar la calidad de los datos se concentra en la utilización de ORME-DQ, un método de trabajo complejo que busca establecer un equilibrio lógico entre la integridad y la viabilidad práctica en el procedimiento de la DQ.

Otros estudiosos como Stenström, Mustafa y Parida (2014) hicieron referencia a la importancia que tiene la DQ dentro de la planificación y la ejecución del mantenimiento en el estudio de los datos. Y argumentaron que una empresa con una calidad de datos decadentes no puede brindarles un servicio eficiente a sus clientes, toma decisiones débiles, y genera el impacto contrario al esperado cuando implementa las estrategias. Así pues, los autores plantearon que una buena evaluación de la DQ se puede efectuar mediante un estudio bien definido de la relación existente entre los datos, los metadatos y los prototipos que sirven para representar los datos.

En contraste con estos autores, Valverde (2014) señaló que la DQ se basa en una metodología que resulta de diversos experimentos, y propuso un diseño de un modelo que está fundamentado en definiciones claras y bien referenciadas que, además, cuentan con antecedentes de otros estudios, los cuales guían y soportan el diseño de un modelo confiable. También expuso que el principio de evaluación mediante este modelo se inicia con la medición de las dimensiones de los “atributos”. Cabe resaltar que el modelo propuesto no causa ningún tipo de retraso en la ejecución de las tareas cotidianas de la empresa, por ende, tiene incidencia en la toma de decisiones; además, sugiere que la medición

de la DQ debe ser “subjetiva y objetiva”, o que esta debe ser evaluada desde una perspectiva cualitativa y cuantitativa para tener una valoración efectiva, lo que requiere la extracción previa de los atributos que serán medidos.

De lo anterior se concluye que la gestión empresarial es un proceso complejo que demanda planificación e innovación para impulsar su crecimiento y desarrollo. De allí que el manejo y el control que se pueda tener sobre la información tenga un carácter imperante, porque de ello depende –en gran medida– la seguridad de los procesos que se ejecutan dentro de su seno. Actualmente, son pocas las empresas que gestionan y controlan la información de manera manual, pues la mayoría de ellas ha establecido sistemas de control automatizados o semiautomatizados para resguardar los datos, aunque ello no garantiza la fiabilidad de la información, porque este proceso siempre está sujeto a puntos críticos que deben ser monitoreados y atendidos oportunamente para evitar daños severos a la data que se maneja.

1.2.2 Beneficios de la calidad de los datos (DQ)

Los beneficios de la DQ se miden por la transparencia y la eficacia de la gestión empresarial. Así pues, la calidad se expresa por medio de los datos maestros, lo cual implica que estos deben tener un nivel de integridad alta, ser sólidos, exactos, confidenciales y estar ajustados a los estándares de calidad empresariales. Por tanto, los beneficios que aporta la DQ a una empresa se pueden determinar por lo siguiente:

1.2.3 Gestión de la calidad de datos

Respecto a la gestión de datos Portilla (2018) argumentó que esta debe ser eficiente para asegurar el desarrollo de una organización, dado que, en la actualidad, es una herramienta que le aporta beneficios a la empresa, le agrega valor a los datos, al

crecimiento y desarrollo organizacional ante las demandas internas y externas. De igual forma, Gómez y Piattini (2018) indicaron que la gestión de datos centra su objetivo en la creación y el mantenimiento de un diseño de datos bajo estándares de calidad que se ajustan a los procedimientos reglamentarios de un gobierno de datos. Entre los atributos que integran la gestión de la DQ se encuentran:

- Gobierno de datos.
- Seguridad en los datos.
- Integridad e interoperabilidad de los datos.
- Almacenamiento de datos.
- Arquitectura de datos.
- Administración de base de datos.
- Calidad de datos.
- Administración de datos maestros.
- Delinear una “estructura de metadatos”.
- Administración de documentos, protocolos y contenidos.

1.2.4 Principales tendencias en calidad de los datos (DQ)

Las tendencias actuales en la DQ apuntan a la elevada demanda de las innovaciones tecnológicas que se dan en el contexto empresarial, lo cual conlleva a crear una gestión de datos engranada a la tecnología. Al respecto, de Soto y Cuervo (2006) señalaron que el uso de mecanismos flexibles es fundamental para que las empresas tengan un control seguro sobre sus datos y su análisis, así como establecer un gobierno sobre estos. En ese contexto también aparece la gestión de los metadatos como un complemento de las nuevas tendencias que están a la disposición de las organizaciones empresariales. Ello significa que

la administración de los metadatos proyecta un incremento eminente dentro de las empresas, debido a la necesidad de estas por tener un manejo eficiente de la información que cumpla con las normas, las reglas y los lineamientos que determina el gobierno de datos.

Otro elemento que cobra fuerza en este panorama es el análisis anticipado para optimizar la DQ, el cual sienta un precedente en las experiencias anteriores de la organización. La importancia de esta estrategia es que permite examinar modelos, prototipos y otras tendencias para establecer fallos en la calidad de los datos, proyectar futuros requerimientos y prever los posibles problemas que surgen de imprevisto.

Dentro de todas esas tendencias el gobierno de datos ocupa un lugar privilegiado porque su implementación, a nivel empresarial, permite reglamentar la administración de los datos, así como también organizar, crear políticas y asignar roles de seguridad, esto es: crear un escenario propicio para proteger los datos, garantizar su integridad y privacidad frente al usuario. Estas tendencias para gestionar los datos deben incluir la privacidad de estos, la cual debe prevalecer y estar al frente de las garantías que ofrece el gobierno de datos, puesto que sus políticas contemplan la seguridad y la privacidad como estandarte de todas sus operaciones. Sobre este particular ID Grup (2018) publicó un informe en su portal web (<https://idgrup.com/tendencias-datos-afectaran-las-empresas-2018/>) en el cual explicó las tendencias de la DQ que, según su criterio, marcaron un hito en el año 2018; además de mencionar la influencia que cada una de ellas puede tener en las organizaciones. En concordancia con los planteamientos de ID Grup, Soto (2006) señaló que entre esas tendencias se encuentran:

- Fusión de un conjunto de técnicas para gestionar los datos: consiste en el uso de herramientas tecnológicas para la ejecución de las tareas dentro de una

organización, siendo las de mayor relevancia y demanda la DQ que maneja la organización, el estudio de los datos, el gobierno de datos, la integración y la administración de los metadatos. La finalidad de esta tendencia es suministrarles a las organizaciones métodos, técnicas o nuevos modelos para gobernar los datos de una manera eficiente, confiable y segura.

- Gobierno de datos como estandarte de control: la gobernanza o el gobierno de datos le permite a la organización tener mayor control sobre los datos. Este consiste en la regularización que se establece sobre los datos, lo cual significa que al implantar un gobierno de datos en una empresa las perspectivas sobre el tipo de información serán transparente y ello generará beneficios tanto para la empresa como para el cliente, debido a que aumenta la confianza y el nivel de competitividad.
- Incremento del uso de Chief Data Officer (CDO) como estrategia innovadora para la gestión de datos de alto nivel: esta tendencia se perfila como una herramienta de control de datos con mayores alcances en la organización, pues engloba en una sola estrategia la tecnología, el negocio y la seguridad, con lo cual apoya la administración de los datos como “activo corporativo”. Entre sus funciones se hallan el aprovechamiento del dato y de la gobernanza de datos, es decir, define los procedimientos reglamentarios para resguardar y almacenar los datos y las políticas de seguridad, además de mantener a la organización a la vanguardia tecnológica en materia reguladora de datos.
- Establecimiento de reglamentos regulatorios para avalar la privacidad del dato: la creación de regulaciones fortalece los niveles de seguridad de la información. A través de ella se puede acceder a un gobierno de datos apropiado, así como

monitorear el cumplimiento de las normas para gestionar la calidad de datos y el uso de los datos.

- Expansión en la administración de metadatos: su implementación no solo le ayuda a la organización a usar los datos efectivamente, sino también a entender su esencia, valor e importancia. Así, su aporte a la empresa se cristaliza en el apoyo que las organizaciones le brindan al gobierno de datos, al cumplimiento de las políticas reguladoras y a los requerimientos propios de una gestión de datos adecuada.
- Asignarle un valor a los datos como activos de la organización: permite clasificar los datos según su importancia y jerarquía para asignarles un “valor numérico”, es decir, monetizarlos.
- Estudio predictivo como estrategia para optimizar la DQ: se fundamenta en el uso de metodologías automatizadas para analizar los datos, localizar irregularidades en la calidad y corregirlas.
- La fusión de tecnologías para gestionar la DQ: su principio es la administración de metadatos y la gestión de la privacidad de los datos, lo cual lleva a cabo mediante una articulación con la reglamentación que brinda el gobierno de datos para regular la DQ de los procesos del negocio.

1.2.5 Problemas comunes para garantizar la calidad de los datos

Autores como Arias (2015) han señalado que la gestión de la calidad de los datos se asocia con los riesgos a los que está expuesta la organización, lo cuales se relacionan con la administración de los datos, es decir: con las operaciones que realiza el usuario a través de ellos, la forma como asume el problema, y las estrategias que adoptada para enfrentar la situación. En este contexto, es importante el modo como la organización enfrenta cada

situación problema; establece directrices orientadas a la adquisición de un modelo de gestión de calidad de datos que sea eficiente, se centre en los procesos, les dé un tratamiento adecuado a los datos, y garantice un desempeño funcional y seguro tanto para la organización como para el cliente.

Igualmente, Arias mencionó que uno de los principales problemas que puede afrontar una organización es creer que cuenta con el monitoreo y el mantenimiento de los datos bajo estándares de calidad óptimos, y que estos no pueden ser duplicados o presentar una baja en su estándar de calidad. Pues, no solo esto puede ocurrir, sino que también es probable encontrarse frente a saturaciones de datos que obstaculizan la integración de los datos que recientemente ingresan a la base de datos ya diseñada.

Por su parte, Logicalis (s.f.) publicó en su portal web una explicación sobre el éxito o el fracaso de una organización, lo que se vincula con la gestión de la calidad de los datos, dado que no existe un sistema totalmente seguro, y el riesgo o las amenazas son latentes en este tipo de procedimiento. Lo anterior conlleva a las empresas a tomar medidas para responder ante cualquier eventualidad que vulnere la DQ y la información que se genera de ellos, aunque, en ocasiones, la organización tiende a enfocar su atención en los procesos de administración y deja de lado la calidad del dato que se gestiona.

Otro error que puede asociarse con la DQ es la falta de implementación de procedimientos automatizados para monitorear y detectar oportunamente cualquier debilidad en esta. Ello se debe a que muchas de estas organizaciones manejan el criterio, hasta cierto punto errado, de que una “supervisión y mantenimiento de estándares” eventual de la calidad de los procesos es suficiente para medir y determinar la calidad del dato. Igualmente, se pueden presentar casos en los que la base de datos ya estructurada colapsa por la cantidad de datos que contiene; frente a una situación de esta magnitud cualquier

técnica, modelo o mecanismo que se aplique para examinar la DQ sería inútil, y ello afectaría significativamente el nivel de competitividad de la empresa. En definitiva, este problema puede crear barreras funcionales en la base de datos que afectan la integración de los datos nuevos.

Finalmente, cabe anotar que la falta de un sistema de monitoreo constante de los datos recién integrados o ya existentes genera inconsistencia en la DQ. Puesto que un seguimiento continuo ayuda a depurar y a clasificar los datos de acuerdo con su calidad antes de ser gestionados, así como también contribuye a medir la relación, la oportunidad y la confiabilidad de estos.

1.3 Monitoreo de la calidad de los datos

Como parte de sus políticas internas, las empresas se ven en la necesidad de adoptar modelos o técnicas de monitoreo y medición de la calidad de los datos. Estas estrategias les permiten contar con un control oportuno y corregir cualquier falla que represente una amenaza inmediata o eventual. Sin embargo, no existe un parámetro único que ponga en evidencia la calidad de los datos; por el contrario, unos datos robustos o de baja calidad se manifiestan de diversas formas, lo cual involucra un conjunto de elementos que requieren ser tomados en cuenta como, por ejemplo, la seguridad de los datos maestros. En ese sentido, si no hay un sistema de vigilancia constante, es posible que se presente algún cambio y en caso de no ser percibido a tiempo y posteriormente corregido, lo más seguro es que se desencadenen problemas de mayor magnitud dentro de la organización.

En suma, la evaluación de la calidad de datos depende del monitoreo o vigilancia a seguir, no obstante, como se mencionó anteriormente, las herramientas dispuestas en las organizaciones no están sujetas a ningún tipo de actualización. Lo anterior, en tanto que la

mayoría de ellas continúan sumidas en viejos patrones que no garantizan la optimización de estos datos, y menos aún de la información que propician. Dicho de otro modo, el nivel de confiabilidad con la aplicación de métodos de monitoreo y medición tradicionales no es el esperado para la empresa, puesto que no hay una evaluación modular ni una vinculación de la calidad de datos a herramientas, métodos, técnicas o modelos de búsqueda automática de errores en el flujo de datos, conforme a normativas establecidas en la organización (Royo, 2005).

Al respecto, GS1 Argentina (2008) explicó que toda organización debería perfeccionar sus estrategias funcionales y para ello instaurar métodos modernos para monitorear los procedimientos de gestión de calidad de datos, lo cual permita tasar los resultados. Desde esta perspectiva, los “procesos y análisis de monitoreo” responden a la capacidad de la organización para dar cumplimiento a las metas, gobierno e indicadores claves. No obstante, la esencia de cada una de estas herramientas consiste en aprovechar al máximo el sistema de gestión de calidad de datos durante la fase de evaluación de fallas, tomar medidas preventivas y efectuar los correctivos pertinentes.

El monitoreo, seguimiento o vigilancia que se les aplica a los datos, conlleva a determinar las acciones a seguir y, sobre todo, a analizar si realmente es necesario implementar un modelo para medir la calidad de datos, los pasos y el enfoque óptimo. Para ello, el primer paso a tener en cuenta consiste en registrar o documentar cada procedimiento. Este mecanismo contribuye a la corrección de los problemas de calidad de datos, advertir el efecto sobre la organización y la duplicación del fallo. La medición establece el nivel de calidad de datos, pero adicionalmente pone de manifiesto los factores desencadenantes del problema, con el propósito de solventar la eventualidad. Asimismo, el proceso de medición de datos involucra los datos maestros y aquellos que ya han sido

publicados. Por último, se trae a colación las auditorías internas, por ser un procedimiento relevante para establecer debilidades o problemas de mayor magnitud en las organizaciones. En este escenario, el papel de la auditoría radica en demostrar que la calidad de datos manejados por la organización se ajusta a las reglamentaciones ya establecidas, cumple con los requerimientos de seguridad, privacidad, efectividad, eficacia, precisión y consistencia.

1.3.1 Metodologías para medir/monitorear la calidad de datos

1.3.1.1 *Total Data Quality Management (TDQM)*

Con respecto a la metodología TDQM, Caro, Fuentes y Soto (2013) explicaron que esta tiene como objeto suministrar datos de alta calidad a los usuarios. A tal efecto, proporciona el empleo de políticas idóneas para gestionar la calidad de datos integrales en una empresa, sobre todo nivel de gestión y de alta dirección. Para la aplicación de esta metodología, se parte de la definición del modelo de negocio, y posteriormente se procede a conformar el equipo de trabajo, para lo cual se evalúa la capacitación y conocimientos del personal en el manejo de herramientas, mecanismos y modelos para valorar la calidad de los datos. En cuanto a la consecución de las metas trazadas, la metodología TDQM plantea un proceso cíclico, a saber: la definición, la medición de la calidad, el análisis, y la optimización del proceso de gestión de los datos, tal como se presenta en la Ilustración 5:

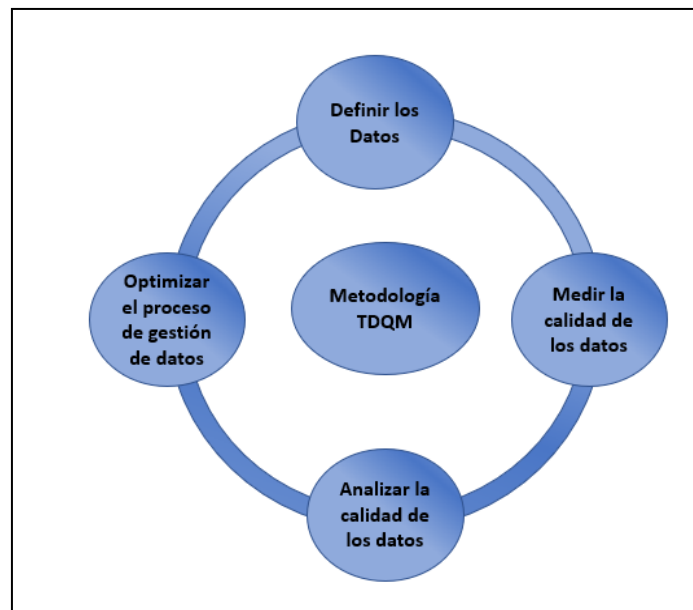


Ilustración 5. Esquematización de la metodología TDQM

Fuente: elaboración propia

Las fases de la metodología TDQM se componen, en primer lugar, por la definición de los datos. En esta primera fase se examina exhaustivamente los datos empleados por la empresa, para lo cual se elabora un prototipo de entidad de relación donde se precisa con exactitud el alcance y los requerimientos. Todo este procedimiento tiene como propósito determinar la relación entre quien provee la información, los usuarios, la administración y la incidencia de todos ellos en la creación de un modelo de gestión. En contraste, la medición de datos parte de la distinción y utilización de métricas que, de manera cuantitativa, sintetizan las eventualidades presentes en la calidad de datos, conforme a su nivel.

Por otra parte, el análisis de la calidad de datos es la fase o etapa con la mayor responsabilidad de todo este proceso cíclico. Esto último, puesto que es aquí donde se determinan las causas que producen el problema DQ, y se calcula el impacto que pueden generar los datos de baja calidad a la organización. Finalmente, la mejora establece los métodos, modelos, técnicas o herramientas útiles en la optimización de la calidad de datos.

Esta identifica los puntos críticos de cada departamento y organiza los movimientos de información que requieren atención prioritaria en el modelo de negocio.

En síntesis, TDQM es considerada como una metodología con un alto nivel de eficacia para la medición y monitoreo de calidad de datos. Sin embargo, para cumplir con las proyecciones y expectativas de su aplicación, se debe partir del compromiso asumido por la gerencia de frente a las políticas establecidas y manejadas por la organización. Adicional a ello, resulta conveniente que la organización admita que la información forma parte de sus activos, que está en constante movimiento y que ese mismo dinamismo puede limitar el uso y generar fallos en el sistema sino hay un control constante del flujo de información.

1.3.1.2 *Data Quality Assessment (DQA)*

La metodología DQA, según lo estableció la National Risk Management Research Laboratory Office of Research and Development (2016), centra su atención en evaluar, en primer lugar, la calidad de datos a través de un registro que se obtiene en el proceso de evaluación; y, en segundo lugar, estima la idoneidad de los sistemas usados por la organización para obtener datos, transferirlos, registrarlos y comunicar resultados. El proceso de evaluación con esta metodología, desde una perspectiva generalizada, es sencillo. En este caso, la evaluación de la calidad del dato se realiza mediante el arqueo y la comprobación de los datos relevantes en organizaciones seleccionadas, se analiza el recurso, la cualidad integral de los datos y la congruencia entre la documentación obtenida y los reportes; y finalmente, se evalúa cualitativamente la administración del sistema de gestión y el registro por niveles de los datos, como se muestra en la Ilustración 6:

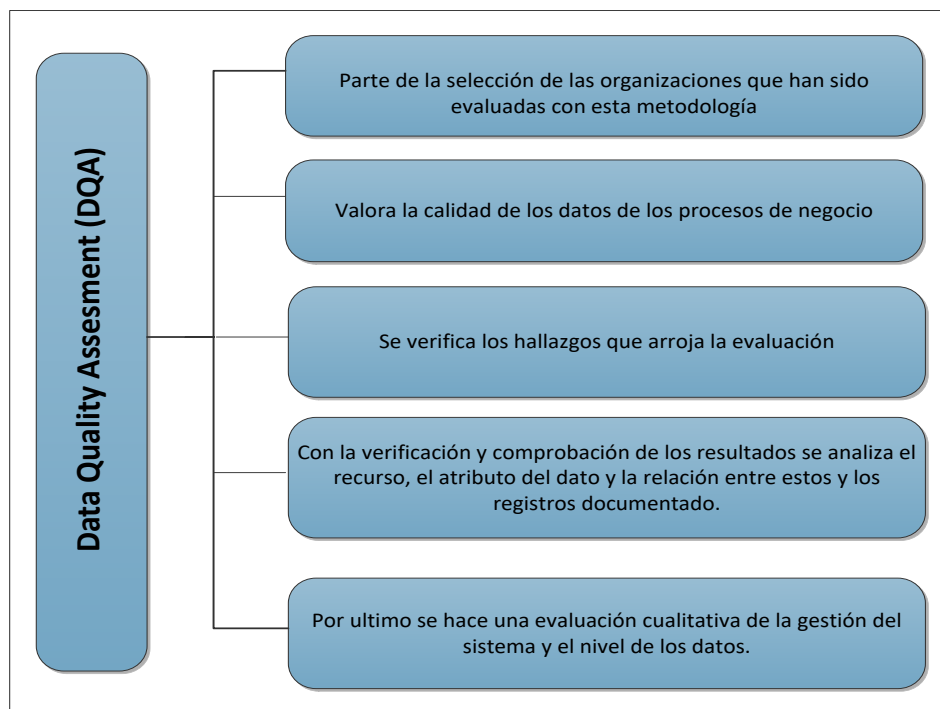


Ilustración 6. Esquemización de la metodología DQA

Fuente: elaboración propia

1.3.1.3 A Data Quality Assessment Framework (DQAF)

La esencia de este tipo de metodología es la Norma ISO/IEC 65:1996, y se cimienta en el Foro de Acreditación Internacional Asociado (IAF). Su función como método de medición y monitoreo descansa en establecer las capacidades de la organización para llevar a cabo una auditoría. Para dicho fin, la empresa debe demostrar que realizó previamente un estudio interno y examinó los potenciales de la calidad de datos respecto a otras organizaciones. Cabe mencionar que esta metodología es de fácil manejo y puede ser implementada en cualquier organización que cumpla con los requerimientos que exige como parte de su funcionalidad y efectividad para monitorear y medir la gestión de datos.

Las organizaciones, como instituciones funcionales, tienden a presentar en algún momento de su desempeño puntos críticos en el manejo y administración de la información,

lo que pone en duda la calidad de los datos. La Methodology for Information Quality Assessment (AIMQ), es una metodología integral que sirve de base en la valoración de la calidad de la información y una estimación desde una perspectiva valorativa que se organiza a través de un estándar de CI. Sobre esto, Lee et al. (2002) afirmaron que la metodología AIMQ se enfoca en satisfacer los requerimientos de quienes administran la información, así como de quienes hacen uso de esta, y dependiendo de cómo sea valorada la información por cada uno de ellos se utilizan cuatro “cuadrantes”.

El primer componente tiene como fundamento la receptividad de la información, esto es, la manera en que es percibida: como producto o servicio. Con base en esa determinación, se analiza si las mejoras efectuadas al sistema pueden valorarse a través de descripciones formales o desde la perspectiva del cliente. El segundo cuadrante o componente es la aplicación de un cuestionario de medición CI. Para ello, se inicia con el diseño del instrumento, para lo cual se establecen primeramente las variables que integran la calidad de datos y se extraen las dimensiones, pero tomando en cuenta el coeficiente intelectual de los administrados y usuarios en cada indicador que se emplee. El tercer elemento que constituye esta metodología son las técnicas de valoración e interpretación de la información. La primera de ellas captura la información que se obtuvo al aplicar el cuestionario de medición, la analiza y comprueba el “coeficiente intelectual” de una organización a través de puntos referenciales de otras organizaciones que han implementado mejores prácticas; y la segunda técnica mide el trayecto entre una evaluación y otra.

1.3.1.4 *Framework and Methodology for Data Quality Assessment (ORME-DQ)*

Para Batini et al. (2004) ORME-DQ es una metodología que busca un equilibrio asequible entre la integridad y la práctica para medir la factibilidad del proceso que administra la calidad de datos. A su vez, se caracteriza por poseer un conjunto de recursos para ejecutar el proceso de medición, y abarca la organización, procesos, el modelado del negocio, asociación de datos, flujos de datos, entre otros.

- Emplea métodos cualitativos y cuantitativos de manera simultánea para medir, evaluar y medir el riesgo. Cabe resaltar que la valoración cuantitativa está cimentada en dimensiones 4DQ, es decir, en la exactitud, integridad, vinculación y actualización.
- Estima las pérdidas económicas de la organización con base en los resultados del análisis de datos, dado que la presencia de datos de mala calidad influye en los costos.

El proceso de implementación comienza con categorización de los riesgos y se sistematizan en una matriz. Posteriormente se caracterizan los riesgos considerando las pérdidas financieras y los datos de bajo nivel, y se valoran las pérdidas obtenidas por la empresa. El paso siguiente consiste en medir los riesgos apoyado en métricas cualitativas y cuantitativas. Por último, se hace el monitoreo de riesgos mediante umbrales bien definidos (Batini et al., 2007). Dentro de los métodos cuantitativos, se tiene en cuenta las dimensiones de calidad de datos para su medición. De acuerdo con Batini et al. (2004), en el módulo de evaluación de la calidad de los datos, se aplican técnicas y algoritmos de calidad en los datos recuperados de las sondas para medir las dimensiones de la calidad de estos. La versión prototipo del marco admite las métricas que se muestran en la Tabla 2:

Tabla 2. Resumen dimensiones de calidad de datos

Fuente: elaboración propia

Dimensión	Descripción	Clases	Descripción
Precisión	En las dimensiones de calidad de datos es la más ampliamente usada (Huang et al, 1998). Es la proximidad entre un valor de datos y el valor de un objeto del mundo real (Batini y Scannapieco, 2006).	Sintáctica Semántica	Por ejemplo, el valor contenido en un atributo nombre “Dego”, es un error sintácticamente incorrecto, pues no corresponde al nombre real “Diego”. Se identifica con los valores discretos <si, no> o <correcto, incorrecto>. La técnica para identificar precisión semántica consiste en buscar los mismos datos en diferentes fuentes y encontrar la información correcta por comparación (también requiere de la estrategia de identificación de objetos para conocer si dos tuplas representan a la misma entidad del mundo real o no).
Complejidad	Para la dimensión de Complejidad, la métrica se basa en el supuesto de un mundo cerrado y el modelo relacional con valores <i>null</i> . Como consecuencia, la métrica calcula la proporción de valores nulos sobre los valores de la tabla.	Sin nulos con OWA Con nulos y CWA	Closed World Assumption (CWA): establece que solo los valores presentes en una tabla relacional r, y no otros, representan los hechos del mundo real. Open World Assumptions (OWA): no se puede establecer ni la verdad ni la falsedad de los hechos que no están representados en las tuplas de r.
Relacionadas con el tiempo	Esta dimensión nos identifica el cambio y la actualización en el tiempo.	Rapidez Volatilidad Oportunidad	Los datos son actualizados en el tiempo. Ejemplo, si el correo electrónico de una persona es actualizado, y este corresponde al correo electrónico personal, entonces la rapidez es alta. Es la frecuencia en la que los datos varían en el tiempo. Por ejemplo, las fechas de nacimiento tienen cero (0) grado de volatilidad; mientras que el nivel de inventario de una compañía posee un alto grado. Los datos siempre están disponibles en cualquier momento para cumplir con una operación puntual.
Consistencia	La consistencia de un	Restricciones de Integridad	Restricciones de interrelación: una restricción de dominio definida sobre un esquema es “edad está entre 0 y 120”. Se puede calcular identificando los atributos que cumplan la restricción o no.

<p>conjunto de datos podría comprobarse con un conjunto de normas o reglas; un conjunto de otros datos de una base de datos; un conjunto de otros datos de otros sistemas, y otros datos de una instancia diferente del mismo proceso.</p>	<p>Data Edits</p>	<p>Restricciones de interrelación: involucra atributos de más de una relación. Una restricción de interrelación dice: “el año de la relación película debe ser igual al atributo año en la tabla Oscar Awards”. Se mide identificando que relaciones cumplen con la condición o no. Se aplican cuando el modelo no es relacional. La regla para definir este tipo de inconsistencias puede ser la siguiente: si el estado civil indica que está casado, la edad no debe ser menor de 18 años. Se mide identificando los conceptos que no cumplen con la condición o no.</p>
--	-------------------	--

1.3.1.5 *Data Warehouse Quality Methodology (DWQ)*

Los autores Jarke y Vassiliou (1997) expusieron que DWQ es una metodología innovadora que nutre el estudio y análisis de la DB de metadatos, basados en estándares formales de calidad de datos y, por ende, de información que contribuye favorablemente con el mejoramiento de adaptación del diseño y cuantificación de los datos en reserva. Por otro lado, aporta recursos técnicos útiles en la difusión de cambios, para brindar soluciones ante la presencia de conflictos. En últimas, la metodología facilita al usuario mejorar la calidad de datos, optimizar el proceso, el modelo de negocio e integrar estrategias que sirvan de guía para la evaluación.

Por su parte, Gonzáles, (2010) refirió que la metodología parte del entendimiento de los datos. Una vez que se comprende la esencia de estos, se procede a optimizar su calidad y, finalmente, los combina por medio de diferentes fuentes para aumentar el valor de los datos como activo de la organización. Asimismo, según lo mencionó el autor, DWQ admite la evaluación de la calidad de los datos en tiempo real. El procedimiento a seguir es el siguiente:

- Selecciona las fuentes de datos internas y externas (provenientes de datos maestros y datos obtenidos).
- Cimentación de los datos: integra los datos, filtros, normalización, herramientas para el manejo de datos.
- Fases de almacenamiento: determina la correspondencia de los datos a través del uso de metadatos.
- Fases de acceso: genera conocimiento partiendo del almacenamiento para garantizar la conexión entre todos los dispositivos que integran la arquitectura del almacén de datos.
- Fase de explotación: pone en evidencia todo el proceso, mediante la generación de un nuevo conocimiento al servicio de la organización.

1.3.2 Modelo de procesos para la evaluación en calidad de datos de las organizaciones

Las metodologías y marcos de trabajo para la evaluación en la DQ se enfocan en contextos genéricos que no proporcionan soluciones a problemas específicos y, por tanto, carecen de practicidad. De tal manera que las organizaciones, al adoptar estas metodologías, se dan cuenta de que no se adaptan a sus necesidades y expectativas. Lo anterior, debido a que se usan conjuntos fijos ya predefinidos para sus dimensiones y mediciones; mientras que, en un marco de evaluación para un contexto específico, se requiere de la definición de sus dimensiones, medidas y un proceso de evaluación en calidad de datos.

El estudio realizado por Wister (2019), aborda la brecha existente entre los métodos genéricos (es decir, independiente del contexto) y un contexto específico, por lo que el

objetivo del estudio se centra en aumentar la capacidad de conocimiento sobre la forma de aplicar las actividades críticas de evolución en calidad de datos en un contexto determinado, para así obtener una evolución completa de su calidad, garantizando mayores capacidades a los profesionales en calidad de datos. Resulta preciso mencionar que este objetivo se logra mediante el diseño de un modelo de proceso genérico, pero con un alto grado de practicidad para la evaluación de calidad de datos.

En tal sentido, para que el modelo del proceso sea práctico, se requiere de roles de los participantes y las actividades en las cuales se encuentran involucrados. Adicionalmente, para que el proceso que se muestra en la Ilustración 7 sea funcional deben identificarse todas las actividades críticas en la evolución de calidad de datos, por lo que hay dos preguntas que deben responderse:

- ¿Cuáles son las actividades críticas de un proceso genérico de evaluación de la calidad de los datos?
- ¿Qué funciones deben asignarse a esas actividades para llevar a cabo eficazmente el proceso de evaluación de la calidad de los datos?

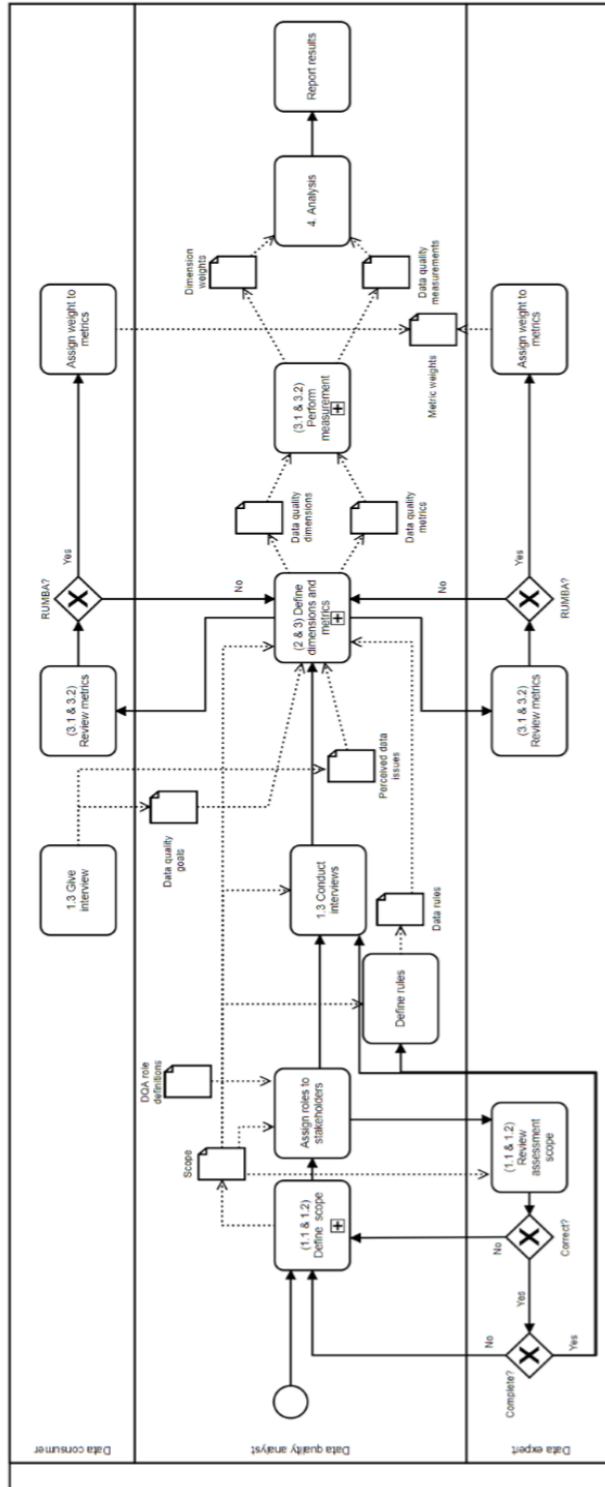


Ilustración 7. Process model for data quality assessment

Fuente: (Wister, 2019)

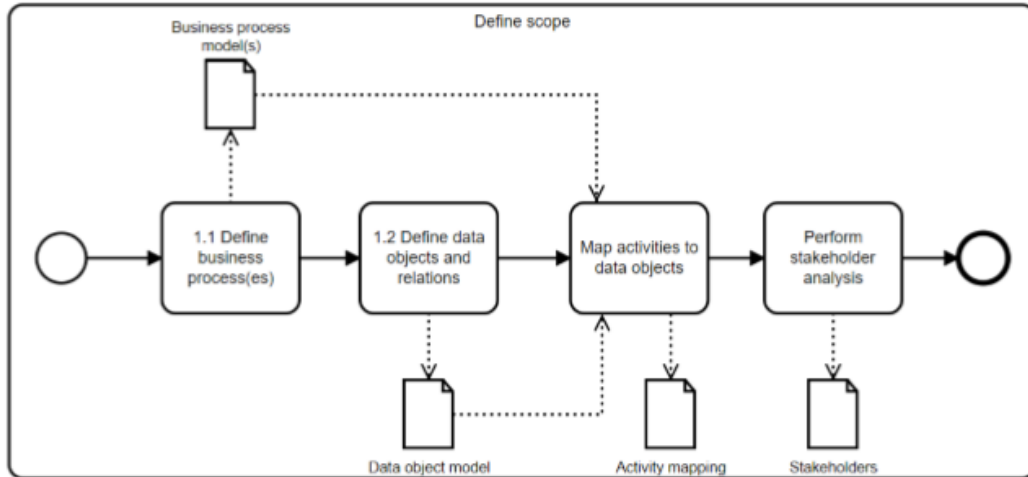


Ilustración 8. Subprocess Define Scope

Fuente: (Wister, 2019)

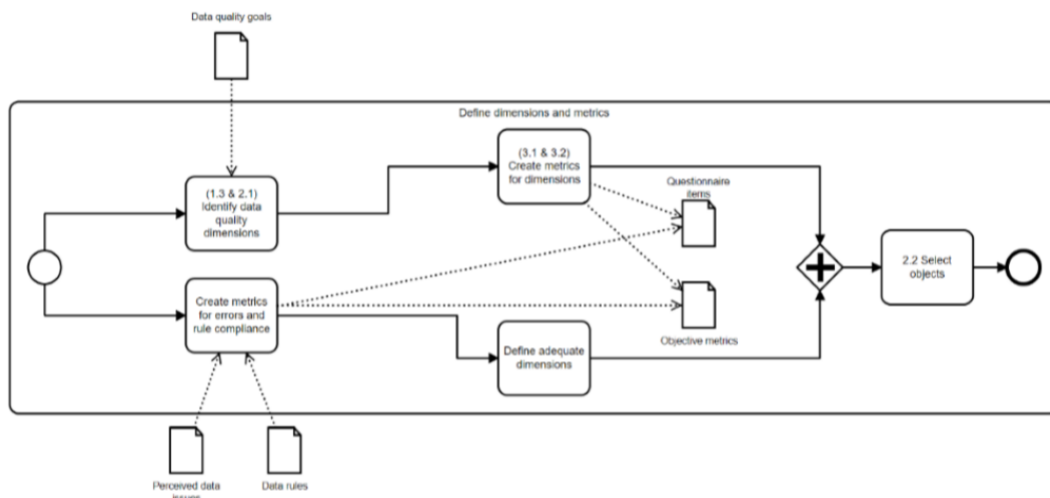


Ilustración 9. Subprocess Define dimensions and metrics

Fuente: (Wister, 2019)

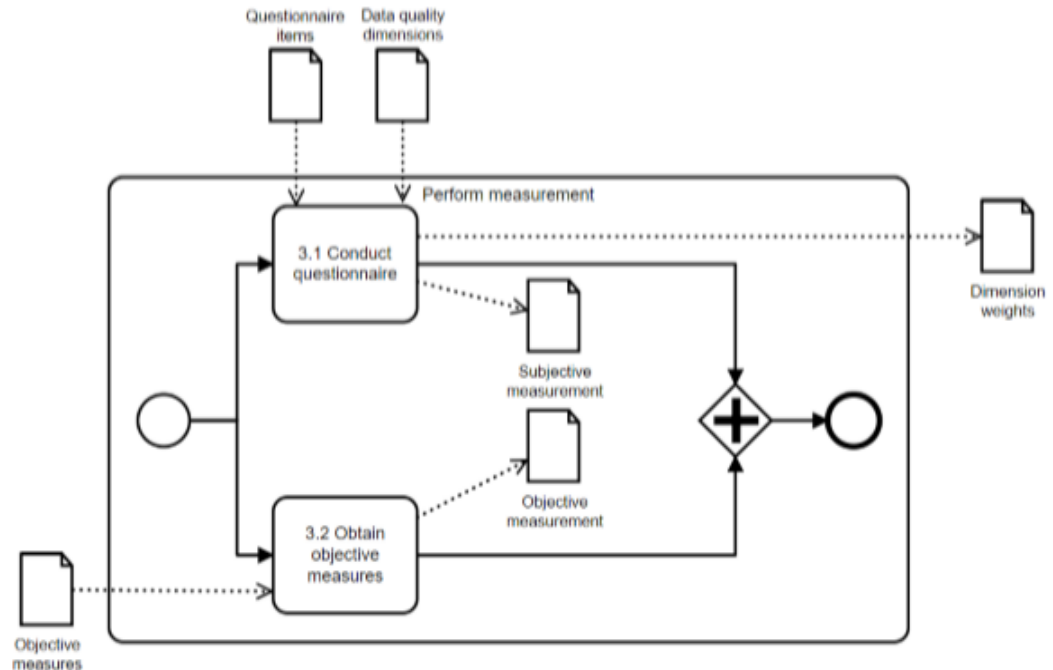


Ilustración 10. Subprocess Perform measurement

Fuente: (Wister, 2019)

En concordancia con Wister (2019), el diseño propuesto se muestra en las ilustraciones 8, 9 y 10. Además, este se explica de la siguiente manera:

1. Definición del alcance: el primer paso descansa en delimitar el contexto, esto es, los procesos de negocio, los datos y sus relaciones, los objetivos y los requisitos de los usuarios o clientes de datos. Los procesos empresariales y los datos a evaluar forman el ámbito de evaluación, uno de los subprocessos es la definición del alcance (ver Ilustración 9), y esta tiene las siguientes actividades:

- Concretar los procesos de negocio implica proporcionar un modelo Business Process Model and Notation (BPMN) del proceso de negocio relacionado con los datos a evaluar.

- La definición de los datos y sus relaciones implica la construcción de un diagrama de clases de los objetos de datos, con sus relaciones, atributos y las interrelaciones entre ellos.
 - Mapeo de los procesos empresariales a los datos con el fin de obtener una definición clara y una mejor comprensión del contexto. Para ello, se añade un mapeo de los procesos empresariales a los objetos de datos, describiendo cómo los procesos consumen, crean o modifican los datos.
2. Definir las dimensiones y métricas: se parte de la premisa de que un modelo de proceso debe incluir tanto un enfoque ascendente (impulsado por el problema) como descendente (impulsado por la necesidad). Los problemas experimentados se identifican mediante entrevistas semiestructuradas con los clientes/usuarios de datos. Estas entrevistas también se utilizan para identificar los objetivos de los consumidores de datos (para incluir un enfoque de evaluación de arriba a abajo). Asimismo, combinando los resultados de los enfoques descendente y ascendente, se puede crear el conjunto completo de dimensiones y métricas. Sobre la base de este conjunto, se pueden seleccionar los objetos de datos (sistemas de información, tablas, atributos, historia, etc.), sobre los que se realizan las mediciones. Los criterios para la métrica son definidos de acuerdo con el acrónimo RUMBA (en inglés); la métrica debe ser razonable, comprensible, mensurable, creíble y alcanzable. Además, los clientes/usuarios de datos y los expertos en datos, asignan pesos a las métricas en función de su opinión sobre la medida en que la métrica representa las dimensiones previstas.

3. Realizar la medición: como las mediciones se realizan de forma subjetiva y objetiva, este subproceso contiene dos caminos paralelos. La medición subjetiva implica la realización de un cuestionario, cuyos elementos se crean durante el desarrollo de la métrica (en el subproceso anterior). Paralelamente a la realización de un cuestionario, se pueden realizar las métricas objetivas (es decir, calculadas sobre los objetos, tablas, atributos e historial de datos seleccionados). Este subproceso da lugar a una medición subjetiva en forma de respuestas a los elementos del cuestionario, y a una medición objetiva en forma de fórmulas calculadas.
4. Análisis y presentación de informes: por último, se combinan los resultados del cuestionario y las mediciones objetivas. Esto, utilizando los pesos métricos, se puede obtener una puntuación final para cada dimensión, y empleando el peso de las dimensiones, es posible obtener una puntuación final de calidad de los datos globales. La presentación de informes incluye la creación de un informe sobre la calidad de los datos (en el que se describen los resultados y una descripción del proceso) y su distribución a los interesados.

Para desarrollar el modelo de proceso propuesto, fue necesario identificar las actividades y funciones críticas de la evaluación de la calidad de los datos. Estas preguntas de investigación se respondieron realizando un examen de la literatura, analizando las metodologías y los documentos encontrados y, posteriormente, sintetizando las metodologías basadas en este análisis. Esta síntesis condujo a la identificación de actividades y funciones críticas para la evaluación de la calidad de los datos. Sin embargo, esta síntesis es bastante subjetiva y en términos de reproducibilidad, podría dar resultados algo diferentes si se hiciera la misma síntesis en una investigación diferente.

Para que el modelo de proceso fuese práctico, se requirió de una interpretación más práctica de esas actividades. Esta traducción de la abstracción a la práctica también depende de la interpretación subjetiva del investigador, lo que quiere decir que hay más formas de traducir la abstracción a una práctica). Por ejemplo, a la actividad crítica “definir datos y relaciones” se le dio una interpretación práctica mediante la creación de un modelo de objeto de datos utilizando un diagrama de clases UML (Unified Modeling Language). Sin embargo, hay más formas de definir los datos y sus relaciones. Las variaciones del modelo de proceso que contienen diferentes interpretaciones prácticas son un tema interesante para seguir mejorando el modelo en futuras investigaciones.

Por otra parte, el modelo de proceso se evaluó utilizando un conjunto de objetivos de solución definidos al principio del estudio. Estos objetivos de solución se basan en el problema que el artefacto pretende resolver, y se obtuvieron mediante un razonamiento directo del investigador. No obstante, es posible que haya más objetivos de solución que sean interesantes para la evaluación de esta investigación, por ejemplo, conjuntos predefinidos de objetivos de solución para el diseño de artefactos de sistemas de información en la literatura (Wister, 2019).

1.3.3 Herramientas tecnológicas para calidad de datos

Existe una variedad de aplicaciones de software dedicadas a la calidad de datos y enfocadas al preprocesamiento, perfilamiento y limpieza de estos, con la finalidad de garantizar la integración o análisis de información. Los autores Ehrlinger, Rusz y Wößl (2019), en su estudio identificaron 667 herramientas, de las cuales seleccionaron 17 de calidad de datos para una evaluación detallada. Tres de ellas se basan en SAP (SAP Information Steward, solución DQ de ISO Professional Services y dspCompose de

BackOffice Associates GmbH) y como no accedieron a una instalación de SAP, no incluyeron esas herramientas en la encuesta. De esa manera, fueron evaluadas 15 con respecto a tres áreas de funcionalidad: (1) perfil de datos, (2) medición de calidad de datos en términos de métricas y (3) monitoreo continuo de la calidad de los datos, teniendo en cuenta que para este estudio era importante la independencia del dominio y que ofrecieran versiones libres o de pruebas. Las herramientas evaluadas fueron:

1. Aggregate Profiler.
2. Apache Griffin.
3. Ataccama ONE.
4. DataCleaner by Human Inference.
5. Datamartist by nModal Solutions Inc.
6. Experian Pandora.
7. Informática Data Quality.
8. IBM InfoSphere Information Server for Data Quality.
9. InfoZoom by HumanIT Software GmbH.
10. MobyDQ.
11. OpenRefine & MetricDoc.
12. Oracle Enterprise Data Quality.
13. Talend Open Studio for Data Quality.
14. SAS Data Quality.
15. Data Quality Solutions dedicated to SAP.

La investigación se basó en la aplicación de unas encuestas y uno de sus principales objetivos fue la característica en el área funcional de monitoreo continuo en calidad de datos. Dentro de la encuesta de investigación en herramientas de calidad de datos los

requisitos evaluados puntualmente fueron enfocados en el área funcional (3) Monitoreo Continuo en Calidad de Datos (CDQM):

- (39) Programación de una tarea de métrica de calidad de datos o perfilamiento de datos en períodos definidos por el usuario.
- (40) Almacenamiento de las mediciones de calidad de datos y resultados de perfilamiento de datos.
- (41) Recuperación de las mediciones de calidad de datos o resultados de perfilamiento de datos.
- (42) Comparación entre varias mediciones de calidad de y resultados de perfilamiento de datos.
- (43) Visualización de las mediciones de calidad de datos y resultados de perfilamiento de datos a lo largo del tiempo.

Para cada requisito se calificó de la siguiente manera: (✓) el requisito es cumplido, (-) el requisito no es cumplido, o (p) el requisito es parcialmente cumplido (ver Ilustración 11). La cobertura de cada requisito se describe en forma textual con un enfoque en el que la justificación de cumplimientos es parcial.

Cabe anotar que para las dos herramientas DQ (Ataccama ONE y Talend OS) se cuenta con la disponibilidad de otras versiones más avanzadas que soportan el CDQM según sus proveedores, pero estas no fueron investigadas.

		Aggregate Profiler	Apache Griffin	Ataccama ONE	DataCleaner	Datamartist	Experian Pandora	Informatica DQ	InfoZoom & IZDQ	MobyDQ	OpenRefine & MetricDoc	Oracle EDQ	SAS Data Quality	Talend Open Studio
39	Task scheduling	p	✓	—	✓	p	✓	✓	p	✓	—	✓	p	—
40	Storage of results	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
41	Retrieval of results	—	✓	✓	✓	—	—	✓	✓	—	—	✓	✓	✓
42	Comparison	—	✓	—	✓	—	✓	✓	✓	—	—	—	p	—
43	Visualization over time	—	✓	—	✓	—	✓	✓	✓	—	—	—	p	—

Ilustración 11. Capacidades de monitoreo de la calidad de los datos

Fuente: (Wister, 2019)

Las soluciones empresariales más completas para CDQM en los requisitos 40 y 41 son proporcionadas por Informática Data Quality (IDQ) y SAS Data Quality, que permiten la exportación de perfilamiento de datos. Durante la importación, todos los ajustes y fuentes de datos necesarios se recargan desde el momento del análisis. Por otro lado, la programación de tareas del requisito 39 (CDQM) también es ampliamente apoyada por las herramientas. Aggregate Profiler cumple este requisito solo parcialmente, en tanto que solo es posible programar las reglas de negocio y ninguna otra como por ejemplo las tareas de perfilamiento datos. En Datamartist, InfoZoom y SAS Data Quality la programación de tareas es un poco más compleja para los usuarios, dado que se requiere escribir por línea de comandos un archivo de lote para su programación. En el caso del Datamartist, la herramienta debe cerrarse para permitir la ejecución del archivo de lote programado.

Para visualizar las comprobaciones realizadas a la calidad de datos de forma continua (ya sean tareas de perfilamiento de datos, definición de reglas de usuario o métricas de calidad de dato), IDQ se basa en las llamadas “tarjetas de puntuación” que pueden personalizarse para mostrar la información correspondiente. Por su parte, Apache

Griffin, Experian Pandora y SAS Data Quality posibilitan que las alertas sean definidas cuando se produzcan errores específicos o cuando se infrinja una regla establecida. MobyDQ no ofrece ninguna visualización (lo que se considera una mejora a futuro de esta herramienta), pero se basa en bibliotecas externas en su implementación como Ubisoft. SAS cumple parcialmente los requisitos 42 y 43 ya que sus tableros contienen únicamente el número o porcentaje de disparadores por fecha, fuente o usuario, pero no se pudieron trazar los valores concretos (por ejemplo, 80 % de integridad). Las soluciones más completas para el área funcional (3) CDQM, de uso general, son IDQ y DataCleaner by Human Inference. Con respecto a las herramientas de código abierto, solo Apache Griffin proporciona un soporte completo para CDQM y la versión comercial de MobyDQ la cual se encuentra desplegada en Ubisoft.



Ilustración 12. Cuadrante Mágico de Gartner 2019 sobre herramientas de calidad de datos

Fuente: (Informática, s.f.)

De acuerdo con la investigación, en la selección de herramientas en calidad de datos se proporcionó un buen resumen del mercado, puesto que fueron encontradas en el estudio nueve herramientas comerciales y de código cerrado, así como cinco herramientas gratuitas y de código abierto mencionadas por Gartner (ver Ilustración 12). Los proveedores de cuatro herramientas han sido nombrados líderes en el Cuadrante Mágico de Herramientas de Calidad de Datos 2019 (Informática, SAS, Talend, Oracle) y dos de ellos se encuentran entre los cuatro proveedores que controlan el mercado (que son SAP, Informática, Experian y Syncsort; sin embargo, no se realizó ninguna prueba para SAP ni Syncsort). En general, y

en concordancia con los requisitos experimentados con IDQ, tiene un nivel de madures alto como herramienta de calidad de datos.

El mejor soporte para perfilamiento de datos lo proporciona Experian Pandora, que permite el perfilamiento a través de toda una BD e incluso a través de múltiples fuentes de datos conectadas. Todas las demás herramientas dan lugar al perfilamiento de datos solo para columnas seleccionadas o dentro de las tablas de específicas. Adicional a ello, percibimos que Oracle EDQ, Talend OS y SAS Data Quality tienen menos soporte para perfilamiento de datos y/o monitoreo de calidad de datos. Aunque en el cuadrante (DataCleaner) fue eliminado del estudio de Gartner en 2019 debido a su enfoque en los datos de los clientes, la evaluación arrojó un buen soporte en perfilamiento de datos y un fuerte apoyo en el monitoreo de calidad de datos.

Sin embargo, al comparar las herramientas de propósito general y de libre acceso de calidad de datos con las herramientas Talend OS y Aggregate Profiler, el antiguo interfaz de usuario de confianza es muy intuitivo y tiene una buen de ayuda. Aggregate Profiler, por otra parte, tiene un soporte más rico y avanzado para el perfilamiento multicolumna y limpieza de datos; pero no siempre está claro qué algoritmos se utilizan para realizar las modificaciones datos y la documentación no está actualizada. En consecuencia, se instalaron tres herramientas de código abierto (Apache Griffin, MobyDQ y OpenRefine) desde GitHub, por lo que se requirieron conocimientos técnicos para su configuración. MobyDQ y Apache Griffin tienen claramente un enfoque diferente en el CDQM. A su vez, IBM ISDQ demostró que las herramientas comerciales pueden ser muy complejas y de alto tiempo en su instalación debido a la creciente complejidad de los módulos individuales y las dependencias entre ellos.

Tras haber analizado diferentes herramientas, el presente estudio tuvo en cuenta la herramienta IDQ, puesto que en la evaluación realizada (Ilustración 11: Capacidades de monitoreo en calidad de datos) y según el cuadro mágico de Gartner (Ilustración 12: Herramientas de calidad de datos), en el cuadrante de líderes la mejor herramienta es Informática DQ y la más relevante.

1.3.3.1 *Informática DQ*

Dentro de los productos ofrecidos por IDQ, el módulo dedicado a la calidad de datos, en el cual se despliegan todas sus capacidades como la funcionalidad de monitoreo, Informática (2020) dice que uno de los beneficios del monitoreo de datos es el ahorro de la organización en tiempo y dinero, pues de no tenerlo se gastaría en comprobar y transformar los datos. El monitoreo de los datos promueve la agilidad comercial, por lo que las nuevas iniciativas que implican datos pueden emprenderse inmediatamente sin necesidad de una fase de preparación de datos que requieren mucho tiempo. El objetivo del producto IDQ es maximizar su la calidad de los datos. La herramienta en monitoreo de datos trae un módulo de administración con el cual se pueden programar las diferentes tareas de monitoreo. Esta se basa en la configuración inicial de perfilamiento de datos y reglas de negocio del dominio, y mediante unos procesos desatendidos se recolectan los datos monitoreados que son almacenados para ser visualizados en un tablero de control por los usuarios finales; igualmente, generan alertas a los responsables de los datos y administradores del monitoreo.

1.3.3.2 Conclusiones de herramientas tecnológicas para calidad de datos

- La herramienta líder IDQ carece de un componente de automatización de procesos robotizados para garantizar el manejo con mayor eficiencia y alto grado de integración con otras tecnologías. Además, no cuenta con un cubrimiento en diferentes procesos de negocio con diferentes reglas de mediciones y criterios de aceptación por los negocios a monitorear.
- Los resultados del estudio sobre las herramientas de medición y monitoreo de la calidad de los datos revelaron características interesantes de las herramientas de la calidad de los datos. Por tanto, se infiere que es de suma importancia que las herramientas se fortalezcan en la medición automatizada y el CDQM.
- De acuerdo con los clientes, se necesita un perfilamiento más automatizado inicial que abarque todas las tablas, fuentes de datos, y por tal razón generen alertas en tiempo dinámico.
- Al mismo tiempo, es necesario cerrar la brecha entre las medidas de calidad de datos prefabricadas y el mejoramiento de los cálculos y algoritmos realizados. El estudio identificó que vale la pena investigar el potencial de los datos automatizados listos para ser perfilados junto con una clara definición de los parámetros utilizados que podrían ser modificados después de la ejecución inicial.

En síntesis, lo más relevante para el proyecto, y teniendo en cuenta las conclusiones anteriormente planteadas, la automatización del monitoreo y generación de alertas es de gran relevancia. Para ello, se debe plantear una propuesta genérica que contemple cualquier proceso de negocio de forma dinámica y unas mediciones flexibles a través del tiempo.

1.4 Automatización de procesos RPA

1.4.1 ¿Qué es RPA?

La Automatización Robótica de Procesos (RPA) es una tecnología de software usada para disminuir la intervención humana en los procesos de negocios empresariales, los cuales son repetitivos en los sistemas de información. En ese orden de ideas, “La automatización robótica de procesos (RPA) es una tecnología que permite configurar “Robots Software” para la automatización de tareas manuales o semiautomáticas de manera integrada con los sistemas actuales de la organización” (Ernst y Young, 2018, p. 3). El RPA es una herramienta que garantiza la optimización de los procesos, pues se adapta a los requerimientos de la organización, genera alarmas, facilita el monitoreo y detecta las fallas en tiempo real.

¿Para qué sirve?

La tecnología RPA sirve como un apoyo a las personas, puesto que las libera de grandes cargas de trabajo repetitivo. Esto conlleva a que se optimicen las tareas de creatividad y oportunidad en el negocio por parte de los colaboradores. La automatización robótica de procesos se considera como automatización suave, en tanto que no es un fenómeno físico; es un software de carácter informático programado que puede ejecutar procesos computacionales en lugar de humanos (Isaac et al., 2017). Unos ejemplos de RPA son los *chatbots* que se encuentran en la parte frontal, este hace preguntas más definitorias y luego ejecuta la acción al instante.

La automatización se lleva a cabo en la capa de aplicación y base de datos: ejecuta tareas masivas o grandes transacciones para las necesidades de nivel de la empresa durante todo el día (UiPath, 2015). Dentro de estos está también el proceso de pedidos que puede

involucrar muchas tareas manuales que absorben mucho tiempo, como comprobar direcciones, introducir datos, imprimir la factura, la etiqueta de envío, actualizar el inventario del almacén y volver a hacer un pedido si sus existencias son bajas. A su vez, RPA puede automatizar y coordinar este proceso. Al ser un robot de software y no un robot físico, la única cosa que no podrá hacer por usted es coger y empaquetar cajas (KOFAX, 2020). Por último, un proceso que consume mucho tiempo es la generación de informes financieros, en especial para finanzas y contabilidad, sobre todo a final de mes, de trimestre o de año, cuando todas las personas disponibles están recopilando información frenéticamente. Finalmente, RPA puede automatizar la recopilación y agregación de datos en mucho menos tiempo, liberando ancho de banda y cerebros para analizar esta información en busca de datos clave (KOFAX, 2020). En la Ilustración 13 se pueden observar las soluciones de la automatización de los procesos:



Ilustración 13. Soluciones de automatización de procesos

Fuente: (Deloitte, 2017)

¿Cuándo se recomienda?

Primero que todo tenemos que tener claro en qué casos se puede aplicar un RPA y en qué casos no. RPA se puede usar para automatizar tareas concretas, repetitivas, sencillas, y hoy en día con las tecnologías cognitivas, le da un valor agregado para toma de decisiones en ciertas tareas de conocimiento humano y que exige la toma de decisiones. Es ideal aplicar RPA a temas como procesos de negocio o flujos de trabajo que presentan un alto nivel repetitivo y sus pasos obedecen a un patrón de operación común.

No obstante, existe una tierra de nadie. Cuando se integra con otro software de automatización para mejorar sus capacidades centrales, RPA puede utilizarse en más situaciones y convertirse así en un valioso componente de una estrategia de automatización que incluya captura de datos, reglas de negocio y flujo de trabajos.

(IBM, 2018, p. 3)

En general, hoy en día RPA se está convirtiendo en la mano derecha tanto de las áreas de IT, como para los usuarios finales quienes son los directamente beneficiados; además de esto, les brinda la mano a las corporaciones reduciendo tiempos y costos en operación. En los procesos que son de misión crítica, donde el recurso humano es requerido y su valor en conocimiento aporta demasiado a la corporación, en estos procesos es cuando el RPA libera de carga este recurso permitiéndole su crecimiento como individuo y para la organización.

Beneficios

Los beneficios que brinda la RPA son los siguientes:

Rendimiento

- “RPA tiene el potencial de impactar 40% de las funciones de negocio. Gartner estima que para las funciones de TI puede ser de 60%” (KPMG, 2017, p.1).
- “Robots de software trabajan 24/7, no toman vacaciones y ejecutan tareas a mejor velocidad” (KPMG, 2017, p.1).

Costos

- “Las herramientas de automatización robótica son hasta 65 % menos caras que el tiempo completo de trabajo de los empleados (Capgemini, 2016, p. 8).

Satisfacción

- “Empoderar a los empleados para realizar más tareas de valor agregado” (Capgemini, 2016, p. 6).

Auditoría

- Es auditable todas las tareas y su trazabilidad que realiza una herramienta RPA.

Implementación

- “Estandariza y optimiza procesos, mejorando la calidad y el costo de entrega” (Deloitte, 2017, p. 12).
- “Ciclos de desarrollo rápido, proporcionando ROI en 3-6 meses” (Capgemini, 2016, p. 6).

Para el abordaje de este proceso de automatización, se debe tener en cuenta su clasificación, a saber: “automatización básica de procesos robóticos, automatización de

procesos avanzada y automatización cognitiva” (KPMG, 2017, p.1) las cuales se explican a continuación:

- Automatización básica de procesos robóticos: clase 1, “fácilmente diseñadas y rápidamente probadas, y necesitan poca inversión antes de ser utilizadas. Esta tecnología únicamente puede usar datos estructurados” (KPMG, 2017, p. 1).
- Automatización de procesos avanzada: clase 2, “aprovecha tecnologías más avanzadas que incorporan elementos de autoaprendizaje para abordar la automatización de procesos menos estructurados y, por lo general, más especializados” (KPMG, 2017, p. 1).
- Automatización cognitiva: clase 3, “incorpora capacidades avanzadas de autoaprendizaje en tecnologías tales como procesamiento de lenguaje natural, inteligencia artificial, aprendizaje automático y análisis de datos” (KPMG, 2017, p. 1).

En la Ilustración 13 se muestra el grado de avulsión de los modelos de procesos de calidad como lo son Kaisen, Six Sigma, Outsourcing, Offshoring, entre otras; y las tecnologías emergentes como Deep Learning, Asesores Inteligentes, Informática Afectiva, Ciudadanos de Ciencia de Datos, entre otros. Estas están soportadas con procesos de automatización robotizados RPA. Para analizar el contexto de implementación de RPA, Deloitte (2017) propuso identificar la tipología del proceso a intervenir:

- Procesos específicos: “Son aquellos procesos transaccionales que forman parte de una función más grande dentro de la empresa, los cuales son sencillos y repetitivos” (Deloitte, 2017, p. 7).

- Procesos multifuncionales: “Son procesos similares que se ejecutan a través de múltiples funciones en una organización, por ejemplo, el realizar conciliaciones bancarias y conciliaciones de facturas, requeridas para el cierre mensual/anual en una empresa” (Deloitte, 2017, p. 7).
- Proceso punta a punta: “procesos completos que se llevan a cabo a través de múltiples áreas, como lo es el proceso de Compra y Pago” (Deloitte, 2017, p. 7).

Existen unos requerimientos mínimos que debe cumplir el análisis del proceso a ser automatizado mediante RPA, los cuales son:

- Proceso repetitivo.
- Reglas establecidas.
- Los componentes del proceso no deben modificarse con frecuencia.
- En lo posible los datos deben ser estructurados.

¿Cómo aplicar/desarrollar RPA?

En el desarrollo de un RPA, los costos son relativamente bajos al igual que su desarrollo. Asimismo, la prueba e implementación son muy cortas en el tiempo, su configuración es muy rápida y la adecuación en los procesos de negocio es de muy bajo impacto. El inicio de las tareas, al ser automatizadas, debe ir de la mano de los procesos de negocio. Esto quiere decir que el aporte de un BPMN es importante en tanto que permite identificar qué tareas son de bastante relevancia en los procesos de negocio y que le aporte un ROI a la organización y muy corto tiempo.

Unos de los pasos recomendados por IBM (2018) son los siguientes:

1. Elija automatizar algunas tareas sencillas y vea si sus parámetros encajan en las capacidades del software RPA que va a probar.

2. Recopile información sobre el tiempo que tarda cada una de ellas, tanto individualmente como de forma conjunta en toda la empresa.
3. Registre los pasos de la tarea mediante el software RPA.
4. Efectúe las actualizaciones necesarias para refinar el script.
5. Ponga el robot en producción durante un periodo de prueba.
6. Mida los resultados y analice el impacto tanto en el flujo de trabajo como en los empleados.
7. Evalúe la ganancia de eficiencia, ahorro de tiempo y mejora de la calidad.

Sumado a lo anterior, dentro de la implantación de un RPA hay que tener en cuenta los roles (Deloitte, 2017) dentro de las funciones de negocio, las organizaciones deberán adaptar tanto su estructura tecnológica como su estructura humana. Además, deben modificar los roles actuales y adoptar e implementar nuevos. En términos de organización, es necesario contar con 5 actores que permitirán obtener los mejores resultados:

1. Desarrolladores de procesos que indiquen las tareas que debe desarrollar el RPA.
2. Un gestor de robots que asigne y monitoree tareas.
3. El robot que es el software instalado en el ambiente de trabajo e interactúa directamente con las aplicaciones del negocio.
4. Los usuarios son aquellos que resuelven las incidencias o situaciones que el robot escala.
5. La aplicación o plataforma mediante la cual el robot interactúa con el usuario.

En la Ilustración 14 se muestran todos los actores que participan en el flujo de trabajo:

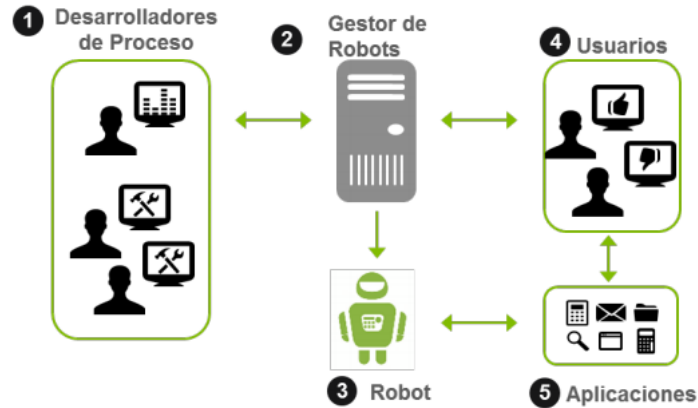


Ilustración 14. Flujo de roles

Fuente: (Deloitte, 2017)

Considerando lo anterior, se deben tener presente las características clave en la implantación de un RPA:

Adopción cultural: en esta parte es importante la participación activa del nivel ejecutivo, es decir, quien apoya el proyecto RPA. Asimismo, garantizar el modelo operativo de robotización de una forma robusta y eficiente, además dedicar un quipo para el manejo de la gestión del cambio.

Compromiso de IT: desde el principio, IT debe asumir una función de gobernanza solididad, es el encargado de propiciar los requerimientos de seguridad, escalabilidad, auditabilidad, recuperación del negocio en caso de presentarse fallos y que se dé la gestión del cambio de una forma apropiada según la necesidad.

Capacidad interna del RPA: el RPA es una capacidad estratégica y tiene como objetivo evolucionar, escalar el apalancamiento operativo e incrementar el valor de los negocios.

2 Estado del arte

El presente apartado se orienta a la revisión bibliográfica de casos de estudios en los cuales se hace referencia al monitoreo de la DQ en los procesos de negocio. En este sentido, su finalidad es tener un acercamiento a investigaciones anteriores que sirven de base para sustentar la propuesta y dilucidar el escenario desde una perspectiva teórica, tal como se presenta a continuación.

2.1 Caso de estudio 1: Allianz

2.1.1 Información del caso

- Nombre del caso: gobierno de datos y gestión de DQ en el sector de seguros.
- Autor(es): Baghi and Abraham 2013
- Referencia bibliográfica: ISBN 978-3-7375-7592-8
- URL: https://www.researchgate.net/publication/303940986_Corporate_Data_Quality_Prerequisite_for_Successful_Business_Models/link/575faa4308ae9a9c955fc221/download

2.1.2 Descripción de la empresa

Allianz SE es una empresa multinacional establecida en 1890, la cual se ubica en el sector de aseguramiento y de servicios financieros, con sede en Múnich, Alemania. Posterior a su fundación, se fusionó como filial de varias divisiones comerciales de riesgos globales y de marcas de marina y aviación. Por ello, en 2016 fue constituida como Allianz Global Corporate & Specialty (AGCS). Para 2013, esta empresa reportaba unos ingresos

financieros de 110.77 billones de EUR y con un lucro de 6.34 billones de EUR, además, contaba con 147 627 empleados para este mismo año.

Cabe señalar que AGCS proporciona servicios de seguros para empresas en 160 países. El alcance de los servicios incluye, entre otras cosas, la transferencia de riesgos y diversas soluciones no tradicionales para la gestión de riesgos, seguros internos de empresas, servicios de reclamos y pólizas de seguros internacionales.

2.1.3 Antecedentes y problemática

AGCS está obligada a someterse a una auditoría por el Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin, autoridad de supervisión financiera federal alemana) cada dos años. Después de la auditoría en 2010, BaFin solicitó que AGCS presentara un concepto de DQ y desarrollara una gestión de DQ en toda la empresa (según lo prescrito por Solvency II en “Implementación y cumplimiento de los requisitos legales y oficiales”, entre otras directivas). Así, el análisis de la situación inicial, antes de la implementación de esta especificación, condujo a los siguientes problemas:

- Falta de claridad en las responsabilidades para la gestión de la DQ: mientras hay flujos de DQ individuales en varios departamentos, el concepto de responsabilidad no es amplio para toda la empresa.
- Ausencia de estructuras de toma de decisiones en los problemas de DQ: los problemas de DQ nunca se corrigieron o se corrigieron por primera vez después de varios niveles de escalamiento, lo que llevó a la compañía a estar en un estado constante de apagar incendios, con respecto a la DQ.
- La carencia de la organización presenta problemas en la entrada de la DQ: no existe una oficina o un responsable de la coordinación de gestión de DQ en la empresa.

2.1.4 Descripción de la solución

Como reacción al resultado del informe de auditoría de BaFin, el jefe financiero de AGCS dio inicio oficial al proyecto Solvencia II en 2011. Por tanto, establecieron un equipo para el gobierno de datos y se redactó un proyecto con los siguientes objetivos:

- Definición de los requisitos más importantes para la DQ y la evaluación de las necesidades de subsidios.
- Definición de los procesos y las instrucciones de manejo para asegurar la DQ.
- Diseño de medidas para la medición y el monitoreo de la DQ.

De este modo, dichas metas fueron incorporadas en un plan de trabajo. Las tareas de trabajo fueron:

- Diseñar una guía para toda la empresa para la gestión de la DQ.
- Definir roles y responsabilidades para la gestión de la DQ.
- Definir los criterios para la medición de la DQ.
- Diseñar procesos y procedimientos para la gestión de la DQ.
- Seleccionar e implementar un sistema de software para la medición permanente de DQ.

Acorde con lo anterior, el proyecto se implementó en tres fases entre 2011 y 2013. La primera incluyó la definición y la asignación de los roles de DQ a los empleados, así como la precisión de una comprensión por parte de toda la compañía sobre el término de datos de calidad. La segunda fase analizó la situación real de los productores de datos para comprender en dónde se generan los datos, en qué sistemas, en qué pasos del proceso y qué se hace con los datos después. En el tercer paso, el equipo de Gobierno de Datos determinó el conjunto de reglas para la DQ que deben cumplir los usuarios con los datos. Por otra

parte, para hacer que la DQ sea medible (y controlable), AGCS definió tres características de la DQ:

- **Integridad:** los datos se considerarán completos cuando sean suficientemente granular para identificar tendencias y describir el riesgo asociado de manera integral.
- **Precisión:** los datos son precisos cuando representan adecuadamente la realidad, por lo tanto, están libres de errores.
- **Idoneidad:** los datos son adecuados cuando son apropiados para los fines comerciales, lo que significa el análisis y la evaluación de riesgos.

Estas características de DQ forman la base para la medición de calidad de estos en AGCS. Para automatizar las mediciones, las tres características se mantienen en las reglas comerciales. En la Tabla 3 se representan algunos ejemplos de reglas de negocio para las tres características.

Tabla 3. Business Rules at AGCS

Fuente: (Otto y Österle, 2015)

<i>Ejemplos de reglas de negocio</i>	
Integridad	La entrada LOB puede no estar vacía
Exactitud	La entrada de fecha debe tener el formato adecuado
	El valor en la entrada Premium puede no ser negativa
Idoneidad	La entrada de la fecha reservada no puede ser de más de 40 días después del ingresado en la entrada Bound Date (un indicador de la calidad del proceso)

Por otro lado, AGCS utilizó una solución de software DataFlux de SAS para monitorear la DQ. La solución accede a los datos de los sistemas operativos, analiza la DQ

y representa los resultados en un software, llamado cockpit, de manejo de la DQ personalizada.

Si bien el software cockpit de manejo de DataFlux de SAS solo puede probar las reglas de DQ que ya se han establecido, la nueva estructura de gobierno de AGCS también puede garantizar que las nuevas reglas se implementan rápidamente, como datos usuarios o productores, para encontrar nuevos errores. Asimismo, el equipo de Gobierno de Datos crea nuevas reglas que pueden implementarse en la herramienta DataFlux, después de la aceptación por parte del grupo directivo de Gobierno de Datos.

2.1.5 Resultados obtenidos

Para comenzar, cabe decir que el llamado ciclo DMAIC (*Define, Measure, Analyze, Improve y Control*) sirvió, en particular, como plantilla para la gestión de procesos de la DQ. En este sentido, el software cockpit proporciona dos vistas de las mediciones de DQ. Una vista es destinada a usuarios especializados (como consumidores de datos o productores de datos), y la segunda es para usuarios técnicos de las fuentes de datos (administradores del sistema). Desde la vista de usuarios, cockpit muestra los requisitos específicos de la DQ, como Solvencia II lo prescribe a nivel normativo con las reglas de negocio que se han definido para cumplir estos requisitos; los usuarios evalúan los informes según el nivel jerárquico de acceso y así pueden evaluar los resultados del monitoreo en DQ, mientras se tienen en cuenta sus categorías de riesgo predefinidas a través de los sistemas fuentes.

Además, el software proporciona una vista de las reglas comerciales que se han roto y los registros responsables del incumplimiento. Gracias a esta función de desglosa, AGCS ha podido rastrear datos erróneos en las especificaciones de las reglas de negocio sobre los

sistemas fuentes que lo originan e, incluso, los datos defectuosos; todo esto es registrado por cockpit. Por otra parte, la vista de cara a los administradores del sistema les muestra las fuentes de datos, los registros inconsistentes, las reglas comerciales fallidas y, además, se puede hacer el seguimiento del ciclo de vida de los datos del sistema.

2.1.6 Lecciones aprendidas

- Los enfoques comprobados para la gestión de la calidad (como el ciclo DMAIC) pueden transferirse a la gestión de la DQ en toda la empresa. En esto, el monitoreo de la DQ se convierte en un proceso que debe llevarse a cabo continuamente en lugar de solo una vez.
- Los enfoques de gestión de riesgos ayudan a evaluar la DQ.
- Monitorear la DQ y el trabajo de gobernanza de datos de la mano, sin Roles ni responsabilidades claras, medidas para asegurar y mejorar de DQ, no funcionará.
- En el complejo proceso y panorama del sistema, la DQ debe medirse tanto desde el punto de vista de los consumidores de datos como del de las fuentes de datos.

2.2 Caso de estudio 2: Bayer CropScience

2.2.1 Información del caso

- Nombre del caso: control de DQ en la industria agroquímica
- Autor(es): Brauer, Berthold 2009
- Referencia bibliográfica: ISBN 978-3-7375-7592-8

- URL: https://www.researchgate.net/publication/303940986_Corporate_Data_Qualit_y_Prerequisite_for_Successful_Business_Models/link/575faa4308ae9a9c955fc221/download

2.2.2 Descripción de la empresa

Bayer CropScience AG es un subgrupo de Bayer AG¹⁹. Bayer AG está organizada en tres subgrupos operativos (Bayer Cuidado en Salud, Bayer Industria Agroquímica y Bayer Ciencia de Materiales) y tres empresas de servicios (Bayer Servicios de Negocio, Bayer Servicios Tecnológicos y Currenta Parque Químico).

Bayer CropScience (BCS) está activo en los campos de protección de cultivos, control de plagas no agrícolas, material de siembra y biotecnología de cultivos. Con 22 400 empleados e ingresos de 8800 millones de euros en el año fiscal 2013, Bayer CropScience es el líder del mercado en el campo de los productos agroquímicos. Cabe mencionar que Bayer CropScience se creó en 2002, mediante la adquisición de Aventis CropScience por Bayer AG. Por otra parte, Bayer CropScience se ha organizado en dos áreas comerciales operativas:

- Protección de cultivos y semillas: en esta área comercial se da la fabricación de productos para la protección de plantas, es decir, herbicidas, fungicidas, pesticidas y de tratamientos de semillas. El subgrupo de semillas planta materiales de siembra para los cultivos claves, como algodón, canola, arroz y vegetales.
- Ciencias ambientales: esta área de negocios se especializa en productos no agrícolas y ofrece artículos para el control de plagas, desde el nivel de hogar y jardín al nivel de la silvicultura.

De igual manera, las tres características que distinguen el mercado de Bayer CropScience están focalizadas en la regulación, la cual está dedicada a obtener permisos para la venta de su producto según las normas de cada país, para efectos de regularizar y de las licencias sanitarias según el caso; en los gastos de investigación y el desarrollo, el cual está asociado a nuevos productos y, por consiguiente, a altos niveles de inversión, en especial en la industria farmacéutica. Por último, en la estacionalidad, la cual se da en el mercado de materia prima, en donde la fluctuación de precios es proporcional al cambio de climas en las diferentes zonal mundiales.

Como es el caso con Bayer CropScience, muchas compañías en esta industria reaccionan a esto a través de la integración de la cadena de suministro desde el agricultor hasta las tiendas minoristas, distribuidores y mayoristas, hasta los fabricantes de medidas de control de plagas y primas proveedores de material (Wallich 2013).

2.2.3 Antecedentes y problemática

En 2007, Bayer CropScience tomó la decisión de armonizar los procesos comerciales a nivel global; dentro de los objetivos estaba:

- Mejorar los requisitos de capital y planificación de liquidez, a través del uso de un proceso universal para planificar y gestionar la producción y los procesos de adquisición de materiales, a partir de la planificación regional de ventas.
- Perfeccionar el sistema de informes, mediante el uso de una base unificada.

Al respecto, uno de los primeros pasos fue consolidar los sistemas usados por las filiales nacionales de más de 120 países en tres sistemas regionales (Europa, Asia-Pacífico y las Américas). Como parte de este proyecto se centralizaron los datos maestros de materiales, clientes y proveedores, el cual fue llamado Golden Box y su objetivo era

centralizar y distribuir los datos maestros a los sistemas regionales. Después de armonizar los sistemas, el proyecto Future System Landscape (FSL) inició con uno de los objetivos, que era establecer un sistema ERP (*Enterprise Resource Planning*) global.

Luego de implementarse el sistema ERP central y al estar armonizados los datos maestros locales, se inició la fase de integración de los procesos de planificación de la cadena de suministros para la región de Asia y el Pacífico.

Ahora, las dificultades con la provisión de datos para el rodaje surgieron al comenzar con el taller de lanzamiento. Los empleados del proyecto descubrieron problemas, como aquellos con la consolidación de requisitos para ingredientes con la determinación de precios para productos y con el programa y la planificación de cartera. Como parte del taller, lograron rastrear estos problemas a datos erróneos, por ejemplo, en la jerarquía del producto. A partir de esto, verificaron una conexión entre la falta de DQ y varios procesos de negocios, como el código de afiliación de un producto al área de negocio, la unidad de negocio y el segmento de negocio. Es menester señalar que la DQ de la jerarquía del producto es un requisito previo para tres procesos de negocio: planificación, informes y producto, como se evidencia en la Ilustración 15, pues la falta de la DQ ocasiona diversos problemas en el negocio como se muestra en la Ilustración 16.

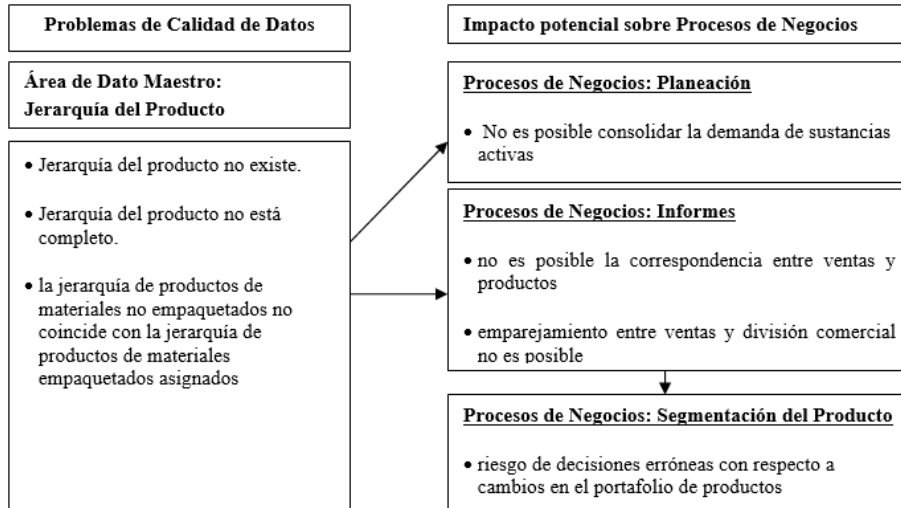


Ilustración 15. Importancia de la calidad de los datos de la jerarquía de productos para negocios procesados en Bayer CropScience

Fuente: (Otto y Österle, 2015)



Ilustración 16. Causas de problemas en la calidad de los datos en Bayer CropScience

Fuente: (Otto y Österle, 2015)

Los ejemplos incluyeron la falta de conocimiento de la importancia de los datos por parte de los empleados, la ausencia de especificaciones para la adquisición y el mantenimiento de los datos, la carencia de responsabilidades y la falta de medición de la DQ. Hasta ese momento, Bayer CropScience no tenía un control de DQ periódico o cualquier medida preventiva para la gestión de DQ. La mejora de datos había sido reactiva, medir en casos individuales, por ejemplo, después de errores que podrían rastrearse; al respecto, observó una falta de DQ en los sistemas de informes.

La idea de que la DQ era un requisito previo necesario para el proyecto FSL y, por lo tanto, los objetivos operativos de la armonización del procesamiento comercial en la compañía tomaron la decisión de comenzar una gestión de DQ en toda la compañía como proyecto.

2.2.4 Descripción de la solución

Con la implantación del software cockpit, que está enfocado en funcionalidades de medición y monitoreo de DQ, los requisitos funcionales definidos fueron:

- Medición de DQ.
- Presentación gráfica de los resultados medidos.
- Almacenamiento de los resultados obtenidos durante un periodo de doce meses para poder identificar tendencias.
- Soporte para vistas específicas de usuarios y roles.

En este sentido, el cockpit entró como apoyo en DQ al proyecto principal FSL, mientras que el proyecto de DQ persiguió cuatro objetivos:

- Crear conciencia: sensibilización de los departamentos especializados y filiales nacionales en temas de la DQ, además de la demostración de la necesidad de una gestión de DQ en toda la empresa.
- Creación de transparencia: demostración de las conexiones entre datos defectos y problemas comerciales, y la justificación de medidas para mejorar la DQ.
- Reforzar la colaboración: crear aceptación de la iniciativa entre los usuarios responsables en las áreas especializadas para asegurar su apoyo.
- Identificar la necesidad de actuar: determinación de la necesidad de actuar en diversas áreas para poder mejorar la DQ, mantenerla estable y de manera sostenible

De esta forma, la introducción del software cockpit se produjo, inicialmente, en quince filiales nacionales en la región Asia-Pacífico. Al principio, solo los datos maestros de material eran tomados en consideración, al hacerlo, BCS determinó que podría ser extendido a más filiales nacionales y transfirió la responsabilidad de la DQ a los gerentes de las filiales; además, se dispuso de un coordinador de datos maestros para cada país.

Es importante afirmar que la DQ sirve como apoyo de los procesos de la empresa, por tal razón, se determinaron las reglas para la DQ como requisitos para el proceso del negocio. Para eso, Bayer inició unas entrevistas con los expertos en los procesos. En este orden de ideas, para garantizar la compatibilidad y la integridad de las reglas, fueron determinadas tres propiedades:

- Medibilidad: la regla se puede medir técnicamente, lo que significa que todos los datos medidos estarán disponibles y comparables a través de las propiedades formuladas por la regla.

- Relevancia para el negocio: existe una relación causal entre la regla y los objetivos comerciales de la empresa, es decir que el efecto técnico de romperse puede rastrearse según la integridad de la regla.
- Conformidad con los negocios: la regla está cerca de las rutinas comerciales diarias y según las actividades que se deben realizar en la operación diaria, rutinas de negocios.

2.2.5 Resultados obtenidos

- A partir de los exámenes individuales se formó un índice de calidad de datos (DQI) de las reglas de comparabilidad de los valores medidos. Se calculó el índice como el cociente del número de esos registros que no rompieron ninguna regla y el número total de registros. Con estos indicadores se mostraron en unos tableros, mientras se señalaba la unidad de negocio con el mejor valor DQI, a la par que las peores unidades de negocio permanecían anónimas. Esta representación comparativa generó un cierto nivel de competencia y motivó a los empleados a mejorar sus datos, para alcanzar un nivel superior en el próximo mes.
- Los empleados de Bayer pudieron determinar las causas de la baja calidad de sus resultados y tomar las medidas correctivas del caso.
- La herramienta de DQ guardó los resultados de DQI anteriores, para que Bayer pudiera observar el desarrollo de la DQ a lo largo de tiempo y así realizar comparaciones estratégicas para tomas de decisiones en los negocios.
- Simplicidad: una métrica única y fácil de entender, como el DQI, es más simple de comprender y es más definitivo que una multitud de resúmenes de métricas.

- Relevancia: las conexiones causales claramente definidas entre la DQ y los problemas de negocios aseguraron que el enfoque se centra en lo más importante, los problemas de DQ, y garantizó el apoyo de los empleados y la administración.
- Conocimiento del tema: este conocimiento del negocio diario fue necesario para la identificación de los defectos en los datos críticos para este.
- Gestión: el apoyo de la dirección de la empresa aumentó la visibilidad y la aceptación de las mediciones de DQ en la empresa.
- Visualización: la presentación de los resultados de medición, que se dirigen a grupos específicos, se entendió fácilmente y fue un requisito previo para la aceptación de las mediciones de DQ por parte de los empleados.
- Comparabilidad: la comparación de los valores de medición entre varios países aumentó la motivación de los empleados.

2.2.6 Lecciones aprendidas

- La DQ es un requisito previo para los procesos comerciales centrales, como los financieros y la planeación de la producción.
- La gestión de la DQ no es un proyecto individual, sino que debe ser anclado en la organización al utilizar roles de gobierno de datos.
- La medición periódica de la DQ es un requisito previo para su gestión y, por lo tanto, su mejora.
- La DQ debe ser un componente de los acuerdos de objetivos con los empleados.

2.3 Caso de estudio 3: Johnson & Johnson

2.3.1 Información del caso

- Nombre del caso: institucionalización de datos maestros para la gestión de datos en los productos de la división consumo.
- Autor(es): Otto 2013
- Referencia bibliográfica: ISBN 978-3-7375-7592-8
- URL: https://www.researchgate.net/publication/303940986_Corporate_Data_Quality_Prerequisite_for_Successful_Business_Models/link/575faa4308ae9a9c955fc221/download

2.3.2 Descripción de la empresa

Johnson & Johnson es una compañía Fortune-50 con más de 275 subsidiarias en 60 países de todo el mundo. Johnson & Johnson se compone por tres divisiones de negocios: farmacéutica, dispositivos médicos, diagnóstico y productos de consumo. En 2013, la compañía empleó a 128 000 personas y obtuvo ingresos de 55 000 millones de euros. Además, las divisiones de farmacéutica y las de dispositivos médicos y diagnóstico se gestionan de forma centralizada. En contraste, la división de productos de consumo se ha dividido en cuatro regiones geográficas (América del Norte, América del Sur, Asia-Pacífico y Europa).

2.3.3 Antecedentes y problemática

Además de otros objetivos de la compañía, Johnson & Johnson tuvo un crecimiento continuo a través de adquisiciones corporativas. Una adquisición que fue observada en todo

el mundo ocurrió en 2006, cuando Johnson & Johnson compró Pfizer, por 16 600 millones de dólares. Mientras tanto, Johnson & Johnson había subcontratado dos tercios de su producción para concentrarse en sus competencias básicas. Es preciso aclarar que el caso de estudio se centra en las actividades de gestión de datos en los productos de la división de consumo (América del Norte).

No menos importante, debido a la considerable cantidad de compras de la compañía, los procesos de negocio de Johnson & Johnson no habían sido armonizados en gran parte, sino que fueron muy diferentes en las divisiones de negocios y de empresas subsidiarias. Así, por ejemplo, no había un proceso para las fijaciones de precios para toda la empresa y tampoco existía una gestión de datos general. En cambio, había cinco grupos de gestión de datos, que trabajaban independientemente el uno del otro.

En 2005, Johnson & Johnson comenzó un gran proyecto para introducir el software SAP ERP en toda la empresa. El objetivo era usar una aplicación estándar para el sistema de planificación, producción y comercialización de productos de la división de consumo. El proyecto también incluyó la introducción de herramientas de software para creación de datos maestros.

Sin embargo, los procesos de gestión de datos individuales, como la creación y el mantenimiento de los datos maestros de material, eran incompletos e indefinidos en sus detalles. De igual manera, estos procesos todavía estaban en organización a nivel local o regional; en consecuencia, no eran uniformes, porque los diseños de procesos y de sistemas en este proyecto no habían sido lo suficientemente coordinados entre sí: la inversión en el proyecto no proporcionó las ventajas esperadas. Con respecto a la gestión de datos, Johnson & Johnson tuvo que combatir tres dificultades significativas:

- Primero, varios procesos de negocios sufrieron errores que indicaron problemas en DQ: con frecuencia se enviaron facturas erróneas a los clientes, los camiones tenían que esperar en las terminales para cargar hasta que los materiales correspondientes fueran aprobados y los retrasos en la producción se debieron a que los pedidos no se habían realizado de manera oportuna en los sistemas de fabricación. Además, la información sobre el estado de los productos nuevos en el ingreso no era claro en el flujo de la responsabilidad de todo el proceso.
- En segundo lugar, los operadores de grupos de datos globales, como la Sincronización de datos Global Red (GDSN) se quejó de la calidad de los datos transferidos por Johnson y Johnson. Los clientes también criticaron con frecuencia la DQ en la logística, tales como los pesos y las dimensiones del producto. En este sentido, uno de los clientes más importantes de Johnson & Johnson informó que la empresa era una de las más pobres como proveedores estratégicos para datos logísticos.
- En tercer lugar, Johnson & Johnson juzgó su gestión de datos como ineficiente. Por lo tanto, los empleados en la gestión de datos utilizaron aproximadamente el ochenta por ciento de su tiempo para el análisis de errores de datos y la corrección de problemas de datos.

Por ello, GS1 puso a disposición su equipo CubiScan, que se utilizó para escanear todos los productos físicamente. En un mes, cada producto activo había sido analizado de esta forma. Los resultados presentados a la gerencia evidenciaron unos importantes problemas de DQ. Así, se expuso el hecho de que menos del treinta por ciento de los datos

para el peso y las dimensiones de los productos no estaba dentro del cinco por ciento permitido dentro del rango de tolerancia a errores.

2.3.4 Descripción de la solución

Para dar solución a las anteriores problemáticas expuestas, Johnson & Johnson inició un proyecto en tres fases que se exponen a continuación.

Fase de establecimiento

Esta fase se inició en la primavera de 2008. La gerencia de Johnson & Johnson decidió introducir un departamento para toda la empresa, llamado maestro de datos empresarial (EMD), para garantizar una DQ suficiente en la empresa y evitar errores en los procesos de negocio en el futuro.

Al gerente designado para el departamento de EMD se le atribuyó la tarea de desarrollar la nueva organización en ocho meses. Así, fue responsable de convencer a todas las divisiones de negocios para entregar sus propias actividades descentralizadas en la gestión de datos y colocar esas actividades desde la nueva responsabilidad central del EMD. Debido a que la iniciativa fue apoyada por los clientes de la gerencia, la cuestión no tuvo que ser negociada, sino que la discusión se limitó a cómo realizarla. Durante esta fase de comprensión de la propiedad de los datos, estos se convirtieron gradualmente en “nuestros datos”, como conciencia del valor de los datos compartidos por toda la empresa. En definitiva, los gerentes a nivel de vicepresidente, para todas las divisiones de los negocios, reconocieron la necesidad de esta iniciativa compartida.

Por otra parte, el equipo de EMD definió los roles y las responsabilidades para el uso y el mantenimiento de los datos maestros, por lo que determinó la responsabilidad y se identificaron 420 atributos que fueron nombrados. Posteriormente, el equipo de EMD

desarrolló reglas para el mantenimiento del atributo para un tipo de material, luego el equipo comenzó con las reglas especificadas en sistemas en uso (como SAP) y, posteriormente, desarrollaron las reglas comerciales adicionales, a través de consulta con los expertos para los procesos de negocio. Un ejemplo de lo anterior es: la regla de negocio determinó que el atributo de peso neto tenía que mantenerse para cada producto terminado en los sistemas Johnson & Johnson.

Fase de desarrollo

En mayo de 2009, la nueva organización central de EMD comenzó a funcionar. De este modo, asumieron la responsabilidad de la disponibilidad oportuna de los datos para un nuevo producto necesario, al nivel de la calidad. Al principio, las personas responsables de los datos se asignaron de acuerdo con la clase de datos y de la división de negocios; mientras que los miembros del equipo EMD recibieron oficinas compartidas en cada una de las ubicaciones de la empresa. De este modo, al comienzo del proyecto, todo el equipo de EMD, compuesto por 27 personas (16 empleados internos y 11 empleados externos), había sido reubicado en la sede de la compañía temporalmente.

La fase de desarrollo fue apoyada por comités de datos maestros y un comité directivo. El comité directivo evaluó periódicamente los procesos de datos maestros, con respecto al cumplimiento de las normas, el logro de los objetivos de calidad y la provisión puntual de los datos. Esto tomó lugar en los doce departamentos que participaron en la creación de datos maestros.

Además, Johnson & Johnson estableció una nueva gestión de datos en toda la empresa, el cual incluía el soporte por flujo de trabajo en la creación de datos de productos y un sistema de monitoreo de DQ. Este proceso fue asignado al recién creado equipo de EMD.

Fase de madurez

A mediados de 2011, la gestión de datos en Johnson & Johnson había logrado un alto nivel de madurez; se habían integrado los procesos para la gestión de datos en las rutinas comerciales diarias y fueron aceptadas por todos en la empresa. Por ello, el sistema central de información de gestión de datos continuó mejorando. En conclusión, Johnson & Johnson estaba en condiciones de acceder a datos del producto rápidamente y con un buen nivel de calidad.

2.3.5 Resultados obtenidos

Hoy en día los procesos comerciales en Johnson & Johnson utilizan datos maestros (como datos del producto) de manera consistente. Para esto, el EMD creó una comprensión única de los datos esenciales en toda la empresa y, además, construyó los requisitos para la gestión de datos para todos los procesos de negocio.

Al respecto, el software cockpit es una herramienta de análisis y de informes para la DQ. Con ella se comprueba si los registros de datos maestros han roto las reglas definidas o violan la calidad requisitos. Mientras que los informes de gestión proporcionan indicadores clave de rendimiento para la DQ.

2.3.6 Lecciones aprendidas

Johnson & Johnson logró muchos de los objetivos de gobernanza de datos establecidos en solo tres años. El establecimiento del departamento de EMD suavizó el camino hacia el objetivo superior de mejorar la DQ a través de la empresa.

Ya no se discute que el factor motivador esencial para la iniciativa vino de afuera de la empresa, pues hubo motivadores para la gestión de la DQ dentro de la empresa; sin

embargo, un factor motivador fue la crítica de los clientes, quienes se expresaron con fuerza.

En este sentido, era importante que los gerentes de la empresa respondieran con rapidez a la llamada de atención basada en las críticas de los clientes. Si bien habían considerado problemas de DQ como fenómenos que ocurren en toda la empresa, sin ningún contexto, Johnson & Johnson luego se dio cuenta de que este problema era relevante para toda la compañía y tendría que ser abordado a nivel corporativo general.

Ahora, sobre el seguimiento diario de la DQ y la documentación del análisis, cabe mencionar que Johnson y Johnson podría rastrear los efectos de ciertos incidentes en la DQ. A través de esto, la compañía podría adaptarse mejor a los eventos en el futuro e intentar reducir los efectos negativos de estos en la DQ. Además, fue importante reconocer que la gobernanza de datos tuvo un efecto positivo en la DQ, que fue claramente demostrado por la medición y el monitoreo continuo de esta.

Igualmente, la mejora permanente de la DQ requiere una organización focalizada para la gestión de DQ en toda la empresa. En cambio, la validación digital automatizada de la DQ, mediante reglas comerciales, debe combinarse con pruebas físicas parcialmente manuales (en este caso, con la ayuda del dispositivo CubiScan), pues probar el objeto del mundo real garantiza la exactitud de los datos.

Finalmente, los flujos de trabajo y las áreas preparatorias (áreas de preparación), antes de la aceptación de los datos en el sistema ERP, evitan la adquisición de datos erróneos y aseguran la DQ en el sistema productivo.

2.4 Caso de estudio 4: HILTI

2.4.1 Información del caso

- Nombre del caso: gestión universal de los datos de los clientes en la industria de las herramientas y los cierres.
- Autor(es): Baghi and Ebner, 2013
- Referencia bibliográfica: ISBN 978-3-7375-7592-8
- URL: https://www.researchgate.net/publication/303940986_Corporate_Data_Quality_Prerequisite_for_Successful_Business_Models/link/575faa4308ae9a9c955fc221/download

2.4.2 Descripción de la empresa

Hilti suministra productos, sistemas y servicios tecnológicos de vanguardia para herramientas y tecnologías de consolidación, para la industria de la construcción en todo el mundo. La sede del grupo Hilti está en Schaan, en el Principado de Liechtenstein. El espectro de productos incluye los sistemas de medición por láser, los sistemas de anclaje e instalación, las herramientas de perforación y demolición, los sistemas de perforación con punta de diamante y los equipos de corte y pulido. Además, Hilti AG está activa en todo el mundo, con 21 000 empleados en más de 120 países. La compañía opera plantas de producción e instituciones de investigación y desarrollo en Europa, Asia y América Latina.

Por otra parte, la organización corporativa de Hilti incluye las divisiones de investigación y tecnología corporativa, cadena de suministro y corporativa central, así como las unidades de negocio. Cada área de negocio se ha dividido en varias líneas de productos. El Departamento de Alcance de Mercado (MR) es responsable de todas las

actividades de ventas y *marketing* de la empresa, que responde ante el jefe de ventas y de *marketing* global. En este sentido, el negocio de Hilti se ha basado tradicionalmente en un modelo de comercialización directa. Así, dos tercios de los empleados trabajan en el Departamento de *Marketing* y tienen contacto diario con hasta 200 000 clientes en todo el mundo. Además de la comercialización directa, Hilti tiene cuatro canales de venta adicionales, que representan puntos de contacto con los clientes. Asimismo, Hilti tiene sus propios puntos de venta (Hilti Center), tiendas en empresas asociadas (ProShop), centros de atención al cliente y una tienda online, además de la fuerza de ventas externa.

Ahora, cabe exponer que la visión (2015) de Hilti define los objetivos estratégicos de la empresa y considera que la satisfacción del cliente es una condición importante para el objetivo de crecimiento sostenible de la rentabilidad deseada. Para asegurar la satisfacción del cliente en todos los canales de venta y trabajar eficientemente al mismo tiempo, Hilti requiere de los datos del cliente de alta calidad. Además de los datos del sistema ERP, esto también afecta al sistema central de gestión de relaciones con los clientes (CRM) de Hilti, que apoya, por ejemplo, los procesos de comercialización y de ventas, así como la gestión de contactos y campañas. De este modo, los atributos de los datos maestros de los clientes afectan directamente al buen funcionamiento de varios procesos comerciales.

2.4.3 Antecedentes y problemática

A pesar de la importancia significativa de la corrección de los datos de los clientes, Hilti no tuvo una gestión de DQ en toda la empresa hasta 2006. Por esa razón, la compañía determinó los tres riesgos siguientes en la gestión de los datos de los clientes.

- Uso de los datos: Hilti utiliza datos de varios sistemas, así como de varias entradas de datos en el mismo sistema, para calcular los precios específicos de los clientes.

Los datos erróneos en estas entradas llevarían a cálculos incorrectos y pondrían en peligro la satisfacción del cliente y la capacidad de generar beneficios.

- Medición de la DQ: debido a que la DQ de los clientes no se mide periódicamente, los defectos de los datos generalmente aparecieron por primera vez como problemas en los procesos comerciales. Las consecuencias fueron la investigación de las causas y las medidas correctivas tomadas por los empleados de Hilti en el Departamento de Gestión de Datos. Además, no existían directrices sobre las normas de DQ a través de las fronteras de la empresa, porque los empleados habían mantenido predominantemente los datos de los clientes a nivel local, por efectos de la presentación de informes o de conversaciones y reuniones con los clientes.
- Proceso de mantenimiento de datos: no se había definido un proceso de creación y de mantenimiento de datos maestros de los clientes que se ajustara a los requisitos de la fuerza de ventas externa, en cuanto a la velocidad, la fiabilidad y la movilidad.

En este orden de ideas, el Departamento de Gestión de Datos de Clientes (CDM) esperaba conseguir las siguientes ventajas para las actividades de venta y comercialización de Hilti, a partir de una mejor gestión de la DQ.

- Aumento de la satisfacción del cliente y mejora de la lealtad de este.
- Facilitación de la planificación estratégica mediante mejores análisis de los clientes (como la segmentación de los clientes).
- Reducción de los costos de gestión de los contactos.
- Mejor apoyo a los procesos de logística y facturación.
- Reducción de los gastos de mantenimiento de datos causados por registros erróneos de datos de clientes.

Para lograr esos objetivos, Hilti puso en marcha en 2006 el proyecto Herramienta de calidad de datos de clientes, para la gestión sostenible de la DQ. Para ello, el Departamento de Alcance del Mercado financió el proyecto y sugirió una duración de un año. En este caso, el epónimo de Herramienta de calidad de datos de clientes fue solo uno de los varios nuevos de un total de cuatro medidas exhaustivas. Las cuatro medidas llamadas Definir, Prevenir, Reconocer y Corregir (ver Ilustración 17) cubrieron todo el ciclo de vida de los datos de los clientes; en las cuales se incluyen actividades tanto proactivas como reactivas para la gestión de la DQ.

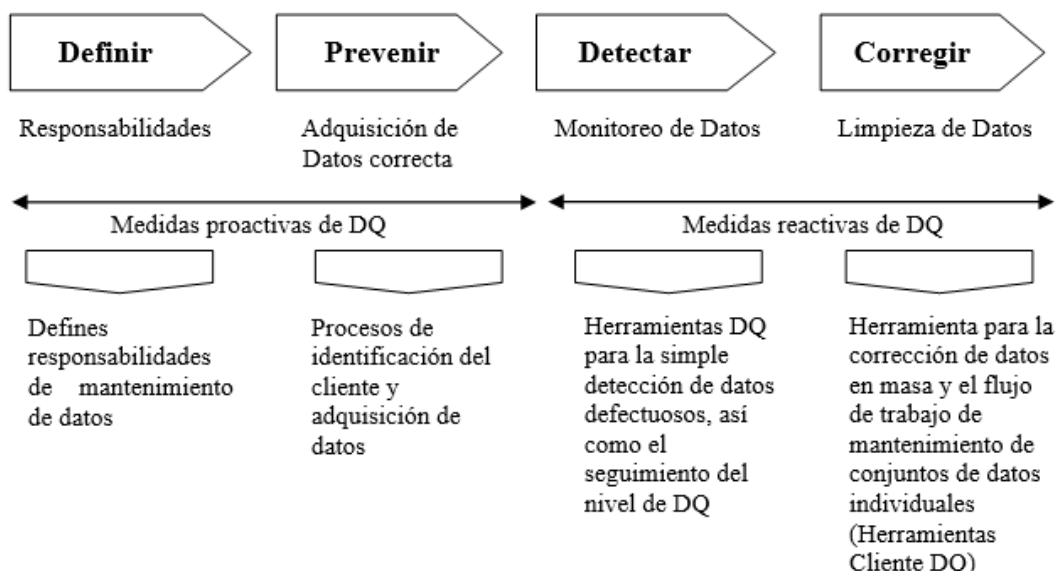


Ilustración 17. Medidas para la gestión proactiva y reactiva de los datos del cliente

Fuente: (Otto y Oesterle, 2015)

Al principio del proyecto, el Departamento de Alcance del Mercado reunió a un grupo de expertos que se encargaría de la cuestión de la gestión de los datos de los clientes. Estos expertos en procesos locales (EPL) identificaron, primero, los problemas de DQ y derivaron de ellos las reglas de negocio para evitar estos problemas. Los EPL determinaron que la integridad y la precisión eran las dimensiones más importantes de la DQ para Hilti.

Además, el equipo del proyecto definió los siguientes objetivos para la iniciativa de DQ del cliente.

- Creación de conciencia sobre la necesidad de la gestión de la DQ por parte de la dirección de la empresa, con la ayuda de sencillos indicadores clave que permiten realizar comparaciones internas de rendimiento.
- Creación de transparencia en cuanto a la DQ de los clientes.
- Establecimiento de funciones de control y de vigilancia para seguir las tendencias y la evolución de la DQ de los clientes de manera continua.
- Inicio de las actividades adecuadas de corrección de datos y de mejora de la calidad para las esferas de acción identificadas.
- Establecimiento de un proceso continuo para el mantenimiento de los datos de los clientes.

2.4.4 Descripción de la solución

Responsabilidades

Como primera medida proactiva, Hilti definió nuevas funciones y responsabilidades para los diversos procesos de gestión de datos, como la corrección de los defectos de los datos y la presentación de informes sobre el estado del DQI (véase la siguiente sección sobre la Monitoreo de los Datos).

En este caso, los denominados gestores de país tendrían el máximo nivel de responsabilidad en la gestión de los datos de los clientes y delegarían las tareas en los expertos en procesos locales (LPE). Ellos, a su vez, asumirían la responsabilidad de la gestión de los datos de los clientes a nivel local y coordinarían las actividades de corrección de datos y de mejora de la calidad a nivel nacional y regional. Según el tipo de entrada de

datos, colaborarían con diversos departamentos funcionales locales (como los de Ventas, Logística, Comercialización o Contabilidad) e introducirían las medidas correctivas necesarias.

Los EPL también crearon archivos de corrección de errores que podían ser descargados de la intranet por los departamentos funcionales. Para ello, se definieron varias medidas de corrección de datos. Los EPL aplicaron medidas de corrección en masa, mientras que los cambios en las entradas de datos individuales fueron sugeridos por los empleados de la fuerza de ventas externa y se activaron después del proceso de aprobación respectivo; todas estas actividades se integraron en los diversos procesos comerciales existentes. Asimismo, la comunicación entre los EPL y los departamentos funcionales fue bidireccional. Esto significa que los departamentos funcionales también podían iniciar procesos correctivos, cuando detectaban problemas en sus procesos que podían rastrear hasta la DQ defectuosos. En algunos casos (como los cambios masivos), los departamentos funcionales también tenían derecho a solicitar que el LPE realizara determinados cambios.

Adquisición de datos correctas

Como segunda medida proactiva para la gestión de los datos de los clientes, Hilti reelaboró el proceso de creación de los datos de los clientes. Este proceso sería, entonces, soportado por un flujo de trabajo del proceso de creación de datos, que implementa las reglas de negocio previamente desarrolladas en el sistema o los sistemas de validación de DQ automatizados. Además, en los flujos de trabajo se utilizaron datos de proveedores de datos externos (como los datos de direcciones) para su validación durante el proceso de creación. Esto redujo el riesgo de que se introdujeran datos erróneos en el sistema (principio de primera vez correcto) durante la creación de un registro de datos del cliente. Por último, los principales usuarios del flujo de trabajo fueron los centros de servicio al

cliente, que fueron responsables del primer paso en la creación de los datos de los clientes. A continuación, el flujo de trabajo tenía varios pasos de aprobación.

Monitoreo en calidad de datos

Una de las dos medidas reactivas para la gestión de la DQ en Hilti fue un instrumento de vigilancia continua de la DQ, que creó transparencia en cuanto a la DQ de los clientes de toda la empresa. En primer lugar, esto requería un DQI que cruzara los límites internos de la empresa y que fuera reconocido por todos los departamentos participantes. En segundo lugar, se requería una herramienta correspondiente que calculara el DQI y posibilitara los informes de evaluación.

El DQI evaluó los dos atributos (exactitud e integridad) que se habían identificado como particularmente importantes para los procesos de los clientes de Hilti al comienzo del proyecto. El cálculo del DQI se basó en las reglas de negocio que medían los atributos críticos del cliente. Como valor objetivo, se consideró que un DQI de al menos el noventa por ciento en todos los canales y sistemas de venta (CRM, ERP e inteligencia empresarial) era un indicador de que los datos de los clientes eran suficientemente completos y estaban libres de errores; estos valores objetivo se aplicaban a todas las unidades de negocio locales.

Por su parte, el departamento interno de TI de Hilti desarrolló su propia herramienta (la herramienta de seguimiento de la DQ) para calcular el DQI y para supervisar la DQ. Anteriormente, los sistemas de inteligencia empresarial (BI) y de planificación de recursos empresariales (ERP) de Hilti se consideraban plataformas potenciales. Sin embargo, estas opciones se rechazaron porque la plataforma de BI no contenía todas las entradas de datos necesarias (como contacto) en el momento del proyecto y se consideró que el proceso periódico de análisis de datos podría limitar el rendimiento

del sistema, en el caso del sistema de planificación de recursos empresariales. Finalmente, el equipo del proyecto se decidió por una solución basada en Microsoft Access, que calcularía el DQI sobre una base mensual. Aparte de esto, los defectos de los datos existentes se identificarían mediante un sistema de vigilancia periódica. En ese caso, el departamento responsable pondría en marcha las correspondientes actividades de corrección de datos para los atributos afectados. En esto, los administradores responsables de los respectivos procesos de gestión de datos.

Paralelamente al nuevo DQI y al instrumento de vigilancia, Hilti introdujo cambios organizativos. Así, el equipo del proyecto elaboró varios tipos de documentación sobre los problemas de los datos y diseñó un plan de capacitación para contrarrestar las posibles fuentes de errores. Además, todos los procesos relativos a la gestión de los datos de los clientes se orientaron a las nuevas normas comerciales. El DQI y el sistema de monitorización periódica crearon, inicialmente, transparencia en cuanto a la DQ en los diversos departamentos y regiones comerciales de Hilti. Junto con la estructura de gobernanza, esto contribuyó a que los departamentos responsables respondieran con rapidez en caso de que fuera necesario adoptar medidas.

De esta manera, el instrumento de monitoreo se extendió a todas las sucursales después de su elaboración, comenzando con siete países. La herramienta se introdujo en un paso más relacionado con la puesta en marcha simultánea del sistema de planificación de los recursos institucionales. Desde finales de 2009, esta herramienta se ha empezado a utilizar en todas las regiones y países.

La corrección de datos y la herramienta de calidad de datos del cliente

La medida con mayor efecto en las rutinas diarias de la fuerza de ventas externa fue la introducción del flujo de trabajo de la calidad de datos de clientes (flujo de trabajo en

calidad de datos del cliente). Este nuevo proceso utilizaba una aplicación de *smartphone* para la corrección de datos, que también fue desarrollada internamente por Hilti. La aplicación proporciona a la fuerza de ventas externa la capacidad de adquirir datos directamente del cliente e, incluso, iniciar un proceso de corrección de datos obsoletos o erróneos en el sitio. Por lo tanto, el proceso puede atribuirse a las medidas de gestión de la DQ tanto reactivas como proactivas.

Mientras utiliza este flujo de trabajo, Hilti podría hacer cambios en los datos de los clientes de manera directa desde la fuente, en donde el nivel de interés propio y motivación sería particularmente alto para el empleado de ventas. El director de ventas de área y el administrador de *back office* podrían rastrear todas las solicitudes de cambio para los miembros de su equipo y, por ejemplo, supervisar dónde se produjeron más cambios para qué clientes y para qué entradas de datos. Además, el flujo de trabajo aseguraba la consistencia de los datos y entre los diversos sistemas de Hilti.

Como medida reactiva para las correcciones, Hilti introdujo una nueva herramienta para corregir los datos en masa (la herramienta de mantenimiento en masa). Esto implicó un desarrollo personalizado de Hilti, que utilizó el sistema ERP como base. La herramienta de mantenimiento en masa permitió realizar cambios y correcciones en grandes bases de datos en cortos períodos de tiempo. Los EPL, que hacían cambios y correcciones por instrucción de los departamentos funcionales, serían los principales usuarios de esta herramienta.

2.4.5 Resultados obtenidos

Desde la aplicación de las nuevas soluciones para la gestión de los datos maestros de los clientes, la DQ de los clientes en Hilti ha aumentado, como se muestra en el

desarrollo del DQI a lo largo del tiempo. Dado que Hilti sigue añadiendo nuevas reglas comerciales para el cálculo de este valor, es necesario mejorar continuamente la gestión de la DQ para mantener el DQI al mismo nivel. Con ello, el negocio operacional recibe apoyo de varias maneras. Por ejemplo, los empleados de ventas pueden segmentar a los clientes con mayor eficacia y proporcionarles ofertas que se ajusten mejor a sus necesidades. Esto ha tenido efectos positivos en la satisfacción general de los clientes e incluso ha asegurado una mejor lealtad de estos. En este sentido:

- La DQ de los clientes es un requisito previo para un modelo de comercialización directa.
- La DQ de los clientes se puede asegurar mejor en la fuente, es decir, por los empleados de ventas.
- La aceptación de los procesos de DQ se incrementó mediante un enfoque de circuito cerrado, en el que los empleados de ventas se beneficiaron directamente de sus propias mejoras en la DQ.
- Las mediciones de la DQ demostraron que era necesario adoptar medidas y documentaron las mejoras de la DQ.

2.4.6 Lecciones aprendidas

Hilti consideró que los siguientes aspectos eran particularmente importantes para el éxito del proyecto.

- Apoyo de la alta dirección de Hilti: los cambios necesarios en la estructura organizativa fueron posibles gracias a la dirección y a los clientes del proyecto.
- Definición temprana de las reglas de negocio: Hilti definió las reglas de negocio necesarias al principio del proyecto y se aseguró de que se compartieran.

- Definición de los objetivos de DQ de toda la empresa.
- Colaboración a través de las fronteras de la empresa: Hilti consideró muy positiva la intensa colaboración de todos los participantes en el proyecto, tanto entre los empleados a nivel local y la sede central de la empresa como entre los departamentos funcionales y el Departamento de Tecnología de la Información. Desde el principio, los requisitos para las nuevas herramientas se orientaron hacia las necesidades de los departamentos funcionales y los procesos empresariales asociados. Así pues, se incluyeron los usuarios de los datos, por ejemplo, mediante varios talleres intensivos durante todo el proyecto, con el fin de asegurar los niveles necesarios de motivación y de participación activa en el proyecto en los departamentos funcionales. Así, el departamento de alcance comercial evaluó la aplicación y el despliegue de las soluciones en toda la empresa de manera tan fluida como satisfactoria en general.

Para recapitular lo expuesto hasta el momento, cabe señalar que los casos de estudio presentan un tema en común, el monitoreo continuo en la DQ; lo cual se alinea con el objetivo principal de la tesis planteada “Automatización en el monitoreo y generación de alertas para calidad de datos”. Entre lo destacado de los casos, este tema corresponde en una parte al componente de monitoreo, el cual es esencial en cualquier escenario al que se enfrentan las organizaciones para un aseguramiento en DQ exitoso. Esto hace que cobre más relevancia en focalizar soluciones tecnológicas con mayor estrategia para brindar a las organizaciones un respaldo más amplio y acorde a los desafíos globales. En este orden de ideas, en la siguiente tabla se muestran los casos de negocio y el tema en común

(monitoreo) desde el enfoque de DQ en elementos comunes, metodología, tecnología y factores críticos de éxito (ver Tabla 4).

Tabla 4. Cuadro de criterios comunes casos de estudio

Fuente: creación propia

Monitoreo	Elementos comunes	Metodología	Tecnología	Factores Críticos de Éxito (FCE)
Casos de estudio				
Allianz	Diseño de medidas para la medición y el monitoreo de la DQ e implantación de una herramienta para el monitoreo continuo de DQ	Se usó el ciclo DMAIC, el cual es una herramienta metodológica enfocada en la mejora incremental de procesos existentes. Cada paso en la metodología se dirige a obtener los mejores resultados posibles para minimizar la posibilidad de error y es utilizado como plantilla para la gestión de procesos de DQ	Implementación de una herramienta de software cockpit para el monitoreo de DQ y generación de estadísticas en visualizadores para los diferentes niveles de usuario	-Relación con los proveedores -Imagen con los interesados -Reputación de solidez financiera
Bayer CropScience	Requisitos funcionales para la medición de DQ e implementación de una herramienta de control	Iniciación de un proyecto dedicado exclusivamente al tema de DQ con cuatro objetivos: crear conciencia empresarial, transparencia, reforzar la colaboración y, por último, identificar las necesidades para la toma de decisiones, por parte de los negocios, para mejorar la DQ	Con la implantación del software cockpit, se enfocó en las funcionalidades de medición y monitoreo, y en generación de gráficas de resultados para vistas específicas de usuarios por negocio	-Relación con los clientes -Imagen con los interesados -Líneas de productos y servicios -Credibilidad de la información

Johnson & Johnson	Fue importante reconocer que la gobernanza de datos tuvo un efecto positivo en la DQ, que fue claramente demostrado por la medición y el monitoreo continuo de la DQ	Se inició un proyecto en tres fases, con el fin de mejorar la DQ organizacionales: fase de establecimiento (maestro de datos empresariales EMD), fase de desarrollo (flujos de trabajos en la creación de datos un sistema de monitoreo de DQ) y fase de madurez (se evaluó un buen nivel de calidad que lo apropió la organización)	El software cockpit es una herramienta de análisis e informes para la DQ. Con ella se comprueba si los registros de datos maestros han roto las reglas definidas o violan la calidad de los requisitos. Además, muestran informes de gestión que proporcionaban indicadores claves de rendimiento para la DQ	-Relación con los clientes -Imagen con los interesados -Líneas de productos y servicios -Credibilidad de la información
Hilti	Las mediciones de la DQ demostraron que era necesario adoptar medidas y documentaron las mejoras de la DQ	Se inició un proyecto llamado Herramientas de calidad de datos de clientes. Para la gestión sostenible de la DQ, se determinó un modelo de cuatro medidas en el ciclo de vida de los datos del cliente: definir (responsabilidades de mantenimiento de datos), prevenir (adquisición de datos correcta), detectar (monitoreo de calidad de datos) y corregir (limpieza de datos); con el fin de tener un modelo para la gestión proactiva y reactiva de la DQ	Se desarrolló, por parte de TI, una herramienta ajustada a las necesidades para el monitoreo de DQ basada en los DQI	-Relación con los clientes -Imagen con los interesados -Líneas de productos y servicios -Credibilidad de la información -Mejoramiento de DQ de clientes

3 Desarrollo de la solución propuesta

3.1 Modelo de referencia

Para responder al objetivo general de este trabajo de grado, el cual consistía en “diseñar e implementar un mecanismo automatizado para monitorear y generar alertas, teniendo como base la adaptación de un modelo metodológico en aspectos relacionados con calidad de datos contenidos en fuentes de información de procesos de negocios”, se describió un diagrama BPMN del modelo de procesos en el numeral 1.3.2 del Marco Teórico, el cual corresponde a los procesos para la evaluación de calidad de datos de las organizaciones. Este sirvió como referencia para implementar la propuesta, dado que aborda la brecha existente entre los métodos genéricos (es decir, independiente del contexto) y un contexto específico.

Consecuente con lo anterior, las primeras actividades que se llevaron a cabo fueron las siguientes:

1. Definición del alcance: consiste en delimitar el contexto, los procesos de negocio, los datos y sus relaciones, los objetivos y los requisitos de los usuarios o clientes de datos. Los resultados de cada elemento servirán de base o insumos necesarios para la implementación de la solución.
2. Definición de las dimensiones y métricas: se parte de la premisa de que un modelo de proceso debe incluir un enfoque tanto ascendente (impulsado por el problema) como descendente (impulsado por la necesidad). Los problemas experimentados se identificaron mediante entrevistas semiestructuradas con los clientes/usuarios de datos, las cuales también permitieron identificar los objetivos de los consumidores de datos (con un enfoque de evaluación de arriba así abajo) donde se identifican las

necesidades en calidad de datos. La combinación de los resultados del enfoque descendente para identificar nuevas necesidades y del enfoque ascendente que parte de lo existente para establecer las mejoras en relación con calidad de datos, permitió crear el conjunto completo de dimensiones y métricas a partir de este levantamiento de información. Sobre la base de este conjunto se seleccionaron los objetos de datos (sistemas de información, tablas, atributos, historia, etc.) para hacer las mediciones.

3. Realizar la medición: este componente se desarrolló a través del RPA descrito en el numeral 4.2.
4. Análisis y presentación de informes: al igual que el anterior, este componente forma parte de la solución RPA descrito en el numeral 4.2.

En el numeral 1.3.2 se muestra el modelo de proceso de referencia, el cual se toma como línea base en la implantación de la solución.

3.2 Arquitectura de la solución

En este capítulo se presenta la arquitectura implementada para soportar la solución de este proyecto; los componentes y la interacción que tienen entre sí. Para describir la arquitectura se utilizaron las vistas lógicas y funcional, con el objetivo de tener la visualización de las capas abstractas y los componentes funcionales.

3.2.1 Vista lógica

Esta arquitectura, definida para esta solución, es de tipo cliente/servidor y se distribuye en tres capas: cliente, negocio y datos DQ que se explican a continuación:

- Cliente: en esta capa se encuentran dos componentes del tipo cliente. El primero, es un RPA que trabaja en modo desatendido, y obedece a una secuencia de

actividades definidas en unos scripts por medio del uso de web API tipo REST para la conexión a la capa de negocio mediante el consumo de sus servicios. El otro es un formulario de consulta de información en Excel que genera indicadores de forma gráfica, y se conecta a la capa de negocios mediante el consumo de servicios vía REST.

- **Negocio:** en esta capa se construye la lógica requerida por el RPA y es manejada por una federación de servicios web que se orquestan entre ellos, los cuales obedecen a unos patrones de Arquitectura Orientada a Servicios (SOA). Estos acceden a la capa de datos mediante el uso de un componente DAO (Entity Framework), por lo tanto, no se necesita usar un SQL directamente. Todo se hace a través de un modelo abstraído de los modelos de datos para monitoreo.
- **Datos DQ:** esta capa gestiona los datos y las operaciones que son enviados por la capa de negocios, los cuales garantizan la persistencia de la información, las consultas y las actualizaciones. La solución está compuesta por tres bases de datos: Configuración DQ, Bitácora DQ y BRM DQ.

En la Ilustración 18 se diagraman los componentes:

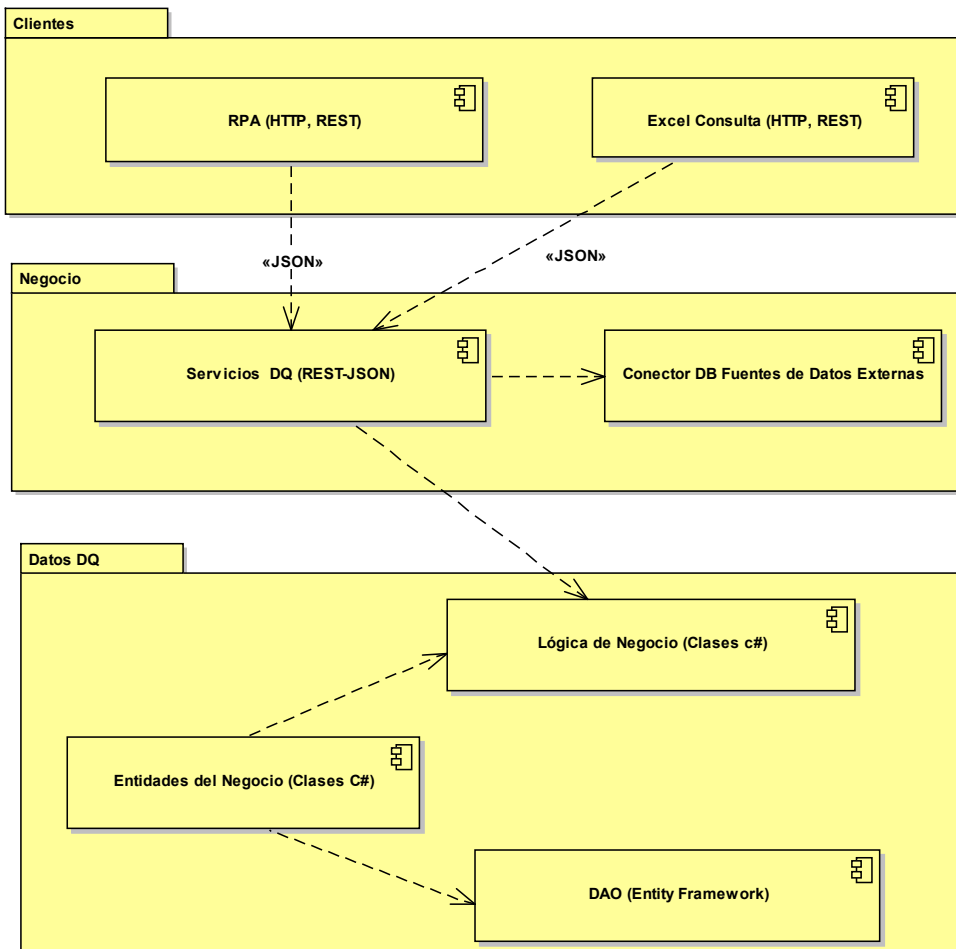


Ilustración 18. Vista lógica

Fuente: elaboración propia

3.2.2 Vista funcional

En esta vista (ver Ilustración 19) se muestran todos los componentes funcionales de la solución propuesta, los cuales cubren la automatización del proceso apropiado como referencia implementado mediante un RPA.

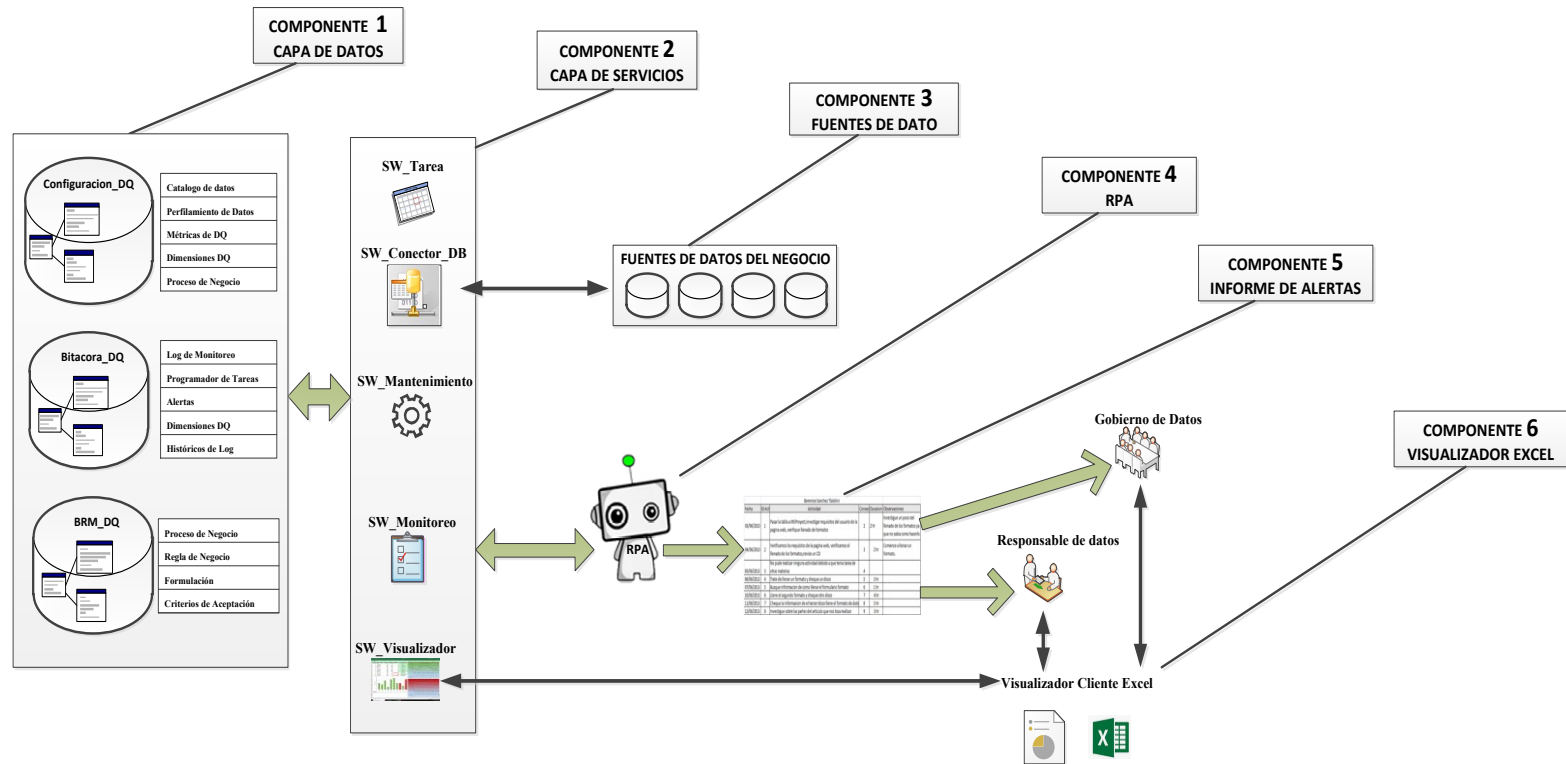


Ilustración 19. Vista funcional de los componentes

Fuente: elaboración propia

3.2.2.1 *Componente 1: capa de datos*

La capa de datos soporta la información recopilada durante las actividades del proceso de referencia que incluye el alcance, las dimensiones, los atributos de calidad, las métricas, las formulaciones, el perfilamiento de los datos y los roles de usuarios. Asimismo, en este componente se describen los modelos de datos para el monitoreo y la generación de alertas en calidad de datos del proceso de negocio y de la programación de tareas para la vigilancia continua mediante un RPA. Ambos modelos se describen a continuación:

3.2.2.2 *Base de datos de configuración DQ*

En esta base de datos se almacena la lógica correspondiente a la estructuración de los procesos de negocio, las dimensiones con sus factores de medición y el perfilamiento de los datos. Este modelo de datos garantiza a la capa de servicios su uso para que el RPA pueda monitorear la calidad de los datos al aplicar las diferentes métricas establecidas por los negocios organizacionales. En la Ilustración 20 se muestra el modelo en mención y se describen las tablas que conforman la base de datos de la solución propuesta:

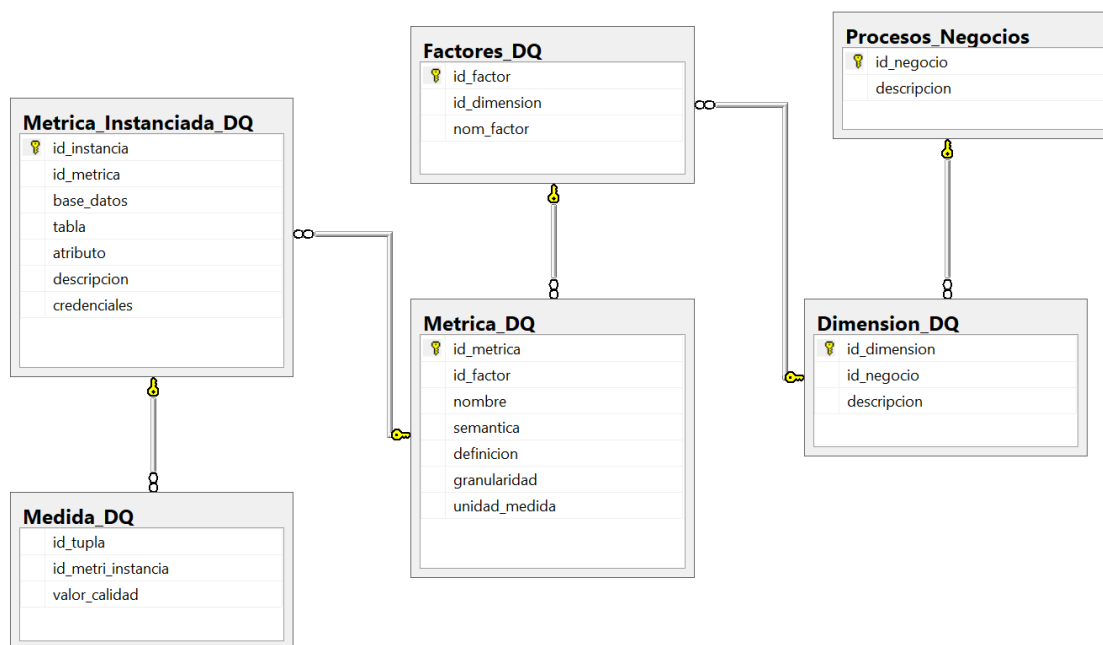


Ilustración 20. Modelo base de datos Configuración_DQ

Fuente: elaboración propia

- **Procesos Negocio DQ:** en esta entidad de datos se reúne la identificación y el nombre del proceso de negocio dentro de la organización, el cual es soportado a nivel digital.
- **Dimensión DQ:** en esta entidad de datos se almacenan las dimensiones identificadas en el proceso de negocio producto del levantamiento previo de las necesidades o mejoras en la calidad de datos. De esa manera se pueden crear dimensiones de forma dinámica, en lugar de utilizar conjuntos predefinidos, para que sean flexibles y estén alineadas con el contexto de las necesidades de los procesos de negocios. Las dimensiones se seleccionan de acuerdo con la funcionalidad de los datos, lo cual requiere tener claridad en los objetivos y la expectativa de los consumidores de estos datos. En este componente se presentan las dimensiones a almacenar más relevantes, las cuales constituyen el foco de numerosas investigaciones sobre la calidad de datos (Scannapieco y Catarci, 2002) y el eje central para el estudio propuesto en el monitoreo de calidad de datos; estas son:

- Precisión
- Completitud
- Relacionada con el tiempo
- Consistencia
- **Factores DQ:** en esta entidad de datos se almacenan los factores correspondientes a cada dimensión, los cuales deben ser medidos e identificados en el proceso de levantamiento de la situación actual de acuerdo con la problemática que se detectó en la calidad de datos.
- **Métricas DQ:** en esta entidad se almacenan las métricas de los factores establecidos que se van a medir, y se define el detalle de cada una, por ejemplo, si la medida está por fuera del valor del rango, tiene registros inexistentes o valores nulos, entre otros.
- **Métrica instanciada DQ:** en esta entidad de datos se almacena un catálogo del factor que va a ser evaluado, cuya función debe corresponder con el nombre de la base de datos y de la tabla, las credenciales y los atributos que identifican al dato dentro de las fuentes de datos que serán monitoreadas, así como también con los aspectos relacionados con accesos a las fuentes de datos.
- **Medida DQ:** en esta entidad se almacenan los registros monitoreados con el resultado de la métrica obtenida y definida en el proceso de negocio.

Base de datos Bitácora DQ: en este modelo de datos se almacenan los logs generados por cada tarea programada para el monitoreo de la calidad de datos y sus alertas, según las definiciones establecidas en las reglas de negocio (en función de las métricas preestablecidas en la base de datos de configuración). El RPA tomará el programador de tareas para ejecutar las definiciones del manejo de alertas para notificar a los interesados las inconsistencias presentadas, almacenar los logs de la ejecución de cada monitoreo realizado y almacenar un histórico

detallado de todas las transacciones realizadas para efectos del monitoreo; las siguientes entidades de datos integran la base de datos que se muestra en la Ilustración 21:



Ilustración 21. Modelo base de datos Bitácora_DQ

Fuente: elaboración propia

- **Log DQ**: aquí se almacenan las transacciones de monitoreo de calidad de datos, el perfilamiento de datos, el proceso de negocio, las métricas aplicadas, las dimensiones, los resultados, los estados de alertas y las fechas con horas. Además, para obtener una parte de la información se establece un mecanismo de servidor vinculado para traer información

con la base de datos BRM DQ, con el fin de completar los datos que alimentan los logs de información.

- **Programador tareas DQ:** en esta entidad de datos se configura todo el plan de tareas como fechas, horas y procesos de negocio. Este programador de tareas es usado por un servicio que es consumido por el RPA para lanzar todos los procesos de monitoreo de calidad de datos y generación alertas. Es importante señalar que esta tabla utiliza un servidor vinculado con la base de datos BRM DQ para completar información requerida en la ejecución del monitoreo.
- **Alertas DQ:** en esta base de datos se fundan todas las alertas que deben generarse para enviárselas a los interesados, como son los procesos de negocios y el gobierno. De tal forma que el RPA consulte la información para el envío de alertas y la visualización de los informes.
- **Históricos logs DQ:** en esta entidad de datos se reúnen todos los logs de monitoreo; los datos del log son las descripciones definidas en cada relación con otras tablas de la configuración del modelo de datos, por esto no se requieren de relaciones o consultas complejas. El objetivo es realizar consultas y extraer la información únicamente de esta tabla en modo de contenido, para obtener mayor eficiencia en la extracción del servicio de visualización de datos.

Base de datos BRM DQ: gestiona las reglas del negocio y es la responsable de integrar los procesos de negocio con la base de datos de un servidor vinculado, el cual es el de Configuración DQ; esto con el fin de completar la lógica requerida por el RPA y suplir las reglas definidas por el negocio con sus criterios de aceptación correspondientes, basado en la formulación del proceso de negocio (ver Ilustración 22).

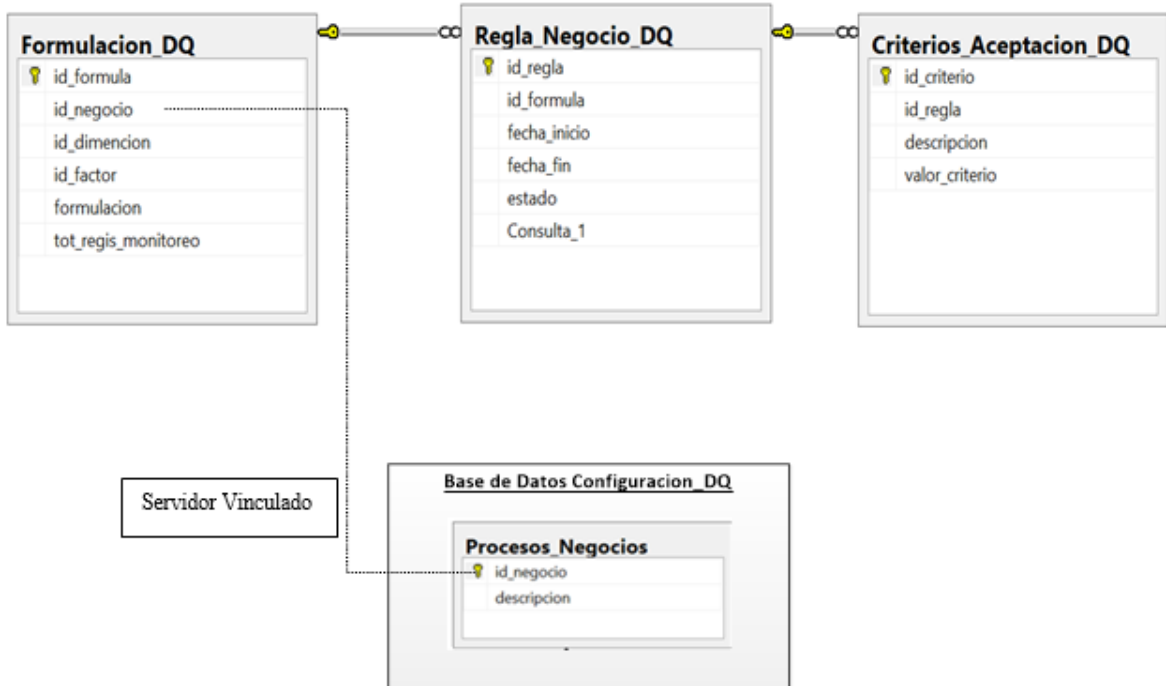


Ilustración 22. Modelo base de datos BRM_DQ

Fuente: elaboración propia

- **Regla negocio DQ:** en esta entidad se acumula la definición de la regla de negocio, la cual contiene la fecha de inicio y finalización establecida por el negocio para su ejecución; asimismo, un campo para establecer la estructura de consulta y un estado para activar o desactivar la regla de negocio, con el fin de asegurarle al RPA las condiciones lógicas para la ejecución del monitoreo en calidad de datos y generación de alertas.
- **Formulación DQ:** en esta entidad de datos se concentrarán todas las formulaciones que corresponden a una Regla de Negocio DQ y así obtener los resultados de los indicadores previamente establecidos con consulta al servidor vinculado (Configuración DQ). El RPA aplica las fórmulas y establece las lógicas para su ejecución. Cabe señalar que esta entidad de datos almacenará las formulaciones de negocios y definirá la cantidad de registros a ser tomados en cuenta para el monitoreo de calidad de datos; las cuales se definen a nivel de dato,

atributo (columna), registro (fila) o relación (tabla o conjunto resultado de una consulta). Las dimensiones establecidas para este caso de estudio son:

- **Dimensión de precisión** tendrá las siguientes mediciones:

- **Precisión de una relación** (tabla o conjunto resultado de una consulta):

$$\% PR_i = (1 - (NTI / NTR)) * 100$$

NTI = Número de tuplas (registros) en la relación que tienen uno o más valores incorrectos.

NTR = Total de tuplas (registros) sobre la relación.

R_i = Relación en la cual se aplica la medición. Esta relación puede ser una tabla o el conjunto de resultados que se obtiene de una consulta.

- **Precisión de un atributo** (columna):

$$\% PA_i = (1 - (NCI / NTC)) * 100$$

NCI = Número de campos incorrectos sobre el atributo.

NTC = Total de campos sobre el atributo.

A_i = Atributo sobre el cual se aplica la medición.

- **Precisión de un registro** (fila)

$$\% PF_i = (1 - (NCI / NTC)) * 100$$

NCI = Número de campos incorrectos sobre la fila. NTC = Total de campos sobre la fila.

F_i = Fila sobre la cual se aplica la medición.

- **Dimensión de completitud de un valor dato:** Esta medida establece la existencia de un dato definido por un criterio (*null*, vacío, errado, entre otros).

- **Completitud por relación** (tabla o conjunto resultado de una consulta):

$$\% CR_i = (1 - (NRI / TRR)) * 100.$$

NRI = Número de registros incompletos.

TRR = Total de registros de la relación.

R_i = Relación evaluada.

▪ **Precisión de un atributo** (columna):

$$\% CA_i = (1 - (NVI / NTV)) * 100$$

NVI = Número de valores incompletos.

NTV = Número total de valores en el atributo.

A_i = Atributo evaluado.

▪ **Precisión de un registro** (fila)

$$\% CF_i = (1 - (NAI / NTA)) * 100$$

NAI = Número de atributos incompletos.

NTA = Número total de atributos de la fila.

F_i = Fila evaluada.

○ **Dimensión relacionada con el tiempo:** cambio en datos en el ciclo de vida de la información en el tiempo.

- Rapidez: la velocidad en que los datos son actualizados en el tiempo.

$$\text{Rapidez} = \text{Age} + (\text{Delivery Time} - \text{Input Time})$$

Age mide la antigüedad de la unidad de datos cuando se recibe. Por otro lado, *Delivery Time* es el momento cuando el dato se entrega al cliente, e *Input Time* es el momento en el cual la unidad de datos se obtiene.

- Volatilidad: frecuencia en la cual cambia en el tiempo. La volatilidad se define como la longitud en tiempo para el cual el dato sigue siendo válido.

- Oportunidad: disponibilidad en un momento determinado para cumplir con una tarea específica.

$$\text{máx} \{0, 1 - (\text{rapidez} / \text{volatilidad})\}$$

Los rangos de oportunidad son 0 y 1. Se debe señalar que 0 significa mala oportunidad y 1 significa buena oportunidad.

Criterios de evaluación:

C1: 0 - 1 día (bueno).

C2: 1 - 3 días (aceptable).

C3: 3 - 5 días (regular).

C4: Más de 5 días (malo).

$ActR_i = (\text{fecha de entrega al usuario} - \text{fecha de ingreso al sistema})$

$R_i = \text{Registro entregado al usuario.}$

$\% Act (C_i) = (NRC / NTR) * 100$

$NRC = \text{Número de registros cumplen } C_i.$

$NTR = \text{Número total de registros.}$

$C_i = \text{Criterio de evaluación.}$

○ **Dimensión de consistencia**

○ Restricciones de integridad: los manejadores de las bases de datos relacionales gestionan la integridad referencial que se definen a nivel de un esquema estructurado y de sus dependencias, como son: i) dependencias de claves principales mediante las cuales se definen valores únicos en un dominio de datos; ii) dependencias de inclusión que corresponden a la llaves foráneas que hacen parte integral de una llave principal que identifica una tupla como única en la relación; y iii) dependencias funcionales a través de un campo de una columna que permite proyectar un valor en función de una regla, por ejemplo, si se conoce el valor de la fecha de nacimiento de una persona se podrá determinar el valor de su edad, lo cual asegura una consistencia de los datos. Con esto como referencia, los manejadores de bases de datos no alcanzan a cubrir los valores que no forman parte de las dependencias relacionales; a su vez, esta se cubre por los *Data Edits*.

○ *Data Edits*: es definido como la tarea de identificar inconsistencias basadas en la formulación de reglas (*Edits*), la regla de inconsistencia en forma de un *Edits* sería:

$\text{Estado Civil} = \text{casado} \wedge \text{edad} < 14$

La métrica para soportar esta inconsistencia es:

$C_1 = \text{Criterio 1 de consistencia para el registro.}$

$C_2 = \text{Criterio 2 de consistencia para el registro.}$

$C_n = \text{Criterio n de consistencia para el registro.}$

$$\text{CoCiRj} = (\text{NRC} / \text{NTR}) * 100$$

NRC = Número de registros que cumplen Ci.

NTR = Número total de registros Rj.

Ci = Criterio de consistencia evaluado.

Rj = Relación sobre la cual se aplica la validación. Esta relación puede ser una tabla o el conjunto de resultados que se obtiene de una consulta.

- **Criterios Aceptación DQ:** en esta entidad de datos se configuran los criterios de aceptación definidos por los procesos de negocios. Aquí se define el valor del criterio de aceptación para que el RPA, dentro de su ejecución, se identifica y válida para establecer si está por fuera del rango de aceptación, y así genere la alerta correspondiente.

3.2.2.3 Componente 2: capa de servicios web

La capa de servicios está en un estilo de arquitectura orientada a los servicios SOA, para que esta funcione como una capa intermedia entre el RPA y la capa de datos que soporta la lógica de las reglas para el monitoreo, la generación de alertas y la visualización del histórico de la información monitoreada. En esta capa se coordinan los servicios para lograr, en tiempo de ejecución, que el RPA construya la lógica para efectuar el monitoreo continuo en la calidad de datos basado en la configuración en la que fueron establecidos para cada proceso de negocio. Lo anterior se plantea con el objetivo de alinear la tecnología con el negocio y obtener la generación de alertas según las reglas definidas e implementadas por el servicio (WS Monitoreo); además de cumplir con los objetivos de SOA como, por ejemplo: agilidad organizacional para la toma de decisiones y ayudar a que los negocios puedan responder, en el menor tiempo posible, ante situaciones donde la calidad de los datos presenta inconsistencias.

Otro de los objetivos consiste en reducir la carga de TI, mediante el servicio “WS Visualizador ()”. Ello se da por medio de sus capacidades, “Envío Alertas ()”, la cual notifica a los interesados de los eventos que presentan problemas de calidad de datos del sistema de información, y lo somete a monitoreo con las definiciones de la lógica del negocio establecidas, dándole una oportunidad al área de TI para realizar las acciones correctivas del caso. El servicio de (SW Tareas) es fundamental para que el RPA programe sus tareas de monitoreo en las frecuencias de tiempo permitidas, y se coordine con el servicio de (WS Conector DB) para establecer una instancia con los repositorios de las fuentes de datos a ser monitoreadas. En la Ilustración 23 se muestra la arquitectura de servicios y, luego, se especifica cada uno de ellos en su arquitectura SOA.

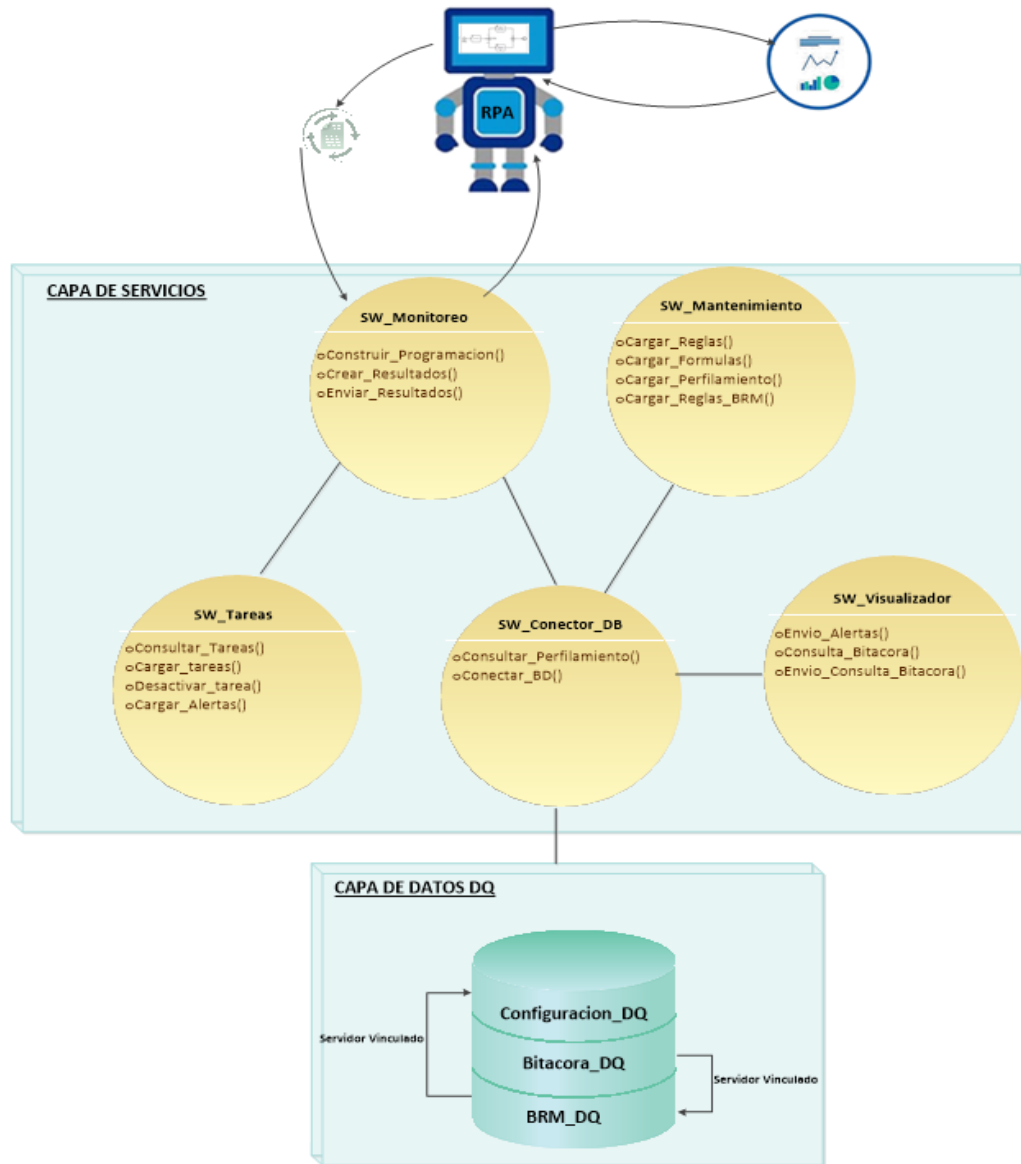


Ilustración 23. Capa de servicios web

Fuente: elaboración propia

En su definición de arquitectura de servicios SOA para el monitoreo mediante el RPA, Ear.T. (2009) planteó los siguientes patrones de arquitectura y sus principios para el abordaje de esta solución:

- **Patrón de centralización de reglas:** las reglas de negocio pueden ser abstraídas físicamente en una parte dedicada de la arquitectura, bajo la gestión de motores y

plataformas de reglas especializadas. Esto centraliza el acceso a la lógica de las reglas de negocio y evita la redundancia. Además, centraliza el gobierno de las reglas de negocio para que puedan ser modificadas y evolucionadas desde un único lugar, para este caso, desde la capa de datos.

- **Principio reutilización de servicios:** los servicios contienen y expresan una lógica agnóstica y pueden posicionarse como recursos empresariales reutilizables.
- **Patrón de capacidad agnóstica:** este patrón posiciona esencialmente cada capacidad como una función independiente, capaz de resolver una preocupación común a múltiples procesos o tareas comerciales. Las capacidades agnósticas bien definidas se encuentran en el centro de los principios fundamentales de la orientación a los servicios, como la reutilización de los servicios y la componibilidad de los servicios.
 - **Principio de contrato de servicio estandarizado:** los servicios, dentro del mismo inventario de servicios, cumplen con las mismas normas de diseño de contratos con un conjunto de capacidades bien definidas.
 - **Principio de servicios de reutilización:** los servicios contienen y expresan una lógica agnóstica y pueden posicionarse como recursos empresariales reutilizables por otros servicios.
 - **Principio de servicio de compuestabilidad:** los servicios son participantes efectivos en la composición, independientemente del tamaño y la complejidad de esta.

- **Patrón de recursos canónicos:** los servicios utilizan el mismo recurso de infraestructura estandarizado para el mismo propósito.
 - **Principio de autonomía de servicio:** los servicios tienen un alto grado de control y lógica en su ejecución.

Los siguientes son los servicios diseñados para la solución propuesta:

SW Monitoreo: es el servicio responsable de ofrecer unas capacidades para la consulta del programador de actividades, mediante la capacidad “Construir Programador ()”. Del mismo modo, genera la lógica de negocio para que el RPA la ejecute. A través de las capacidades de “Crear Resultados ()” y “Enviar Resultados ()” se procesa y persiste la información resultante del monitoreo en la capa de datos para su posterior consulta, configuradas para el monitoreo de calidad de datos y generación de alertas; este servicio se basa en la capa de datos y toma su información de la entidad de datos (Bitácora DQ).

WS Tareas: es el servicio responsable de permitir que el RPA lo consuma para establecer el orden de las tareas y su ejecución, mediante la capacidad “Consultar Tareas ()” y, a su vez, para desactivarlas con “Desactivar Tareas ()”. De tal manera que el RPA tenga claro los procesos de negocio, reglas de métricas, perfilamiento de datos, catálogos de datos, accesos y el BRM en sus procesos de monitoreo de calidad de datos y generación de alertas. Es fundamental señalar que, de igual forma, permite el cargue de tareas y alertas mediante archivos tipo CVS con las capacidades “Cargar Tareas ()” y “Cargar Alertas ()” con el fin de mantener esta información en la capa de datos.

SW Visualizador: es el servicio responsable de la consulta histórica del monitoreo en calidad de datos para que sea visualizado el comportamiento de los diferentes procesos de negocios en su aseguramiento e indicadores, los cuales permitan un apoyo a la organización en la

toma de decisiones; este es consumido por el RPA en sus capacidades “Envío Alertas ()”, “Consulta Bitácora ()” y “Envío Consulta Bitácora ()”.

SW Mantenimiento: es el servicio responsable de la creación, actualización, configuración y parametrización de la lógica que soportará el RPA para garantizar su ejecución y funcionalidad, con el fin de cumplir con las expectativas del negocio la cual se soportará por la capa de datos (Base de datos de Configuración DQ, Bitácora DQ y BRM DQ) y así garantizar la información a las capacidades del servicio “Carga Reglas()”, “Carga Formulas()”, “Carga Perfilamiento()” y “Carga Reglas BRM()”. Como se mencionó, la idea es dar cumplimiento exitoso en el monitoreo continuo en la calidad de datos y generación de alertas; esta carga se realiza mediante el uso de archivos CVS.

SW Conector BD: es un servicio de utilidad, el cual organiza los accesos de los demás servicios que lo consuman para asegurar la conexión e inicio de sesión a las diferentes fuentes de datos establecidas por el catálogo de datos mediante las capacidades “Consulta Perfilamiento ()” y “Conectar DB ()”.

- **Capa de datos:** Descrita anteriormente en el numeral 4.2.2.1.
- **RPA:** este componente es un consumidor de la capa de servicios; se encarga de entregar resultados de inconsistencias en calidad de datos a los actores de los procesos de negocios interesados y al gobierno de datos.

3.2.2.4 Componente 3: fuente de datos

Las fuentes de datos son modelos de información, en los cuales se realizan tareas de extracción de registros de los procesos de negocios para su seguimiento, análisis y generación de resultados.

La extracción de los datos se realiza al tener como base la configuración definida en los modelos de datos (BRM DQ) y (Configuración DQ), las cuales proporcionan el nombre de la fuente de datos, el nombre de la tabla, nombres de campos y un rango de fechas. A partir de esta información, el RPA procede a realizar la consulta para obtener las colecciones de datos almacenados para su monitorización, ejecución de validaciones, reglas, formulaciones, métricas y criterios de aceptación de los procesos de negocios.

3.2.2.5 Componente 4: RPA

Este componente automatiza procesos robóticos y está programado paso a paso; su clasificación es clase 1 por su rápido diseño y pruebas fáciles. El tipo de proceso a implementar es específico, pues es repetitivo por sus características de monitoreo continuo en temas relacionados con calidad de datos. En su programación sigue la lógica que se plantea en la capa de servicios, consume y ejecuta las tareas definidas en las diferentes capacidades expuestas y, por último, como resultado, envía informes de alertas.

La herramienta en la cual se desarrolló el RPA es UiPath, líder en tecnologías de automatización a nivel mundial con una licencia *community*, la cual es de uso libre. Alguna de las características a resaltar es que se integra muy fácil con Microsoft Office Excel, lo cual favorece en dar alcance a la solución de generación de alertas; además, admite invocación de servicios web mediante *scripts* y *requests* HTTP al usar REST en formato JSON.

Pasos en la implantación de un RPA

Identificar proceso a robotizar: es realizar una minería de procesos para determinar cuáles son buenos candidatos para automatización. En esta parte se trabaja de la mano del BPMN y se determina qué procesos serían adecuados para un pilotaje.

Construir su caso de negocio: establecer la automatización que respalda las necesidades del negocio, sus beneficios, los puntos de dolor que se aliviarán, las métricas que identifican que la automatización es valiosa y establecer la estrategia para volver a implementar los recursos existentes después de la automatización.

Determinar el modelo de operación óptimo: identificar el modelo operativo funcional que mejor se identifique para la organización y que brinde un valor agregado. Contar con el equipo adecuado para respaldar la solución y llevar a cabo las responsabilidades para evaluar nuevos procesos a ser automatizados y probar el trabajo automatizado que se despliegue en ambiente de pruebas. Adicional a ello, establecer por parte de TI el rol responsable de administrar el robot.

Identificación de socio de automatización: en especial, los principales proveedores en tecnologías RPA que se adaptan a las necesidades del negocio y además comparar los modelos de precios para tomar la decisión más sólida y de mayor respaldo.

Planificar la hoja de ruta de automatización: identificar de cuánto tiempo debe ser el pilotaje, determinar las etapas después de finalizar el pilotaje, tener claro cuál es la estrategia de escalamiento y garantizar a las partes interesadas que son afectadas por el proceso de automatización los beneficios y cómo va a ayudarlos en sus procesos de negocios.

Para llevar a cabo la implantación de la solución del proyecto de monitorización y generación de alertas de calidad de datos, se tomaron varios elementos que sirvieron como desarrollo de un componente RPA para la solución del proyecto de investigación; entre estos tenemos:

1. Clasificación de procesos a ser automatizados: para esta solución se catalogó como clase 3 por su característica de utilizar de manera intensiva datos estructurados.

2. Definición de la tipología de proceso a intervenir: para este caso fue un proceso específico por sus características de datos transaccionales, los cuales son sencillos y repetitivos como lo es el monitoreo.
3. La metodología utilizada para su implementación y puesta en ejecución fue la de IBM (2018).
4. Se adoptaron los pasos de implementación planteada en este capítulo: pasos en la implementación de un RPA.

En la Ilustración 24 del flujo de trabajo, para el RPA se identifican cinco actividades que corresponden al manejo de la lógica del RPA para el monitoreo:

- Solicitud SOA programador tareas: esta actividad consume el servicio de (SW Tareas) con el objetivo de que el RPA inicie las tareas de ejecución que tiene asignadas y pendientes; se sincronizan con el reloj del sistema y ejecuta toda la lógica de la programación de tareas.
- Solicitud SOA monitoreo de datos: el RPA consume el servicio (SW Monitoreo) y junto con las tareas y consulta extrae la lógica definida por los procesos de negocio, dimensiones, métricas, formulaciones y reglas de negocio para iniciar el monitoreo. Además, envía los parámetros necesarios para instanciar las fuentes de datos y extraer la información requerida.
- Solicitud SOA generación de alertas: consume el servicio SW Tareas; su objetivo es generar las alertas según el caso de criticidad encontrados en el monitoreo de calidad de datos en los diferentes procesos de negocios. Así mismo, envía un archivo de toda la información requerida para construir los informes de alertas y sus destinatarios.

- Generar Excel y enviar correo: esta actividad consume el servicio de (SW Tareas) para identificar la información de la tarea y dar completitud a la ejecución de esta con la generación del archivo Excel; lo adjunta en el correo de envío a los destinatarios interesados en la supervisión de los datos que son monitoreados.
- Actualización de la base de datos (Bitácora DQ) con la persistencia de los datos de resultado de la aplicación de las diferentes reglas de monitoreo; así mismo, actualiza la base de datos de tareas y da por cumplido la ejecución de las diferentes tareas programadas en el ciclo de tiempo y fecha de ejecución.

El flujo que se muestra en la Ilustración 24 es el implementado en la herramienta de UiPath con el uso de los diferentes componentes que ofrece para la ejecución paso a paso del RPA. Todo con el fin de cubrir la solución propuesta en este estudio y cumplir su alcance.

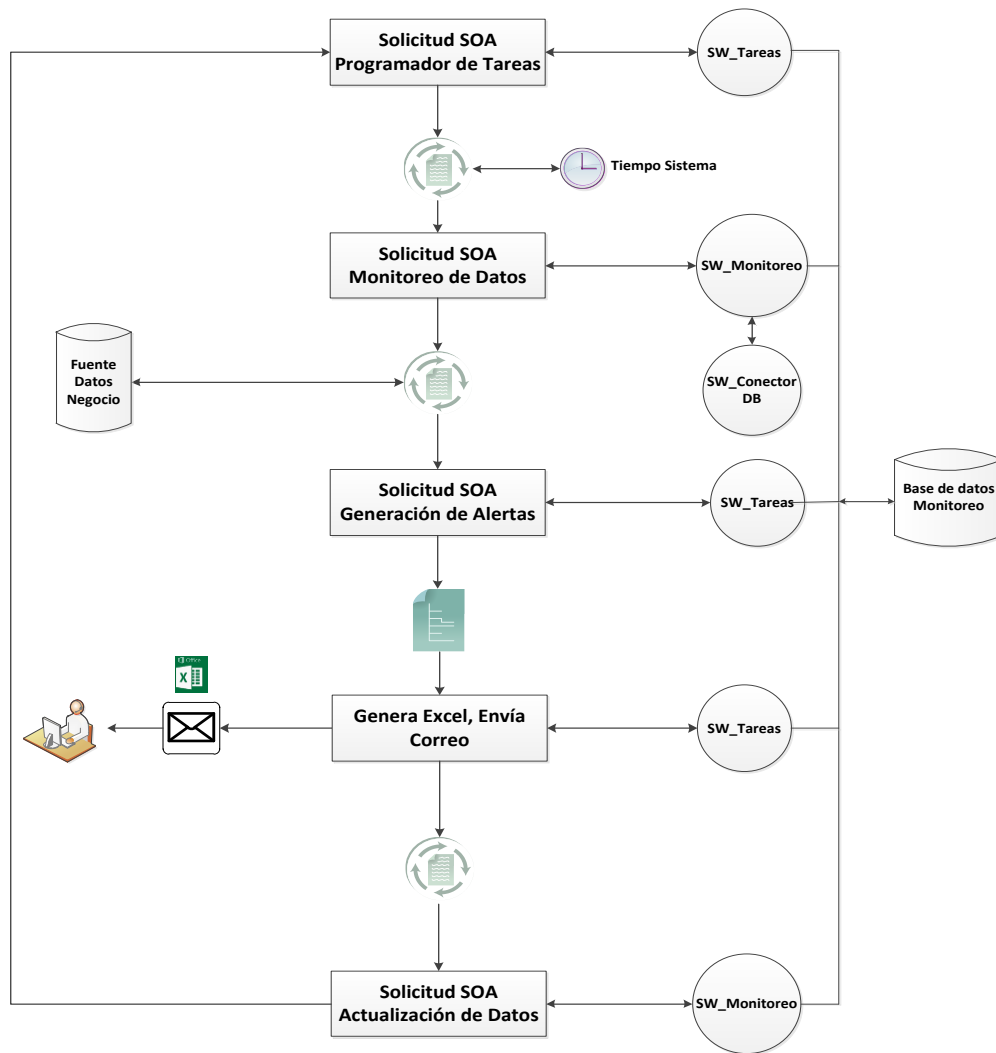


Ilustración 24. Flujo de trabajo del componente RPA

Fuente: elaboración propia

3.2.2.6 *Componente 5: informe de alertas*

El RPA genera un reporte de alertas a los procesos de negocio interesados y gobiernos de datos con sus correspondientes indicadores, según dimensiones de métricas establecidas en la base de datos de Bitácora DQ, que, a su vez, serán enviadas a través de correos electrónicos en un formato Excel. Este entregará los indicadores que presente un porcentaje de inconstancias y

que estén por fuera de los criterios de aceptación de los procesos de negocio, los cuales se encuentran debidamente parametrizados en la base de datos BRM DQ.

Alerta de Calidad de Datos										
Ciclo :										
Fecha de procesos:										
Proceso de Negocio:										
Dimensión de Calidad	Atributo de Calidad	Campo Afectado	Métrica de Calidad	Descripción	Numero Total Registros NR	Numero Registros Errados NRE	Indicador	Sistemas de Información	Formula	Criterio de Aceptación CA

Ilustración 25. Resumen dimensiones de calidad de datos.

En la Ilustración 25 se exponen los datos que contiene el formato de alertas enviados; consiste en un encabezado con la información, compuesta por el ciclo que corresponde al intervalo de tiempo en un formato (HH:MM) de inicio y finalización en el que se ejecuta esta tarea de monitoreo; fecha de proceso que corresponde a la fecha de ejecución de esta actividad; proceso de negocio es el nombre que se encuentra configurado en el modelo de datos de monitoreo que corresponde a la organización.

A nivel de detalle, el formato presenta la información de la alerta generada; las columnas dimensión de calidad y atributos de calidad contienen información relacionada en la Ilustración 25. En la columna (campo afectado) se especifica el campo de negocio en el cual se realiza el monitoreo; la columna (métrica de calidad) detalla la métrica de calidad a ser evaluada; la columna (descripción) hace un breve resumen de la regla de negocio que se aplica a este atributo de calidad; la columna (Número total de Registros (NR)) relaciona el número de registro evaluado en el ciclo de tiempo establecido; la columna (Número de Registro Errados (NRE)) notifica la cantidad de registros hallados con error en la validación de la regla de negocio a ser medida; la columna (indicador) muestra el porcentaje de error, según la fórmula establecida para el dato evaluado; la columna (sistema de información) es el nombre del sistema de información que soporta el dato que es monitoreado; la columna (fórmula) describe la formulación para la

generación del indicador de error y, por último, la columna (criterio de aceptación) señala el porcentaje de tolerancia de error aceptada por el proceso de negocio. Con esta información, los interesados en el proceso de negocio y el gobierno de datos pueden tomar decisiones con las inconsistencias generadas por el monitoreo continuo en todos los temas relacionados con calidad de datos.

3.2.2.7 Componente de visualización: Excel

Este componente es un archivo Excel mediante el cual se consume el servicio de visualización, en él se despliegan las gráficas y los datos según se defina su consulta. Está compuesto por seis campos (fecha de inicio, fecha de finalización, proceso de negocio, dimensión, dato y atributo de calidad) y un botón para ejecutar la consulta de calidad de datos. Los campos obligatorios para realizar la consulta son las fechas que indican el rango de búsqueda, el campo de (proceso de negocio) lista los diferentes procesos de negocios afectados por inconsistencias (se debe seleccionar uno), y en el campo (dimensión) lista las dimensiones correspondientes al proceso de negocio seleccionado anteriormente. Igualmente, se puede seleccionar el nombre del campo del dato y el atributo de calidad como se muestra en la Ilustración 26.

Rango de fechas para consultar			
Fecha de Inicio	1/02/2020	Fecha de Finalizacion	31/03/2020
Seleccione el proceso de negocio			
Proceso de Negocio :	Pedidos de Ventas	Las Fecha , Proceso de Negocio y Dimension son Obligatorias para realaizar la Consulta	
Dimensión :	Precisión		
Dato:	Direccion Envio		
Atributo de Calidad :	Sintáctica		
<input type="button" value="Consulta Calidad de Datos"/>			




Ilustración 26. Componente de visualización en Excel

Fuente: elaboración propia

El servicio (SW Visualizador) extrae información de la base de datos (Bitácora DQ) y presenta gráficas en un libro de Excel que contiene tres hojas: la primera, corresponde a la trazabilidad del proceso de negocio, la dimensión y genera la gráfica “traza dimensiones”; paralelamente, “traza del dato errado” presenta los atributos de calidad en el rango de tiempo establecidos en la consulta; y “traza x ciclos” muestra los atributos de calidad en el rango de tiempo por los ciclos de ejecución realizados por el RPA.

4 Aplicación al caso de estudio

El caso de estudio se realizó con el objetivo de comprobar la conformidad de la solución propuesta en este proyecto de maestría para el monitoreo y la generación de alertas en los temas relacionados con calidad de datos organizacionales. Así pues, con esta investigación se pretendió:

- Definir el contexto del proceso de negocio con los datos evaluados, así como los tipos, relaciones, consumidores y productores de información.
- Definir el método de medición de las dimensiones, e identificar la manera de medir las características de calidad relevantes al mapear el catálogo de datos que será medido.
- Activar el RPA desarrollado para la medición, basado en las formulaciones de cumplimiento de las reglas de los procesos de negocios.
- Analizar y generar los resultados que no cumplen con los criterios de aceptación de los negocios.

4.1 Descripción del caso de estudio

La aplicación se realizó para los datos recolectados en una empresa farmacéutica, ubicada en la ciudad de Barranquilla (Colombia); la cual cuenta con 30 años de experiencia en el mercado y tiene vinculados cerca de 1300 empleados. La empresa es líder en la fabricación de cápsulas duras de grado farmacéutico de alta calidad, es decir, elabora productos farmacéuticos orales (cápsulas, tabletas, suspensiones, etc.) a precios competitivos.

Actualmente, el área de Tecnología de Información (TI) de la empresa farmacéutica debe mitigar la inconsistencia presente en la calidad de los datos. Este requerimiento nace específicamente del área comercial, dado que el proceso de negocio de pedidos de ventas ha

presentado errores en los datos, y algunas veces requieren reprocesos por el grado de dificultad. Este proceso de negocio es altamente complejo porque se compone de varios artefactos de integración, con muchas dependencias funcionales y reglas de negocio; además, los errores en los ciclos de vida de los datos repercuten en otros sistemas de información transaccionales y en procesos de toma de decisiones como los inventarios, los costos y el área Comercial.

4.2 Aplicación de la metodología

4.2.1 Definición del alcance

Para lograr los objetivos del caso de estudio se realizaron reuniones con el gerente de TI y el equipo de tecnología de la empresa farmacéutica, durante las cuales se socializó el proyecto de investigación y se definieron las tareas que se realizarían con sus respectivos responsables. Lo anterior con el fin de coordinar el levantamiento de la información del modelo de proceso de negocio a pilotear.

El contexto de aplicación seleccionado fue el proceso de negocio de pedidos de ventas detallado por las áreas de TI y Comercial. El diagrama BPMN, presentado en la Ilustración 27, inicia con la entrada de los datos en unos archivos de formato Excel, los cuales son diligenciados por las secretarías de cada regional y tienen dos componentes principales: la cabecera y el detalle del pedido. Posteriormente, los archivos son enviados vía correo electrónico y un componente especializado lee el correo y los formatos de pedidos que llegan adjuntos para que sean procesados. Si los formatos cumplen con las reglas de validación iniciales, estos se cargan en la base de datos del Portal web de Pedidos (PyP), los cuales quedan habilitados en el sistema de pedidos web para que los analistas CEDIN los consulten, revisen y generen la solicitud donde se actualizan en el sistema del portal de pedidos web y es integrada en el sistema de información

del ERP INFORFX. Por último, los vendedores consultan, revisan y ajustan la solicitud según el caso, para que esta se consolide como una factura en el ERP (ver Ilustración 27).

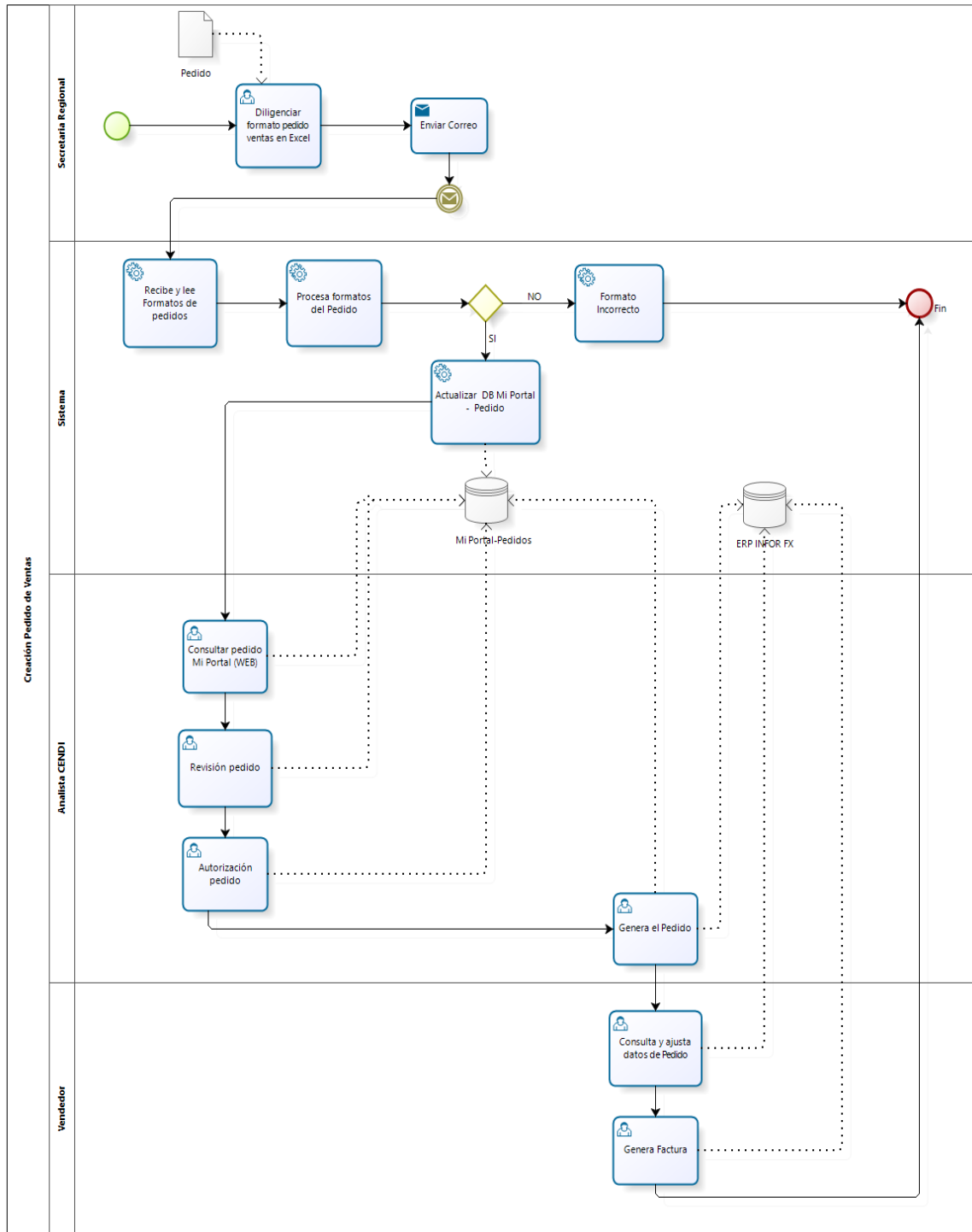


Ilustración 27. Modelo de proceso pedido de ventas

Fuente: elaboración propia

Los sistemas de información identificados en el modelo de procesos de creación del pedido de ventas (ver Ilustración 28) son los siguientes:

- Sistema de información de PyP: es el Portal de Pedidos web, el cual ofrece las funcionalidades para realizar consulta y modificación de pedidos. Este sistema se integra con el ERP INFORFX al autorizar y pasar la información para incorporar como un nuevo pedido de venta en el sistema ERP.
- Sistema de información ERP INFORFX: este sistema es el encargado, dentro de su funcionalidad, de crear un pedido de ventas a partir de los datos enviados por el sistema de información del PyP, el cual es habilitado para su consulta y modificaciones. Además, este pedido pasa posteriormente a convertirse en una factura dentro del ciclo de vida comercial del ERP.

En el proceso de negocio se identificaron unas necesidades importantes relacionadas con la calidad de los datos, las cuales impactan los sistemas de información involucrados y generan reprocesos por inconsistencia en la información. Por eso es necesario aplicar soluciones como las propuestas en este proyecto de maestría, para automatizar el seguimiento continuo de las validaciones a través de formulaciones, métricas y reglas de negocio. Cabe destacar que esto beneficiará a la empresa, al generar las respectivas alertas que identifiquen y midan las inconsistencias en los datos sensibles que afectan la toma de decisiones tanto de los consumidores como de los expertos de los datos del negocio y los analistas de calidad de datos del área TI.

4.2.2 Componente capa de datos

Se seleccionaron los datos sensibles en temas concernientes a la calidad de datos del proceso de negocio para la creación de pedidos de ventas. Para ello se construyeron artefactos

que ayudaron a recopilar la información relevante para identificar los datos, los nombres de los campos y de las tablas, las fuentes de datos, los atributos de los campos y las reglas de negocio implícitas en los datos. A continuación, se especifican estas herramientas:

- **Artefacto descripción proceso de negocio:** tiene como objetivo identificar los procesos de negocio que son impactados por la creación de pedido de ventas, los sistemas de información y la descripción de la problemática a nivel de los procesos involucrados como se muestra en la Ilustración 28.

Proceso de creación de pedidos de venta				
Ítem PN	Procesos de Negocio (PN)	Sistema de Información	Descripción	Problemática
1	CORMECIAL- LOGISTICA	Mi Portal	Creación de pedidos	Compleitud pedidos con información faltante, información errada (códigos de vendedores, direcciones envíos), llega sin orden de compra
2	FINANCIERO - LOGISTICA	ERP INFORLX	Asignación , confirmación y facturación	Se presentan errores en temas de dirección envío , teléfono, listas de precios (el precio no está sincronizado ente PyP y ERP) y productos, cuando los comerciales mandan formatos desactualizados con artículo que ya fueron dados de baja

Ilustración 28. Artefacto proceso de negocio

Fuente: elaboración propia

- **Artefacto de perfilamiento de datos:** se recopiló de todos los campos que contienen la información de los datos monitorizados. Aquí se identifican: el nombre del campo, la descripción del campo, el tipo de dato (texto, fecha, número), la longitud del campo, el nombre de la tabla, el nombre de la base de datos y una descripción de la regla de negocio. Con esta infomación se perfilan los datos sensibles y relevantes para el proceso de negocio que se detallan en la Ilustración 29.

Perfilamiento de datos a ser evaluados							
Ítem PN	Nombre del Campo	Descripción	Tipo de Dato	Longitud	Nombre de Tabla	Base de datos	Regla de Negocio
1-2	direccionentr	dirección de envíos de productos de un pedido	varchar	200	Pedidos	Mi_Portal	se controla por el analista posteriormente en la aplicación de Mi Portal
1-2	dirección	este campo corresponde a la dirección del cliente	varchar	200	Pedidos	Mi_Portal	se controla por el analista posteriormente en la aplicación de Mi Portal
1-2	telefonos	este campo se almacena la información del teléfono(s) del cliente	varchar	60	Pedidos	Mi_Portal	Se maneja con el maestro del ERP
1-2	listaprecios	lista de precios es un dato obligatorio	varchar	50	Pedidos	Mi_Portal	cada formato tiene su propia lista de precios y en algunos casos el formato no lo tiene y en este caso el analista seleccionarla
1-2	Flistaprecios	este campo corresponde a la fecha del precio	timestamp	8	Pedidos	Mi_Portal	Casos de venta con precio de periodos anteriores
1-2	codigo	este campo corresponde al código de artículo	varchar	35	detpedidos	Mi_Portal	se controla por de la aplicación de Mi Portal en el ERP le dan manejo si no hay existencias
1-2	valor	este es el valor donde queda el valor que viene desde el formato Excel	decimal	12,7	detpedidos	Mi_Portal	el ERP definió el precio de artículo según su configuración
1-2	ordencompra	en este campo se almacena la orden de compra del cliente y es obligatorio	varchar	23	Pedidos	Mi_Portal	este campo es obligatorio para crear el pedido en el ERP
1-2	vendedor	en este campo se almacena el código del vendedor	decimal	6	Pedidos	Mi_Portal	este no es obligatorio pero el ERP de la un código por defecto
	Valor_Calculado	en este campo se almacena el valor calculado con base en el precio entregado	decimal	12,7	Pedidos	Mi_Portal	En la actualización con el sistema ERP el dato debería ser diferente de cero

Ilustración 29. Artefacto de perfilamiento de datos

Fuente: elaboración propia

- **Artefacto de reglas de negocio de datos:** tiene como finalidad establecer un grupo de reglas de negocio en un contexto definido (pedidos de ventas) y su estructuración en los procesos de negocios, al identificar: los atributos, la descripción de la validación de la regla y los criterios de aceptación. Después de analizar las diferentes reglas y validaciones, se priorizaron y documentaron como se muestra en las ilustraciones 30, 31, 32 y 33.

Reglas de Negocio	
ID Proceso de Negocio	001
ID Regla de Negocio	001
Nombre	validación de estándar de dirección de entrega
Descripción	Esta regla aplica solamente para los tipo de pedidos de venta, la dirección de entrega debe tener el estándar de datos
Rol	Analista del Centro de Distribución
Área	Comercial
Fecha de Creación	9/06/2020
Prioridad	Alta
Campos Relacionados	direccionentr
Sistemas de Información Relacionados	Mi Portal y ERP
Términos asociados	Dirección de entrega de pedido
Porcentaje de tolerancia aceptado	5%

Ilustración 30. Regla de negocio dirección de entrega

Fuente: elaboración propia

Reglas de Negocio	
ID Proceso de Negocio	002
ID Regla de Negocio	002
Nombre	validación de estándar de dirección del cliente
Descripción	Esta regla aplica solamente para los tipo de pedidos de venta, la dirección de cliente no corresponde con el maestro de clientes del ERP
Rol	Analista del Centro de Distribución
Área	Comercial
Fecha de Creación	9/06/2020
Prioridad	Alta
Campos Relacionados	dirección
Sistemas de Información Relacionados	Mi Portal y ERP
Términos asociados	Dirección del cliente
Porcentaje de tolerancia aceptado	15%

Ilustración 31. Regla de negocio dirección de cliente

Fuente: elaboración propia

Reglas de Negocio	
ID Proceso de Negocio	001
ID Regla de Negocio	003
Nombre	validación de artículos con precio cero
Descripción	Esta regla aplica solamente para los tipo de pedidos de venta, todos los artículos del pedido deben tener un precio mayor a cero
Rol	Analista del Centro de Distribución
Área	Comercial
Fecha de Creación	9/06/2020
Prioridad	Alta
Campos Relacionados	Valor_Calculado
Sistemas de Información Relacionados	Mi Portal y ERP
Términos asociados	Precio de venta
Porcentaje de tolerancia aceptado	0%

Ilustración 32. Regla de negocio precio

Fuente: elaboración propia

Reglas de Validación	
ID Proceso de Negocio	001
ID Regla de Negocio	004
Nombre	Diferencia entre el precio de entrada y el definido en el ERP
Descripción	El dato del entrada del precio de un producto puede llegar con un valor diferente al valor definido por el ERP y estos deberían ser iguales, es aceptable un margen de error de entre el 1% al 2% del valor
Rol	Analista del Centro de Distribución
Área	Comercial
Fecha de Creación	9/06/2020
Prioridad	Alta
Campos Relacionados	valor - valor_calculado
Sistemas de Información Relacionados	Mi Portal y ERP
Términos asociados	Precio de venta
Porcentaje de tolerancia aceptado	0%

Ilustración 33. Regla de negocio precio ERP

Fuente: elaboración propia

En el modelo de datos hay dos bases de datos, las cuales soportan toda la lógica parametrizada (Configuración DQ y BRM DQ) en cuanto a las reglas de negocio, formulaciones, criterios de aceptación y métrica; estas son pobladas con la información levantada en los artefactos para ser incorporadas en los modelos de datos como se describe a continuación:

- Configuración DQ: este modelo está compuesto por las siguientes tablas:
 - Procesos Negocios: esta tabla se completa con la información del artefacto descripción (proceso de negocio).
 - Dimensión DQ: esta se modela a partir de la definición que se da en las reglas de negocio de datos, donde se identifican las dimensiones a ser parametrizadas, las cuales fueron encontradas como requerimiento en el levantamiento del negocio, como unas necesidades de mejora en datos que son sensibles para la organización y requieren se identifique el origen del problema. Se seleccionaron 4 datos en el proceso de negocio (dirección de entrega, dirección cliente, precio artículo y diferencia en precio), estos fueron configurados en las dimensiones (Precisión, Completitud, Relacionada con el tiempo y Consistencia).
 - Factores DQ: esta información se configura con la definición que se da en el artefacto (reglas de negocio de datos), en el que se especifica la dimensiones a ser parametrizadas.
 - Métrica DQ: estas reglas se forman teniendo en cuenta la definición que se da en el artefacto de (reglas de negocio de datos y perfilamiento de datos).
 - Métrica instanciada DQ: se complementa a partir del artefacto de (perfilamiento de datos).

- Medida DQ: esta tabla se configura con las métricas y los valores esperados para las mediciones.
- BRM DQ: este modelo de datos está compuesto por las siguientes tablas:
 - Regla Negocio DQ: esta tabla es parametrizada a partir de la definición realizada en los procesos de negocios y el área TI; se tiene en cuenta la información suministrada por las reglas de negocio de datos; además, se construye la lógica de consulta.
 - Criterios Aceptación DQ: esta tabla se complementa por el artefacto (reglas de negocio de datos), en el que se define el parámetro del valor (criterio de aceptación) por el negocio y, a su vez, se encuentra relacionado con la tabla (Regla Negocio DQ).
 - Formulación DQ: esta tabla se parametriza con base en el artefacto (reglas de negocio de datos) y se relaciona con la base de datos (Configuración DQ). Como resultado se obtienen los identificadores de relaciones con las tablas de (Formulación DQ), la (Dimensión DQ) y el (atributo de calidad) en la tabla (Factores DQ). Asimismo, se establece la formulación para generar los (indicadores) en la tabla del (Histórico Log DQ), y consultar el número total de registros que van a ser monitoreados o muestreados según el caso (tot regis monitoreo) de la tabla (Formulacion DQ).
 - Configuración DQ en la tabla (Procesos Negocios): esta es una tabla que sirve como vinculación al modelo de datos de (Configuración DQ) y el modelo de datos (BRM DQ) para completar información requerida por la tabla (Formulación DQ).

4.2.3 Fuentes de datos

El criterio de selección para la muestra de los datos corresponde a las fechas en que el volumen de las transacciones de pedidos de ventas es más alto a juicio de los expertos del proceso de negocio, es decir, los días finales de mes y mediados del mes siguiente. Con este escenario

establecido se seleccionaron las fechas para realizar el cargue de datos, como se muestra en la Ilustración 34.

Fechas de cargue	Dia 1	Dia 2	Dia 3	Dia 4	Dia 5
Descripción	27/02/2020	28/02/2020	14/03/2020	17/03/2020	18/03/2020
Cantidad de Registro de Carga	30	50	40	100	60

Ilustración 34. Número de registro de cargue de datos pedidos de ventas

Fuente: elaboración propia

La fuente de datos se tomó del modelo de datos de pedidos de ventas, caso de estudio del sistema de información del portal de pedidos web. Esta contiene dos tablas relacionales de la base de datos: 1) el encabezado del pedido (VMPedidos) que registra datos relevantes para el perfilamiento del cliente, así como la logística de entrega de productos e inventarios y 2) el detalle de pedidos (VDPedidos) como se muestra en la Ilustración 35 del modelo de pedidos de ventas.

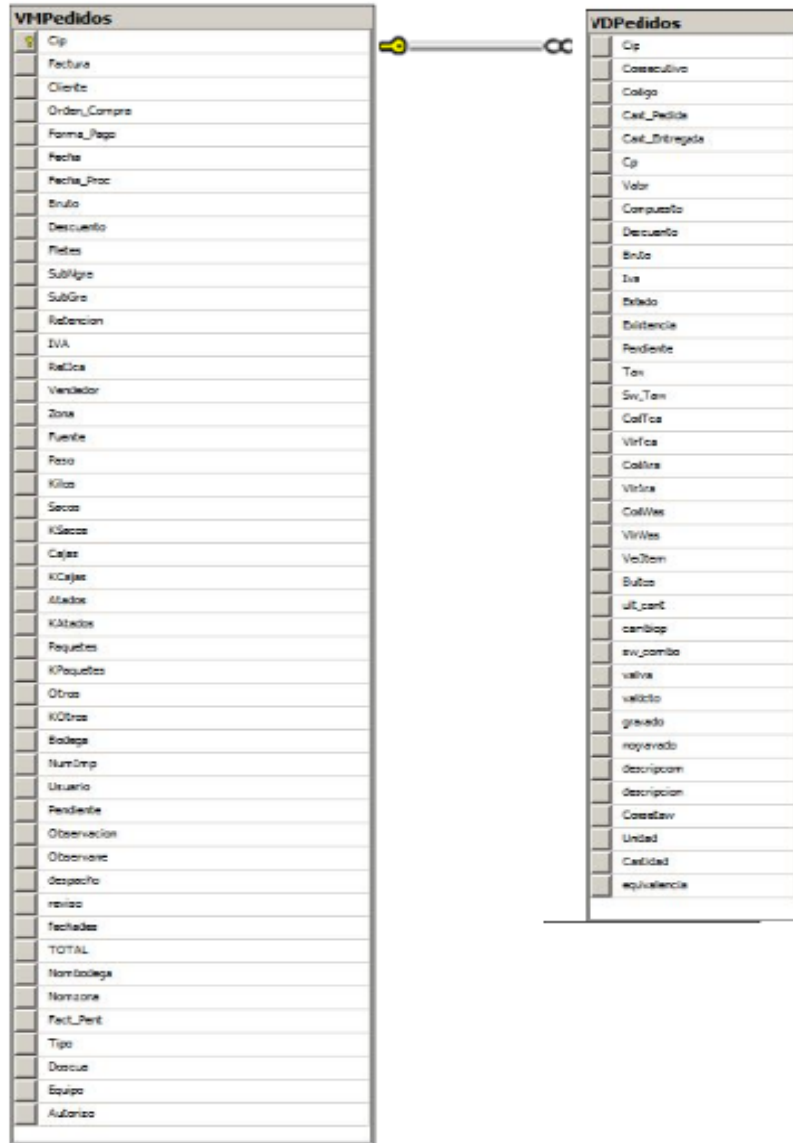


Ilustración 35. Modelo de datos de las entidades de pedidos ventas

Fuente: elaboración propia

El cargue de la información de los archivos de pedidos de ventas se guardó en un formato CSV, y se recopilaron en cinco fechas: el 27 y el 28 de febrero de 2020; y el 14, 17 y 18 de marzo del 2020. Las ilustraciones 36 y 37 ejemplifican los datos de los archivos CSV empleados para el cargue de información.

Código	Empresa	Dirección	Tipo	Fecha	Lugar	Monto
BPE68082	Clinica Nueva	Daigona 21 Nro 22	TE	2020-02-27 13:08:00	Dago 30	67,838,472.00
BPE68083	H & h s	Barrio Abril	TE	2020-02-28 13:18:00	Local 23Av 40	5,438,472.00
BPE68084	Quirufano	Zona Industrial C	TE	2020-03-14 14:49:00	Crr 30 ·77	38,780,342.00
BPE68085	3M	Via 40	TE	2020-03-17 14:54:00	Via 40 # 13-45	4,538,002.00
BPE68086	BAYER	Calle 12 # 13 45	TE	2020-03-18 15:37:00	Calle 20 # 46 - 3	30,384,000.00

Ilustración 36. Cabecera del pedido de venta

Fuente: *elaboración propia*

Cip	Consecutivo	Codigo	Cant_Pedida	Cant_Entregada	Cp	Precio
BPE68085	1	202570	12.000	12.000	1	206.00
BPE68085	2	496030	1.000	1.000	1	36000.00
BPE68086	1	202570	12.000	12.000	1	206.00
BPE68086	2	496030	1.000	1.000	1	36000.00
BPE68087	1	202573	10.000	10.000	1	281.00
BPE68087	2	207604	100.000	100.000	1	446.00
BPE68087	3	207604	50.000	50.000	1	446.00
BPE68088	1	202573	120.000	120.000	1	281.00
BPE68088	2	496030	2.000	2.000	1	36000.00
BPE68088	3	207604	50.000	50.000	1	446.00

Ilustración 37. Detalle del pedido de ventas

Fuente: *elaboración propia*

4.2.4 Ejecución del RPA

Los ciclos de procesamiento para el monitoreo y la generación de alertas se determinaron en conjunto con el proceso de negocio y el área.

Ciclo	Hora Inicio	Hora Finalización	ID Proceso de Negocio	Frecuencia	Estado
1	8:00	10:00	1	Todos los Días	Activo
2	14:00	17:00	1	Todos los Días	Activo

Ilustración 38. Ciclos de ejecución de RPA

Fuente: *elaboración propia*

Como se muestra en la Ilustración 38, el RPA se ejecutó dos veces con una frecuencia diaria para llevar a cabo el proceso de negocio para pedidos de ventas.

4.2.5 Informe de alertas

Las dimensiones configuradas para identificar las inconsistencias fueron: Precisión, Completitud, Relacionada con el tiempo, Consistencia y los datos establecidos del negocio (dirección de entrega, dirección cliente, precio artículo, diferencia en precio). El RPA procesó y generó las alertas identificadas en el universo cargado con la información del negocio, y reportó la dimensión o las dimensiones que presentaban errores. En este proceso se identificó que la única dimensión con un margen de error por encima de los criterios de aceptación fue la de Precisión, este se presentó en el dato dirección de entrega y fue reportado por el RPA en las alertas.

Este informe que envió el RPA, reporto los resultados consignados en las ilustraciones 40 y 41. En este apartado se presenta el resumen del resultado de dos alertas en la ejecución de los dos ciclos en las muestras cargadas de los datos del negocio. A continuación, se presentan algunas observaciones:

- El ciclo 2 fue ejecutado de 14:00 p. m. a 17:00 p. m. del 27 de febrero 2020 para el proceso de negocio pedido de ventas, reporta problemas en las siguientes dimensiones: precisión, atributo de calidad sintáctica; lo cual afecta el campo de dirección de envío del sistema de información Portal web de Pedidos.
- De una cantidad total de 30 registros consultados, 17 presentan error en la dimensión precisión.
- Lo anterior genera un indicador en rojo con un porcentaje del 57 %, el cual sobrepasa el porcentaje del 5 % del umbral definido en el criterio de aceptación del proceso de negocio para la dimensión precisión en el dato de dirección de envío y específicamente en el atributo de calidad sintáctica como se observa en la Ilustración 39.

Alerta de Calidad de Datos										
Ciclo : 2 (14:00 P.M - 17:00 P.M)										
Fecha de procesos: 27/02/2020										
Proceso de Negocio: Pedido de Ventas										
Dimensión de Calidad	Atributo de Calidad	Campo Afectado	Métrica de Calidad	Descripción	Numero Total Registros NR	Numero Registros Errados NRE	Indicador	Sistemas de Información	Formula	Criterio de Aceptación CA
Precisión	Sintáctica	dirección de envíos de productos de un pedido	Falta de estandarización	Valores registrados en diferentes formatos o en un formato diferente al definido como estándar	30	17	57%	Portal Web Pedidos	$((1-(NRE/NR))*100)/100$	5%

Ilustración 39. Informe de alerta de ejecución RPA monitoreo de calidad de datos ciclo 2

Fuente: elaboración propia

- El ciclo 1 fue ejecutado de 7:00 a. m. a 10:00 a. m. el 28 de febrero 2020 para el proceso de negocio pedido de ventas y reporta problemas en las medidas de precisión, atributo de calidad sintáctica, lo cual afecta el campo dirección de envío del sistema de información Portal web de Pedidos.
- De una cantidad total de 21 registros consultados, 7 presentan error en la misma dimensión de precisión o cual indica que este dato de dirección de envío esta generando una inconsistencia repetitiva.
- Lo anterior genera indicador en rojo con un porcentaje del 33 %, el cual sobrepasa el porcentaje del 5 % del umbral definido en el criterio de aceptación del proceso de negocio como se muestra en la Ilustración 40.

Alerta de Calidad de Datos										
Ciclo : 1 (7:00 A.M - 10:00 A.M)										
Fecha de procesos: 28/02/2020										
Proceso de Negocio: Pedido de Ventas										
Dimensión de Calidad	Atributo de Calidad	Campo Afectado	Métrica de Calidad	Descripción	Numero Total Registros NR	Numero Registros Errados NRE	Indicador	Sistemas de Información	Formula	Criterio de Aceptación CA
Precisión	Sintáctica	dirección de envíos de productos de un pedido	Falta de estandarización	Valores registrados en diferentes formatos o en un formato diferente al definido como estándar	21	7	33%	Portal Web Pedidos	$((1-(NRE/NR))*100)/100$	5%

Ilustración 40. Informe de alerta de ejecución RPA monitoreo de calidad de datos ciclo 1

Fuente: elaboración propia

Esto quiere decir que, independientemente del número de registros recibidos en el tiempo total de ejecución del proceso automatizado por el RPA, en el monitoreo de calidad de datos se evidencia una afectación en una dimensión por las reglas identificadas como error para un universo de 280 pedidos de ventas correspondientes a cinco días entre el mes de febrero y marzo de 2020 durante los dos ciclos de ejecución, identificando que dentro de las mediciones configuradas la que presenta errores es la de precisión y específicamente el atributo de calidad sintáctica.

El RPA encontró estas inconsistencias y generó las alertas para el proceso de negocio pedido de ventas. En este caso de estudio es importante hallar y generar las alertas de esas inconsistencias, puesto que, de 51 pedidos realizados, 24 de ellos presentaron errores en el campo de dirección de envío. Hallar estas dificultades permite tomar las medidas correctivas y ajustar la información real de las direcciones de envío, para evitar los reprocesos en el área Logística de entrega de productos y cumplir con las entregas de forma oportuna y a tiempo.

El 54 % y 33 % representa un porcentaje alto de error en el dato dirección de envío frente al universo (280 pedidos de ventas), lo cual indica que en la dimensión de precisión y del atributo sintáctico en la métrica correspondiente a falta de estandarización del dato dirección envío, se evidencia que la entrada del dato está mal digitada o en un formato diferente; un indicador que obliga a tomar decisiones de forma inmediata para mitigar este error y así lograr bajar el porcentaje de alerta.

Cabe resaltar que el eje central de la generación de estas alertas es el RPA, el cual le da un valor particular a la organización permitiéndole identificar errores que antes no eran perceptibles o no tenían una visibilidad en los negocios. De tal manera que las mediciones configuradas de forma dinámica generan nuevos escenarios de errores; esto permite impulsar un

ciclo de mejora continua que debe realizar la organización frente a estas alertas, para mejorar la calidad de la información de manera significativa en los procesos del negocio. Asimismo, permite demostrar que en un modelo automatizado se puede llegar, poco a poco, a una madurez de alta calidad en los datos y asegurar que la calidad se mantiene en el tiempo.

4.2.6 Visualizador de Excel

La consulta realizada corresponde a los siguientes parámetros de entrada como se muestra en la Ilustración 41, se realiza la consulta para la dimensión precisión puesto que en los datos cargados fueron los que reportaron alertas de inconsistencia en esta dimensión y atributo de calidad, por esta razón son los disponibles en el visualizador.

Rango de fechas para consultar			
Fecha de Inicio	1/02/2020	Fecha de Finalizacion	31/03/2020
Seleccione el proceso de negocio			
Proceso de Negocio :	Pedidos de Ventas	▼	Las Fecha , Proceso de Negocio y Dimension son Obligatorias para realaizar la Consulta
Dimensión :	Precisión		
Dato:	Direccion_Envio		
Atributó de Calidad :	Sintáctica	▼	
<input type="button" value="Consulta Calidad de Datos"/>			

Ilustración 41. Visualizador consulta de Excel de inconsistencias

Fuente: elaboración propia

Los resultados se describen en las siguientes observaciones:

- La Ilustración 42 muestra las fechas en que se generan errores con los porcentajes de criticidad, con el fin de que los roles interesados en la mejora de la calidad de los datos puedan tomar decisiones o poder mostrar a cualquier ente de control el seguimiento que se hace. En la Ilustración 42 se entreven los errores encontrados en el atributo de calidad sintáctica, ya que se evidencia que el porcentaje de error es del 54 % correspondiente al día 27/02/2020 y 33 % del día 28/02/2020, rango de fechas seleccionado del 1 de febrero al 31

de marzo de 2020. Los anteriores datos demuestran que están por encima del criterio de aceptación establecido por el negocio, el cual es del 5 % para el dato dirección de envío definido para la dimensión precisión, en la regla del atributo de calidad sintáctica. Estos datos fueron monitoreados por el RPA, encargado de generar las alertas para la visualización del negocio y gobernanza de información.

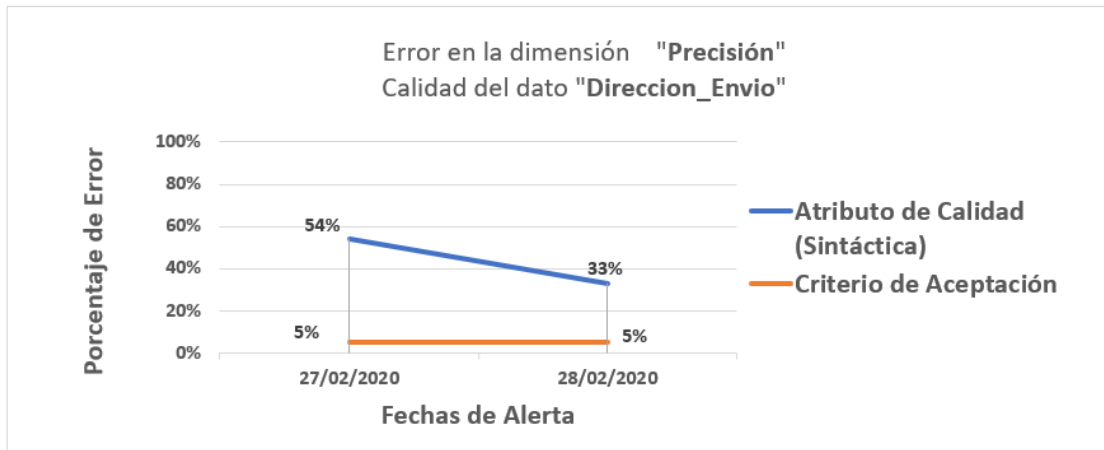


Ilustración 42. Resultado de errores en un rango de tiempo

Fuente: elaboración propia

- La Ilustración 43 muestra la trazabilidad del error del dato en una línea de tiempo y en los cortes de las fechas donde se produjo el evento del error. Se puede visualizar la tendencia de la dimensión en el atributo de calidad; con esto el RPA, al persistir esta información monitoreada, le da una relevante importancia al negocio para aplicar las mejoras y detectar los eventos que ocurren en esta línea de tiempo. En la Ilustración 43 se exhibe la trazabilidad de errores presentados en una línea de tiempo correspondiente del 1 de febrero al 31 de marzo del 2020, en el que se identifican las fechas en los cortes del error producidos en el dato dirección de envío”; con esto se muestra la tendencia de la dimensión precisión con sus porcentajes de error en los puntos encontrados: uno del 54 % correspondiente al día 27/02/2020 y el otro el día 28/02/2020 con el 33 %. La anterior información indica que, en

estos dos días, se presentó algún evento que generó estos tipos de errores. Según la línea de tiempo, es un error constante; por tal razón se debe realizar el análisis y tomar las medidas correctivas del caso y estar al tanto del reporte del RPA.

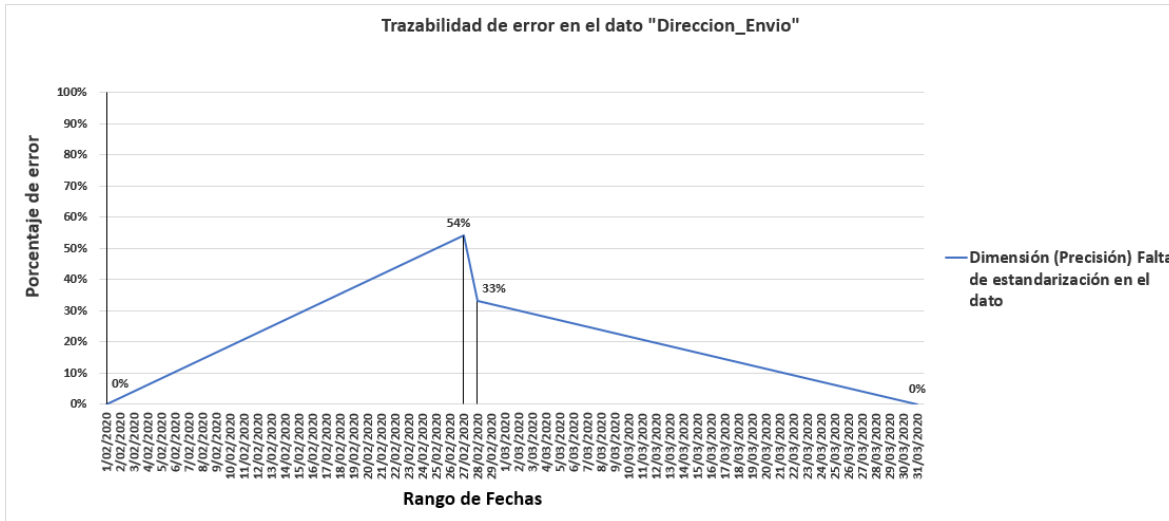


Ilustración 43. Resultado trazabilidad de error por dimensión precisión en un rango de fechas

Fuente: elaboración propia

- La Ilustración 44 muestra el porcentaje de error con las fechas y ciclos en los cuales se presenta el error. Con esta información, el RPA, al persistir esta información monitoreada y los ciclos de su ejecución, le permite al negocio identificar puntualmente las horas en que se produce este error, con el fin de que se llegue a una mayor exactitud e identificación del problema. En la Ilustración 44 se muestran los errores presentados en las fechas y ciclos de tiempo correspondientes del 1 de febrero al 31 de marzo del 2020; se identifica las fechas del error producidos en el dato dirección de envío y muestra la tendencia de la dimensión precisión del atributo de calidad semántica con sus porcentajes de error: uno del 54 % el 27 de febrero de 2020, producido en el ciclo de ejecución de las 7:00 a. m. a 10:00 a. m.; y el otro del 33 % el 28 de febrero de 2020, error producido en el ciclo de ejecución de las 2:00 p.

m. a 5:00 p. m. De acuerdo con la información detalla en estos dos días, puntualmente en estos horarios, se presentó algún evento que originó estos errores.

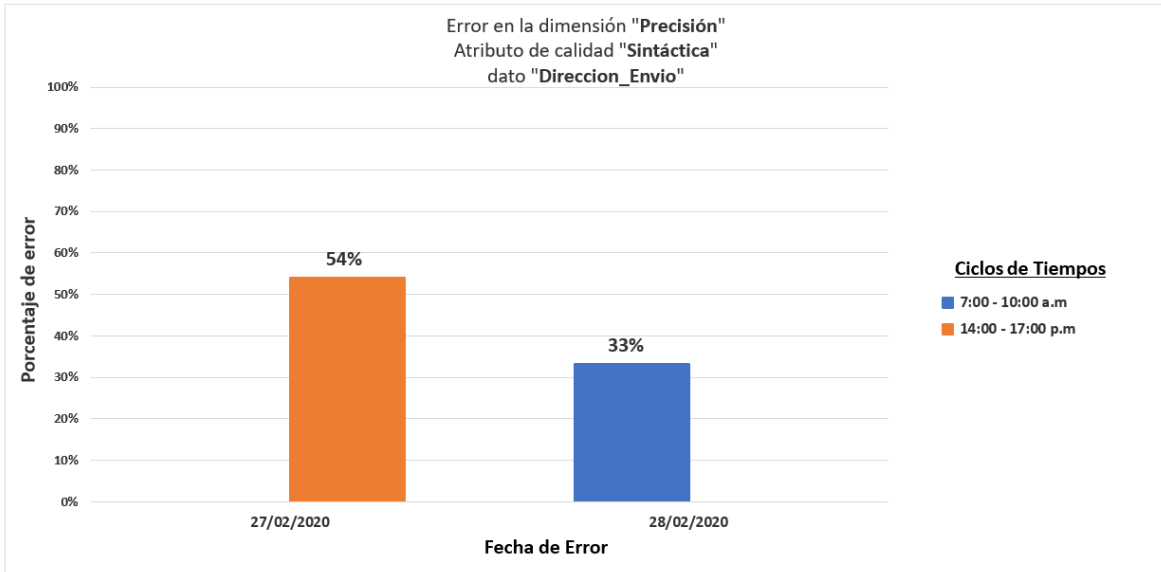


Ilustración 44. Error en ciclo de tiempo del dato errado

Fuente: elaboración propia

El visualizador es un componente que recoge información a demanda para consultar y que muestra solamente las dimensiones, campos y atributos de calidad que presentan errores; del mismo modo, se apoya en lo que el RPA a persistido en el modelo de datos de las Bitácora DQ, que es la que garantiza que la información monitoreada por el RPA esté disponible para su consulta y visualización.

5 Conclusiones, recomendaciones y trabajos futuros

5.1 Modelo de procesos para la evaluación en calidad de datos de las organizaciones

El modelo del proceso de evolución en calidad de datos permitió tener como referencia un marco metodológico soportado por las diferentes metodologías de gobernanza de datos, marcos de trabajo en las buenas prácticas y metodologías en temas relacionados con calidad de datos. De manera que la propuesta de automatización para tener un modelo de monitoreo continuo en la evolución de calidad de datos, se adaptó y así mejoró la propuesta de valor a la organización.

5.2 Automatización de procesos RPA

El componente eje de la solución fue el RPA, con el cual se logró dar el alcance necesario a la solución propuesta. Este componente, en virtud de su potencial tecnológico, aportó la estructuración de las diferentes actividades requeridas; permitió una integración con el estándar de otros componentes de una forma fácil de implementar y así se consiguió cumplir con los objetivos de la propuesta de investigación. Sin duda, la tecnología permitió dar la flexibilidad para un sólido monitoreo y así obtener la generación de alertas de calidad de datos en todos los aspectos propuestos por la solución.

5.3 Desarrollo de la Solución Propuesta

La arquitectura definida para la solución se complementó con el modelo de evaluación en calidad de datos, la cual fue el marco de trabajo base para la definición de los diferentes componentes tecnológicos y la interacción con los componentes funcionales. A partir de estos

elementos y sus definiciones, se logró alcanzar y poner en funcionamiento el piloto al conseguir que el RPA ofreciera todo su potencial en forma dinámica y eficiente.

5.4 Aplicación al caso de estudio

En la implementación de la propuesta del caso de estudio, se lograron satisfacer los siguientes aspectos: cubrir las expectativas planteadas por la propuesta de la solución; el modelo planteado para cada una de las actividades y consecución de defensiones de reglas de negocio; mediciones de datos y el modelo del proceso de negocio a ser validado con la solución. Esto permitió identificar la oportunidad de la solución propuesta con respecto a una automatización más ágil y flexible en la generación de resultados de forma oportuna mediante el RPA.

5.5 Principales dificultades encontradas para el desarrollo de esta investigación

La consecución de las fuentes de datos verdaderas tuvo un nivel de complejidad alto por reserva de la información, lo cual implicó hacer trasformaciones de datos, pero sin perder la esencia de información para efectos de la investigación.

La coordinación de reuniones por parte de los funcionarios de la organización, debido a sus agendas de trabajo, fue un factor limitante para llevar acabo definiciones de pilotaje.

5.6 Contribución de la aplicación

Uno de los aportes más importantes en este documento de la propuesta de la solución para el monitoreo es la generación de alertas en calidad de datos; es decir, la entrega y visibilidad de la información con errores oportuna a los interesados como procesos de negocio y gobiernos de información, en un contexto más de negocio que técnico. Cabe concluir que la instauración de un proceso de automatización, para garantizar una alta calidad de datos y una mejora continua,

genera ventajas y valor a las organizaciones. Esto favorece que las mediciones de calidad de datos sean más flexibles, oportunas e inteligentes y, así mismo, sean alineadas con la dinámica de los requerimientos de datos de hoy en día. Con el uso del RPA, a partir de su modelo de datos centralizado, permite una mantenibilidad y gestión por parte de las organizaciones en la mitigación de errores en datos.

La investigación y propuesta de la solución, desde el punto de vista académico, ofrece un panorama a la Universidad EAFIT en los temas relacionados con la gobernanza de datos desde el aspecto de calidad de datos y sus ventajas en organizaciones públicas o privadas.

5.7 Limitaciones

En el mercado hay herramientas, como se indica en este estudio, que plantean soluciones de temas de calidad de datos de uso específico; el gran reto es lograr tener un aseguramiento en calidad de datos a 360 grados en los diferentes procesos de negocio que interactúan entre ellos.

5.8 Trabajos futuros

RPA con tecnologías cognitivas en calidad de datos organizacionales

Existen unas consideraciones para capitalizar la propuesta implementada, la cual consistiría en la inclusión de nuevas funcionalidades al RPA que le permitan, de forma inteligente, la corrección de datos inconsistentes ya sea individual o en forma masiva y la autogestión de las reglas de negocio mediante el uso de tecnologías cognitivas. Además, un reto sería analizar la posibilidad de agregar Inteligencia Artificial IA, con el objetivo de que un RPA sea capaz de auto aprender del negocio y gestionar reglas de forma autónoma.

6 Referencias bibliográficas

- Aiken, P., Allen, D., Parker, B., y Mattia, A. (2007). Practice Maturity: A Community ' s. *Computer*, (3), 42–50.
- Alcaldía Mayor de Bogotá. (2018). *Política de Gobierno de Datos*. Bogotá: Alcaldía Mayor de Bogotá.
- Arias, A. (2015). La gestión de la calidad: Conceptos básicos. *Facultad de Ciencias de La Comunicacion*, (1), 47.
- Batini, C., Barone, D., Mastrella, M., y Maurino, A. (2004). A Framework and a Methodology for Data Quality Assessment and Monitoring. *Studies in Communication Sciences*, 2, 115–136.
- Batini, C., Barone, D., Maurino, A., y Ruffini, C. (2007). Aframework and a methodology for data quality assessment and monitoring. *Practice-Oriented*, (1), 5–15.
- Buchwald, A., Urbach, N., y Ahlemann, F. (2014). Business value through controlled IT: Toward an integrated model of IT governance success and its impact. . *Journal of Information Technology*, 29(2), 128–147. DOI: <https://doi.org/10.1057/jit.2014.3>.
- Caballero, I., Blanco, M., y Piattini, M. (2014). Optimización del proceso de gestión de información para de la calidad de la información. *Universidad de Castilla*, (5), 1–27.
- Caballero, I., Blanco, M., y Piattini, M. (2014). Optimización del proceso de gestión de información para de la calidad de la información . *Universidad de Castilla*, (5), 1–27.
- Capgemini. (2016). *Robotic process automation (RPA) The next revolution of Corporate Functions*. Obtenido de https://www.capgemini.com/consulting-fr/wpcontent/uploads/sites/31/2017/08/robotic_process_automation_the_next_revolution_of_corporate_functions_0.pdf

- Caro, A., Fuentes, A., y Soto, M. (2013). Desarrollando sistemas de información centrados en la calidad de datos. *Ingeniare*, 21(1), 54–69. DOI: <https://doi.org/10.4067/s0718-33052013000100006>.
- Castillo, A. (2015). *Una gestión de datos para mejorar y dar soporte a la toma de decisiones en los negocios*. Colombia: Instituto Politécnico Nacional.
- Colina, A. (2019). El gobierno de datos: un referente entre el gobierno de TI y la inteligencia de negocios. *ECOCIENCIA*, 6(1), 1–19. DOI: <http://revistas.ecotec.edu.ec/index.php/ecociencia/article/view/186/152>.
- Corrales, D., Ledezma, A., y Corrales, J. (2016). A systematic review of data quality issues in knowledge discovery tasks. *Revista Ingenierías Universidad de Medellín*, 15(28), 125–150. DOI: <https://doi.org/10.22395/rium.v15n28a7>.
- De Soto, A., y Cuervo, E. (2006). Nuevas tenencias en sistemas de información: procesos y servicios. *Pecunia*, 2, 129-158.
- Deloitte. (2015). *Desarrollo de un modelo sea efectivo operativo de gobierno que los equipos de: Una guía para las juntas y administración de servicios financieros*. Obtenido de https://nanopdf.com/download/desarrollo-de-un-modelo-operativo-de-gobierno-que-sea_pdf
- Deloitte. (2017). *Automatización Robótica de Procesos (RPA)*. Obtenido de https://www2.deloitte.com/content/dam/Deloitte/mx/Documents/strategy/Automatizacion_Rob%C3%B3tica_Procesos.pdf
- Félix, J., Mendoza, J., Muñante, D., Quenta, J., y Quispe, H. (2018). *Modelo de gobierno de datos para una entidad tributaria peruana*. Obtenido de Esan Graduate School of Business:

- https://repositorio.esan.edu.pe/bitstream/handle/20.500.12640/1397/2018_MADTI_16-1_03_T.pdf?sequence=1&isAllowed=y
- Garcés, J. (2016). *Caracterización de modelos de madurez en gobierno de datos*. Obtenido de Universidad Pontificia Bolivariana: https://repository.upb.edu.co/bitstream/handle/20.500.11912/2583/INFORME_FINAL_%20Jose%CC%81%20Jaime%20Garce%CC%81s.pdf?sequence=1&isAllowed=y
- Gómez, A., y Piattini, M. (2018). DQTeam S.L. es una spinoff de la UCLM. *Ruidera*, 13, 1–15.
- Gonzáles, R. (2010). *Impacto de la data warehouse e inteligencia de negocios en el desempeño de las empresas: investigación empírica en peru, como país en vías de desarrollo*. Obtenido de Universidad Ramón Llull: <http://manteniment.csuc.cat/>
- Governance, D. (s.f.). *What is Data Governance*. Retrieved from *imperva website*. Obtenido de <https://www.imperva.com/learn/data-security/data-governance/>
- Hernández, C., Mayorga, B., y Peña, L. (2012). Método de madurez para la calidad de los datos. *Punto de Vista*, 2(5), 77–94.
- IBM. (2018). *Automatización de procesos robotizados*. Obtenido de <https://www.ibm.com/downloads/cas/0PZBYB8W>
- Informatica. (s.f.). *Cuadrante Mágico de Gartner 2019 sobre*. Obtenido de <https://www.informatica.com/co/data-quality-magic-quadrant.html>
- Isaac, R., Muni, R., y Desai, K. (2017). *Delineated Analysis of Robotic Process Automation Tools*. Obtenido de First International Conference on Information Technology, Communications and Computing (ICITCC 2017): <https://zenodo.org/record/1134259/files/ICITCC201706.pdf>

- Jarke, M., y Vassiliou, Y. (1997). Data Warehouse Quality A Review of DWQ Project. *Proceedings of the Conference on Information Quality, (9)*, 299–313.
- KOFAX. (2020). *10 ejemplos de uso de RPA para TI, recursos humanos, finanzas, contabilidad y otros departamentos* . Obtenido de <https://www.kofax.es/Blog/10-ejemplos-de-uso-de-rpa>
- KPMG. (2017). *Robotic Process Automation (RPA)*. Obtenido de <https://assets.kpmg.com/content/dam/kpmg/uy/pdf/Folleto-RPA.pdf>
- Lee, Y., Strong, D., Kahn, B., y Wang, R. (2002). AIMQ : a methodology for information quality assessment. *Information & Management, 40* , 133–146.
- Logicalis. (s.f.). *Data Quality : errores comunes en la gestión de la calidad de datos*. Obtenido de <https://blog.es.logicalis.com/analytics/data-quality-errores-comunes-en-la-gestion-de-la-calidad-dedatos>
- Marulanda, C., López, M., y Valencia, F. (2017). Gobierno y gestión de ti en las entidades públicas. *AD-Minister, (31)*, 75–92. DOI: <https://doi.org/10.17230/ad-minister.31.5>.
- Mesa, G., y Cotelo, H. (2015). *Gobierno de datos*. s/e: Costa Rica.
- Mintic. (2016). *Seguridad y privacidad d ela información*. Obtenido de https://www.mintic.gov.co/gestionti/615/articles-5482_Modelo_de_Seguridad_Privacidad.pdf
- Muñoz, J. (2012). *Módulo de evaluación de calidad de datos utilizando reglas de negocio en un contexto determinado*. Obtenido de Universidad de los Andes: <https://repositorio.uniandes.edu.co/bitstream/handle/1992/11770/u622558.pdf?sequence=1&isAllowed=y>

- Otto, B., y Österle, H. (2015). *Calidad de datos corporativos. Requisito previo para modelos comerciales exitosos*. Berlín: epubli GmbH.
- Portilla, J. (2018). Gobierno de datos, un potenciador de los sistemas de gestión de calidad. *Signos*, 9(2), 159-172.
- Royo, C. (2005). *Las competencias como herramientas para el cambio cultural en una organización bancaria*. Obtenido de Universidad de Barcelona: https://www.tdx.cat/bitstream/handle/10803/2667/01.CRM_parte1.pdf?sequence=1&isAllowed=y
- Salah, J. (2017). *Modelo de gobierno y gestión de TI basado en la estrategia de gestión de riesgo para la Secretaría de Educación de Magdalena*. Obtenido de Universidad del Norte: <http://manglar.uninorte.edu.co/bitstream/handle/10584/8079/131482.pdf?sequence=1&isAllowed=y>
- Saltz, J. (Twenty-Fifth European Conference on Information Systems, 1). *Association for Information Systems AIS Electronic Library (AISeL) acceptance factors for using a big data capability and maturity model*. 2017: 6–10.
- Scannapieco, M., y Catarci, T. (2002). Data Quality under the Computer Science perspective. *Archivi & Computer*, 1-12. Obtenido de https://www.researchgate.net/profile/Tiziana_Catarci2/publication/228597426_Data_quality_under_a_computer_science_perspective/links/0fcfd51169a156b61a000000/Data-quality-under-a-computer-science-perspective.pdf
- Serrano, M. (2014). Características de Calidad de Datos de los Almacenes de Datos. *Universidad de Castilla*, (5), 1–31.

- Smits, D., y Van Hillegersberg, J. (2015). IT governance maturity: Developing a maturity model using the delphi method. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 48, 4534–4543. DOI: <https://doi.org/10.1109/HICSS.2015.541>.
- Soto, A. (2006). Nuevas Tendencias en Sistemas de Información . 2, 129–158.
- Stenström, C., Mustafa, A., y Parida, A. (2014). Computerised Analysis of Text Entry Fields. *Luleå: Luleå Tekniska Universitet*, 1, 97–100.
- UiPath. (2015). *RPA and the Back Office/Front Office Debate*. Obtenido de <https://www.uipath.com/blog/rpa-and-the-back-office/front-office-debate>
- Valverde, G. (2014). Experimentos de enseñanza: una alternativa metodológica para investigar en el contexto de la formación inicial de docentes. *Revista Electrónica "Actualidades Investigativas en Educación"*, 14(3), 1-20. DOI: <http://redalyc.org/pdf/447/44732048014.pdf>.
- Wister, V. (2019). *A process model for organizational data quality assessment*. Obtenido de Eindhoven University of Technology: https://pure.tue.nl/ws/portalfiles/portal/118105234/Master_Thesis_Joost_van_Wierst.pdf