

Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado

**MAESTRÍA EN GESTIÓN ESTRATÉGICA DE
TECNOLOGÍAS DE LA INFORMACIÓN Y LAS
COMUNICACIONES**

TRABAJO FINAL DE MAESTRÍA

Gobernanza de Datos. Aplicación en una empresa de
Telecomunicaciones de la Argentina.

AUTOR: MARÍA ALEJANDRA REGUEIRO PÉREZ

DIRECTOR: DR. CLAUDIO FELIPE FREIJEDO

COHORTE 2018
JUNIO 2021

«No hace falta un gobierno perfecto, solo hace falta que sea práctico»

Aristóteles

«Estamos inundados de información y sedientos de conocimiento»

Robert Bolt

Agradecimientos

A mi familia, por su paciencia y apoyo para que pueda realizar este trabajo.

A todos los profesores de la Maestría en Gestión Estratégica de TIC, por transmitirnos sus conocimientos y experiencias con sumo profesionalismo, a las autoridades de la Maestría, especialmente a Claudio Freijedo, director de mi trabajo final de Maestría, y a Virginia Chaina, quienes me guiaron, acompañaron, y me brindaron valiosos consejos, aportes y correcciones.

A todos mis compañeros, excelentes profesionales, por compartir también sus experiencias y conocimientos conmigo durante la cursada.

Resumen

La aparición de Big Data ha posibilitado que las tecnologías de analítica avanzada e inteligencia artificial desarrollen su máximo potencial. Gracias a estas herramientas las organizaciones pueden obtener beneficios de la monetización de sus datos, en consecuencia, su valor como activo estratégico ha aumentado considerablemente. A su vez, y dado que las fuentes de datos son cada vez más variadas y voluminosas, resulta más difícil mantener un nivel de calidad aceptable y su integración supone un esfuerzo cada vez mayor. Adicionalmente, se acrecientan los riesgos asociados a la confidencialidad de la información, motivo por el cual, se han actualizado los marcos normativos vigentes relacionados con la protección de datos personales; lo que también implica que las organizaciones deban adaptarse para dar cumplimiento a nuevas exigencias legales. Todos estos factores complejizan la arquitectura de datos de las organizaciones, y, por ende, el proceso de gestión de la información.

Por lo expuesto, en la actualidad resulta relevante evaluar cómo la implementación de un programa de Gobierno o Gobernanza de Datos impacta en los procesos de gestión de la información de las organizaciones en cuanto a su eficiencia y valor agregado. Con el presente trabajo se pretende responder a dicho interrogante mediante el análisis y la descripción de las características y funciones principales de la Gobernanza de Datos, mediante la identificación y descripción detallada de los requisitos y funcionalidades que es deseable que brinden las herramientas informáticas disponibles para acompañar su implementación, y la identificación del impacto producido por la utilización de éstas en la eficiencia de la gestión de la información y el gobierno de datos en una empresa de telecomunicaciones de nuestro país.

Para ello, se realizó una investigación de tipo teórica y descriptiva, mediante la lectura de bibliografía especializada, la realización de entrevistas a expertos y la observación de campo durante la trayectoria profesional.

Palabras clave: gobierno de datos, calidad de datos, monetización de datos.

Abstract

The emergence of Big Data has enabled advanced analytics and artificial intelligence tools to reach their maximum potential. Thanks to this, organizations can obtain the benefits of data monetization; consequently, its value as a strategic asset has increased considerably. As data sources become more varied and voluminous, maintaining an acceptable level of data quality is becoming increasingly more difficult, requiring additional efforts around data integration. Moreover, the risks associated with data confidentiality have increased. As a result, regulatory frameworks related to personal data protection have been updated, which means that organizations must now adapt to comply with new legal requirements. All these factors increase the complexity of an organization's data architecture and, therefore, the information management process.

Consequently, it is of relevance to evaluate the impact of a Data Governance Program on the information management processes of organizations, in terms of their efficiency and added value. This paper aims to answer this question through the analysis and description of the main characteristics and functions of Data Governance; through the identification and detailed description of the ideal requirements and functionalities, and what the currently available computer tools provide to support their implementation; and the analysis of the impact produced by the use of these tools in the management of information and data governance in a telecommunications company in our country. To achieve these goals, theoretical and descriptive research was conducted; in particular, the review and analysis of specialized bibliography, interviews with subject-matter experts, and on-the-field observations gained throughout the author's professional career.

Keywords: data governance, data quality, data monetization.

Índice

Introducción.....	7
Planteamiento del tema.....	7
Objetivos.....	8
Objetivo principal	8
Objetivos Específicos	9
Metodología.....	9
Capítulo 1. Marco Teórico.....	10
Conceptos iniciales	10
Factores de la cultura basada en datos	16
Primer factor: Alta Dirección.....	16
Segundo factor: Datos.....	16
Tercer factor: herramientas informáticas	16
Cuarto factor: Organización.....	17
Quinto Factor: Proceso de decisión	17
Sexto factor: Gestión del cambio	17
Niveles de madurez de la cultura basada en datos	18
Gobierno Corporativo, Gobierno de IT, Gobierno de Información y Gobierno de Datos	20
Morfología y componentes del Gobierno de Datos	28
Mecanismos de Gobernanza	31
Mecanismos estructurales	32
Estructura organizacional	32
Roles y responsabilidades.....	33
Propiedad de los datos	34
Custodia de los datos	35
Instrumentos para la designación de roles y responsabilidades.....	35
Matriz RACI	35
Coordinación y alineamiento	36
Mecanismos de Procesos	39
Mecanismos Relacionales.....	42
Arquetipos de Gobernanza.....	42
Factores críticos de éxito del Gobierno de Datos	44
Herramientas informáticas para Gobierno de Datos.....	45
Funcionalidades principales.....	45
Linaje	45

Glosario de Negocio	47
Catálogo de Datos	48
Custodia, propiedad y administración de los objetos de datos	48
Calidad	49
Gestión de metadatos	49
Soluciones disponibles en el mercado	50
Capítulo 2. Explicación del caso elegido.....	76
Mercado de Telecomunicaciones.....	77
Contexto político, económico y social.....	78
Marco Regulatorio	78
Leyes de Protección de Datos	79
Organigrama de la empresa. Nivel estratégico	79
Cultura basada en datos	81
Arquitectura Empresarial	82
Principios arquitectónicos de los datos:.....	85
Capítulo 3. Implementación del programa de Gobierno de Datos	86
Situación inicial	86
Análisis de los mecanismos utilizados.....	86
Mecanismos estructurales.	86
Estructura organizacional	86
Roles y responsabilidades	87
Coordinación y alineamiento	90
Comité estratégico de Gobierno de Datos	91
Comité operativo de Gobierno de Datos.....	91
Mecanismos de Procesos	91
Modelo de procesos	92
Monitoreo y evaluación	94
Desarrollo.....	95
Mecanismos relacionales	95
Visión compartida.....	95
Colaboración.....	96
Transferencia del conocimiento	96
Conclusiones generales.....	97
Bibliografía.....	101
Anexo I	103

Introducción

El presente trabajo trata sobre cómo abordar la complejidad de la gestión de la información en las organizaciones, y tiene la finalidad de describir la aplicación de un programa de Gobernanza de Datos en una empresa de telecomunicaciones de nuestro país, el marco de trabajo aplicado, las metodologías empleadas, los artefactos, las herramientas organizacionales e informáticas utilizadas y el valor que le aporta a la organización su implementación.

Planteamiento del tema

El rápido incremento de la digitalización de los procesos organizacionales, el aumento de la variedad, velocidad y volumetría de datos, la aparición de *IoT*¹ o “internet de las cosas”, la amplia variedad de aplicaciones y herramientas de explotación de datos conlleva a una complejidad cada vez mayor de la arquitectura de aplicaciones y de datos de las organizaciones, y hacen más compleja la administración del flujo de la información a lo largo de su ciclo de vida.

A su vez, es necesario cumplir con normas de seguridad de la información y marcos legales vigentes que impactan en los procesos específicos de recolección, almacenamiento, procesamiento, actualización, eliminación, retención y distribución de datos. La gestión de la información requiere de la existencia de un marco de trabajo, políticas, normas, procedimientos, estándares y mejores prácticas para procesar y explotar los datos “corporativos” y “estructurados” que se generan y fluyen dentro de la organización, asegurando su calidad, exactitud y no redundancia; así como también de los “no estructurados”, que también aportan valor para la generación de información.

Debido a lo expuesto, surge en las organizaciones, la necesidad de “gobernar” un activo estratégico clave y cada día más crítico: los datos.

¹ Por su sigla en inglés: Internet of things

Por lo tanto, se entiende que la implementación de un programa de gobierno de datos mejora la eficiencia de la gestión de la información en la organización y agrega valor al proceso decisorio y a los datos como activo económico. Por ello, es recomendable que las organizaciones lo diseñen, planifiquen, implementen y evolucionen, de manera que sus componentes estén alineados con la estrategia organizacional y la estrategia de TI para que su aplicación contribuya a la mejora de la gestión de procesos.

Con este trabajo, se pretende responder el siguiente interrogante:

¿Puede la implementación de un programa de Gobernanza de Datos mejorar la eficiencia en los procesos de gestión de la información y agregar valor a la organización?

para ello también, es necesario responder:

¿Qué es la Gobernanza de Datos, y cuáles son sus componentes principales?

¿Encontramos herramientas informáticas disponibles en el mercado que acompañen este proceso?

¿Cómo impacta la utilización de herramientas y las prácticas de Gobierno de Datos en la eficiencia en los procesos de gestión de la información?

Objetivos

Objetivo principal

El principal objetivo del presente trabajo es evaluar el impacto de un programa de Gobernanza de Datos en los procesos de gestión de la información de la organización en cuanto a su eficiencia y valor agregado.

Objetivos Específicos

Dentro de los objetivos específicos que se pretenden alcanzar, se encuentran:

- Describir las características y funciones de la Gobernanza de Datos.
- Identificar y describir qué requisitos y funcionalidades deben tener las herramientas informáticas que pueden acompañar su implementación.
- Identificar el impacto producido por la utilización de las herramientas y prácticas en la eficiencia de la gestión de la información y el gobierno de datos en una empresa de telecomunicaciones de nuestro país.

Metodología

El presente trabajo es una investigación de tipo teórica y descriptiva, con el objetivo de conocer los componentes (organizacionales, tecnológicos, sociales, etc.) y describir cómo es y cuáles son los aspectos relevantes de la aplicación de un programa de Gobernanza de Datos en una empresa de Telecomunicaciones de nuestro país, por qué es necesaria su aplicación, qué valor aporta, y cómo puede llevarse a cabo su implementación.

Se trata de un Estudio de Caso. El tipo de diseño es un diagnóstico de situación, con un análisis global y detallado, sobre el tema mencionado. Su enfoque es cualitativo, y de tipo de diseño prospectivo.

Las técnicas de recolección de datos utilizadas fueron la realización de entrevistas individuales a expertos en el tema, - la observación en el campo durante la trayectoria laboral, junto con la lectura y análisis de bibliografía especializada en el tema tratado.

Capítulo 1. Marco Teórico

Conceptos iniciales

Desde hace tiempo, las organizaciones invierten fuertemente en TIC² en diversas áreas y con diversos objetivos, a fin de obtener las ventajas competitivas que su utilización estratégica puede brindar.

Con el correr del tiempo, se ha profundizado y acelerado la automatización de los procesos de negocios, así como también la digitalización y la transformación digital de las organizaciones, lo que ha provocado un gran incremento en el volumen de transacciones, que, sumado a otras actividades de los usuarios, clientes y consumidores, se traducen en una importante y creciente volumetría de datos, que pueden explotarse, y que constituyen la materia prima para la generación de información y conocimiento. Estos fenómenos de digitalización y transformación digital se vieron potenciados recientemente por la pandemia de SARS-CoV-2; lo cual incrementó la digitalización de las operaciones y procesos de las organizaciones, fenómeno apoyado en la amplia difusión de telefonía móvil y los *smartphones*, que permiten una conexión en línea desde cualquier lugar y a toda hora del día, para la realización de transacciones, y la creación de datos debido a la accesibilidad que proporciona desde cualquier punto geográfico, sumado a otros factores tales como, el auge de las redes sociales, el aumento de la capacidad de procesamiento “en la nube”³, y la aparición de internet de las cosas, se generó un aumento exponencial de la generación de datos masivos.

Los datos que se capturan como resultado de la operatoria “tradicional” de las organizaciones, y que tienen un formato estandarizado, se conocen como “estructurados”. Por su parte, los “no estructurados”; son aquellos que provienen de diversas fuentes, principalmente de las redes sociales, y que no se encuentran en el formato tradicional en el que se registran las transacciones del ciclo operativo de una empresa, ni se recopilan en sus aplicaciones “corporativas”. Este tipo resulta de gran valor para analizar comportamientos y preferencias, como, por ejemplo, la experiencia del cliente con un producto y sus opiniones

² Tecnologías de la Información y las Comunicaciones.

³ En la nube (o cloud computing) se refiere al almacenamiento en servidores accesibles a través de internet.

sobre el mismo. Ambos tipos de datos son susceptibles de adquirir un importante valor económico para las organizaciones. “El potencial empresarial de los datos escala a través de su característica inherente de no agotarse con su utilización. A diferencia de otros bienes económicos, los datos se pueden utilizar para múltiples propósitos al mismo tiempo” (Fadler & Legner, 2020).⁴

Con la confluencia de los factores mencionados, se dieron las condiciones necesarias para aplicar tecnologías que ya estaban anteriormente disponibles en el mercado, - como la inteligencia artificial y el análisis predictivo -, que se sumaron a los reportes operativos existentes y a las soluciones de *business intelligence* o inteligencia de negocios implementadas en las organizaciones, lo que permite no sólo extraer información de este conjunto de datos, sino, además, generar conocimiento y monetizarlo.

Por este motivo resulta sumamente importante aprovechar dichos datos y analizarlos para obtener información. Esta información no sólo va a permitir automatizar decisiones, o reducir los tiempos de los procesos, sino que también, entre otras cosas, permite personalizar la oferta de productos al momento de un contacto con el cliente o predecir su comportamiento futuro, lo cual constituiría un aporte de gran valor para las organizaciones, ya que pueden incrementar sus ventas, aumentar el éxito de sus campañas (de publicidad, de cobros, etc.) y reducir sus indicadores de incobrabilidad, reclamos, devoluciones, etc. También puede aplicarse en procesos de prevención de fraudes y procesos de auditoría continua, detección automática de anomalías, etc. Por esto, este gran conjunto de datos recopilado es procesado, almacenado, explotado y transmitido, a lo largo de toda la organización y de su cadena de valor, lo cual lo hace un activo crítico para su desempeño, y en el mejor de los escenarios, si la organización puede aprovecharlo, para la obtención de ingresos, a partir de su monetización.

Como señala Chinkes, “en la actualidad, en la mayoría de las organizaciones, prácticamente no existe aspecto sobre el cual haya que tomar decisiones del que no se estén guardando datos”. (2008, pág. 26)

Para explotar y obtener ventajas a partir de los datos, se utilizan las herramientas de inteligencia de negocios.

⁴ Traducción propia de: “*The business potential of data scales through its inherent characteristic of being nonrivalrous. In contrast to other economic goods, data can be used for multiple purposes at the same time*”

Según Saroka:

Hoy en día, es vital que las empresas provean un acceso rápido y efectivo a la información de negocios a muchos usuarios para sobrevivir. Para ello, la solución es un sistema BI⁵, que ofrece un conjunto de tecnologías y productos para hacer llegar a los usuarios la información que necesitan para tomar decisiones de negocios, tácticas y estratégicas (2002, pág. 86).

Cada una de las herramientas de inteligencia de negocios responde a distintos objetivos y satisface distintas necesidades de información, y cada organización implementará aquellas que considere apropiadas de acuerdo con la situación de su negocio, de su estrategia y de su cultura.

Los datos ingresan, se almacenan, procesan y se distribuyen a través de diferentes aplicaciones. La integración de éstas en las organizaciones es un problema común debido a los diferentes objetivos, lenguajes, estructuras de datos, procesos a los cuales soportan, hardware sobre el cual funcionan, etc., lo cual requiere de importantes esfuerzos de coordinación y de recursos informáticos, económicos y humanos.

De acuerdo con el marco de trabajo TOGAF 9.2®, propuesto por The Open Group, la arquitectura empresarial es un marco desde el cual analizar conjuntamente los componentes de dicha arquitectura (Negocios, Aplicaciones, Datos y Tecnología). Dentro de éstos, encontramos la arquitectura de aplicaciones y la arquitectura de datos.

La arquitectura de aplicaciones es definida como "un plano de las aplicaciones individuales a implementar, sus interacciones y sus relaciones con los procesos de negocio principales de la organización", mientras que la arquitectura de datos se define como: "la estructura de datos lógicos y físicos que posee una organización y sus recursos de gestión de datos".

⁵ BI es la sigla del término en inglés "Business Intelligence" y hace referencia a las soluciones de inteligencia de negocios.

Los datos tienen un ciclo de vida propio, dentro y fuera de la organización. La planificación de las actividades de recolección, almacenamiento, explotación y distribución de los datos permitirá una gestión adecuada durante todo el ciclo de vida de estos. Ambas arquitecturas (aplicaciones y datos) están fuertemente relacionadas. Existe una gran variedad de aplicaciones con diferentes características y objetivos a través de las cuáles los datos van a realizar un extenso recorrido.

El almacenamiento de datos no sólo sirve para tener soporte e historia de las transacciones de la empresa, como resguardo, o para obtener información agregada, a través de su procesamiento, para comparar de distintos períodos de tiempo, sino, que gracias a las herramientas de análisis masivo de datos o *Big Data Analytics*, se pueden analizar tendencias, y buscar patrones, con el objetivo de obtener predicciones acerca del comportamiento de determinadas variables “objetivo”.

De acuerdo con lo expresado por Mohapatra, Parisa y Banerjee:

Dado que, en los últimos años, el crecimiento de los datos almacenados se ha incrementado enormemente, el desafío consiste en descubrir cómo analizarlos, y encontrar patrones en ellos. Las fuentes de datos crecen no sólo en volumen, sino también en variedad y velocidad, lo que implica que crezca también el potencial de éstos para revolucionar el proceso de toma de decisiones en cada industria⁶. (2014, pág. 1)

Tal como exponen Larose & Larose, en su obra *Discovering knowledge in Data*, se utilizan indistintamente diversos términos para hacer referencia al análisis predictivo basado en datos masivos:

De acuerdo con el Grupo Gartner⁷, la minería de datos es el proceso de descubrir nuevas correlaciones, patrones y tendencias al tamizar grandes cantidades de datos

⁶ Extraído del artículo “The value of Big Data Analytics to the Business”, traducción propia de “Today, most organizations are facing the challenge of figuring out how to analyze data and find insights into newer and unconventional data types. As new sources of data continue to grow in volume, variety, and velocity, so does the potential of these data to revolutionize decision-making processes in every industry”.

⁷ Gartner Inc. es una empresa consultora y de investigación de las tecnologías de la información con sede en Stamford, Connecticut, Estados Unidos.

almacenados en repositorios, utilizando tecnologías de reconocimiento de patrones, así como técnicas estadísticas y matemáticas. Hoy en día, hay una variedad de términos utilizados para describir este proceso, que incluyen *Analytics*, *Predictive Analytics*, *Big Data*, aprendizaje automático y descubrimiento de conocimiento en bases de datos.⁸ (2014, pág. xi)

Dentro de las categorías de BI, la minería de datos o *data mining* se considera como una de las más complejas de utilizar e implementar, pero es aquella que está más íntimamente ligada a la generación de conocimiento. (Briano, Freijedo, Rota, Tricoci, & Waldbott de Bassenheim, 2011, Chinkes, 2008).

Estas técnicas de minería de datos pueden aplicarse tanto para el desarrollo de modelos descriptivos o de modelos predictivos. Estos últimos se diferencian de los primeros y de las otras herramientas de inteligencia de negocios, justamente por su objetivo. Mientras que con los modelos predictivos se busca predecir el comportamiento futuro de las variables analizadas, el resto de las técnicas y herramientas analiza los datos para exponer el pasado y el presente de la organización que los almacena. (Chinkes, 2008).

Mayer-Schönberger y Cukier aseguran que:

Los datos masivos suponen un paso importante en el esfuerzo de la humanidad por cuantificar y comprender el mundo. Una inmensa cantidad de cosas que antes no pudieron medirse, almacenarse, analizarse y compartirse están convirtiéndose en datos. El aprovechamiento de vastas cantidades de datos en lugar de una pequeña porción, y el hecho de preferir más datos de menor exactitud, abre la puerta a nuevas formas de comprender. (2013, pág. 347).

Existen diversas razones por las cuales contar con herramientas de análisis predictivo puede otorgar ventajas competitivas a las organizaciones. El análisis predictivo es parte de las herramientas de inteligencia artificial, cuya aplicación en las organizaciones también tiene

⁸ Traducción propia de “According to the Gartner Group, Data mining is the process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques. Today, there are a variety of terms used to describe this process, including analytics, predictive analytics, big data, machine learning, and knowledge discovery in databases”

un futuro muy promisorio gracias a los datos masivos, los cuáles les permiten a las computadoras contar con una fuente amplia para aprender, generar conocimiento y así mejorar el proceso de toma de decisiones.

Según indica Ramachandran, gracias a la predicción de eventos futuros, este tipo de análisis les permite a las organizaciones, una mayor precisión en toma las decisiones. Por ejemplo, al focalizarse en la identificación de segmentos de clientes, y en la predicción de sus comportamientos futuros, las empresas pueden basarse en los requerimientos del mercado y lograr que sus inversiones sean más rentables y productivas. (2017, pág. 1).

Un estudio realizado en 2011 revela que es probable que exista una conexión potencial entre el proceso de toma decisiones basado en datos y la productividad de las organizaciones (Brynjolfsson, Hitt, & Kim, 2011).

Cuando esto ocurre, es decir, cuando los procesos de toma de decisiones de una organización se sustentan en datos, se les da el nombre, de acuerdo con diversos autores, de “organizaciones basadas en datos” (*data-driven organizations*).

A menudo se afirma que debe establecerse una cultura basada en datos en toda la organización para explotar todo el potencial de la analítica avanzada (Anderson, 2015; Franks, 2014). Una cultura basada en datos está caracterizada por un proceso de decisión que enfatiza las pruebas y la experimentación, donde los datos superan opiniones, y donde se acepta el fracaso, siempre y cuando se aprenda algo de él (Berndtsson, Forsberg, Stein, & Svahn, 2018).

Esta situación de transformar a la organización en una “organización basada en datos” o “*data-driven*” requiere un esfuerzo y un cambio cultural muy importante, no sólo a nivel gerencial, sino también a nivel de procesos y procedimientos, e inversión en tecnología.

Esta transformación es transversal a toda la organización y sus áreas, y requiere de la existencia de ciertas condiciones o factores para sustentarla. Dentro de los factores habilitantes para implementar una cultura basada en datos, los cinco más importantes son: gestión, datos, herramientas, organización, y procesos de toma de decisiones (Berndtsson, Forsberg, Stein, & Svahn, 2018).

La organización debe centrarse en estos factores para desarrollarlos de tal manera que, en conjunto, puedan generar la sinergia necesaria que apalanque los procesos decisorios basados en datos.

Factores de la cultura basada en datos

Los autores citados, explican dichos factores de la siguiente manera:

Primer factor: Alta Dirección

La alta dirección tiene que dejar en claro la importancia de establecer una cultura basada en datos. Su apoyo es clave.

Segundo factor: Datos

El análisis avanzado que utiliza grandes volúmenes de datos es una opción disponible, pero no es mandatorio utilizar esta tecnología para contar con una cultura basada en datos; mientras que sí lo es contar con datos de alta calidad. La utilización de herramientas de minería de datos supone un esfuerzo adicional, debido a la variedad, volumetría, frecuencia y cantidad de fuentes. Es decir, una organización puede estar basada en datos, aunque sus herramientas de explotación sean básicas, siempre y cuando los datos que utiliza como materia prima para su análisis sean de calidad.

Tercer factor: herramientas informáticas

Para facilitar la adopción generalizada de la cultura analítica y basada en datos, la organización debe brindarles a sus colaboradores las herramientas informáticas adecuadas, de manera que los usuarios puedan realizar análisis descriptivos en forma autónoma. Esto requiere que los usuarios estén capacitados en herramientas de explotación y visualización de datos; que tengan disponible su uso y el acceso a los repositorios de información que necesitan.

Cuarto factor: Organización

Normalmente, las organizaciones cuentan con áreas de TI y con áreas de inteligencia de negocios, las cuales proporcionan informes estandarizados periódicamente. Al momento de incorporar un área de análisis avanzado de datos, surge el interrogante del lugar que debe ocupar dentro de la organización, en cuanto a su jerarquía, recursos, presupuesto, poder formal.

Quinto Factor: Proceso de decisión

Cómo se lleva a cabo el proceso decisorio es el factor que evidencia si la organización logró adoptar una cultura basada en datos o no. Como fue mencionado anteriormente, independientemente de las herramientas de explotación utilizadas, y la calidad de los datos utilizados, lo que determina si se logró implementar el cambio cultural es qué elementos sustentan la toma de decisiones: si la información o la intuición.

Sexto factor: Gestión del cambio

La gestión del cambio se puede describir como las herramientas y conceptos estructurados que se utilizan para lograr la voluntad de cambio en personas, organizaciones y corporaciones. Implementar una cultura basada en datos se trata realmente de cambiar la cultura organizacional. Es importante que el resultado de una cultura basada en datos esté arraigado en las operaciones diarias de los empleados (incluida la gerencia), por tanto, la gestión del cambio concierne a todos los facilitadores descritos anteriormente.

Berndtsson, Forsberg, Stein, & Svahn también proponen un modelo de madurez para medir el grado en que una organización alcanza este objetivo. En su modelo, distinguen cuatro niveles.

Niveles de madurez de la cultura basada en datos

Nivel 1:

En esta etapa, las organizaciones son bastante inmaduras en materia de recolección de datos y análisis de éstos, no tienen una unidad dedicada a Inteligencia de Negocios, y la responsabilidad recae en las unidades funcionales de negocio. Las cifras clave del rendimiento sobre lo que sucedió en el pasado son generalmente distribuidas a través de hojas de cálculo. Debido a la incerteza que rodea a dichos informes, las decisiones terminan basándose en instintos y no basadas en datos.

Nivel 2:

Las organizaciones en el nivel dos han dado un paso importante en la recopilación y análisis de datos históricos de manera sistemática. Para tener una sistematización sostenible en la recolección y análisis de los datos, se introduce una unidad dedicada al análisis de la información. Por lo general, existe un almacén de datos, desde el cual se reportan y se pueden generar cuadros de mando sobre el pasado.

Nivel 3:

Las organizaciones en el nivel 3 han pasado al análisis avanzado y utilizan herramientas de minería de datos para descubrir patrones previamente desconocidos. Normalmente, el uso de herramientas de minería de datos conlleva una cultura de aprendizaje de “prueba y error”, con respecto a cómo descubrir nuevos conocimientos empresariales. Este enfoque es diferente de la cultura de generar informes estandarizados y ad-hoc. Por lo tanto, no es raro que las organizaciones tengan unidades separadas para el análisis descriptivo y el análisis predictivo. En esta etapa, la mayoría de los empleados confían en el almacén de datos, - aunque no tanto en las herramientas de minería aún - y se utilizan herramientas de autoservicio para acceder a éstos.

Nivel 4:

En este nivel, las organizaciones cuentan con una unidad especial de análisis de datos para toda la organización, donde ésta es responsable de todo tipo de análisis. Hay un énfasis en la organización en la implementación de negocios lo antes posible, preferiblemente a través de decisiones (semiautomatizadas). Las herramientas de autoservicio para análisis

descriptivo y análisis predictivo están disponibles para todos los usuarios que trabajan con datos.

Factor	Nivel 1	Nivel 2	Nivel 3	Nivel 4
Organización	No existe área de inteligencia de negocios	Se establece una unidad de Inteligencia de Negocios	Las áreas de Inteligencia de Negocios y Analítica están separadas.	Se ha establecido un equipo de analítica extendido a toda la organización
Tecnología	Mayoritariamente hojas de cálculo	El datawarehouse está construido y en uso	Se utilizan el datawarehouse y herramientas de minería de datos	Implementación de negocios rápidamente
Proceso decisorio	Basado en experiencia/opinión del decisor	Los reportes y tableros de comando se generan automáticamente y a demanda	Cultura de prueba y error	Las decisiones son semi-automatizadas
Personas	Poca confianza en los datos y en analítica	Hay posiciones encontradas respecto del uso de analítica	El datawarehouse se explota con herramientas de autoservicio. Hay posiciones encontradas respecto del uso de analítica	Se utilizan herramientas de analítica en forma de autoservicio
Analítica	Descriptiva	Descriptiva	Descriptiva y predictiva	Descriptiva, predictiva y prescriptiva

La necesidad de las organizaciones de crear un entorno donde el proceso decisorio se sustente en datos, y que estos se aprovechen incluso para generar nuevos ingresos, genera la necesidad de adaptar las arquitecturas de datos pensadas para la explotación de los datos

netamente “corporativos” y plantea retos a la explotación de grandes volúmenes, así como también a la calidad y seguridad de la información, y la adopción de nuevas aplicaciones que permitan realizar una explotación más exhaustiva de los datos. Asimismo, crece también la complejidad de su administración, y su vulnerabilidad, que afectan a sus tres características fundamentales de la información; integridad, disponibilidad y confidencialidad. Debido que actualmente las organizaciones registran, almacenan y gestionan grandes volúmenes de datos a bajo costo y sin restricciones, se incrementa el riesgo para la seguridad de la información y la privacidad de las personas, lo que conlleva la necesidad de una administración eficiente del flujo de información a través de la organización y fuera de ella.

Debido a su calidad de activo estratégico y crítico, y la complejidad creciente de su gestión surge la necesidad de “gobernar” los datos dentro de la organización. El aumento de la volumetría de datos, la amplia variedad de herramientas de explotación de datos y la complejidad cada vez mayor de la arquitectura de aplicaciones y de datos de las organizaciones dificultan la administración del flujo de la información a lo largo de su ciclo de vida, y genera la necesidad de implementar mecanismos tendientes a mejorar su gestión.

Gobierno Corporativo, Gobierno de IT, Gobierno de Información y Gobierno de Datos

Según Weill y Ross, existen seis activos clave a través de los cuales las empresas logran sus estrategias y generan valor comercial: humanos, financieros, físicos o materiales, de información y tecnología, de propiedad intelectual y de relaciones (internas y externas).

Los autores hacen referencia a la definición propuesta por la Organización para la Cooperación y el Desarrollo Económicos, en la publicación de 1999; "Principios de la OCDE para la gobernanza empresarial", por la cual, “El gobierno corporativo proporciona la estructura para determinar los objetivos organizacionales y monitorear el desempeño para asegurar que se logren los objetivos”. En consecuencia, para gobernar la administración y el uso de cada uno de estos activos, se crean distintos mecanismos organizacionales tales como estructuras, procesos, comités, procedimientos y auditorías. En cuanto al Gobierno de TI, lo definen como la especificación de los derechos de decisión y el marco de responsabilidad para fomentar el comportamiento deseable en el uso de TI, es decir, los derechos de decisión y responsabilidad que determinan quién toma las decisiones sobre los activos de TI. Existen diversos marcos de trabajo relacionados con

distintas áreas de análisis, que guardan relación directa o indirecta con el Gobierno de Datos, tal como se enumeran en la figura 1.

Gobierno Corporativo	<ul style="list-style-type: none">• COSO• ERM
Gobierno de TI	<ul style="list-style-type: none">• COBIT• ISO/IEC 38505-1: 2017/38505-2:2018 (*)
Arquitectura Empresarial	<ul style="list-style-type: none">• Zachman Framework• TOGAF
Ciclo de vida de desarrollo de sistemas	<ul style="list-style-type: none">• RUP
Gestión de proyectos	<ul style="list-style-type: none">• PRINCE2• PMI PMBOK
Proceso de Desarrollo	<ul style="list-style-type: none">• CMMI
Gestión de servicios de TI	<ul style="list-style-type: none">• ITIL• ISO 2000

Figura 1. Marcos de Trabajos relacionados. Elaboración propia basada en información del framework DAMA y en Piorín, Normas y Marcos relacionados con la implementación de esquemas de Gobierno de la Tecnología de la Información. Buenos Aires (Piorín, 2016)

El Gobierno de TI está fuertemente relacionado con el Gobierno de Información y el Gobierno de Datos. En tal sentido, Cheong y Chang explican lo siguiente:

Se puede concluir que el éxito del marco COBIT depende de la calidad de los datos corporativos subyacentes. [...] Esto muestra que, además de la gobernanza de TI, existe la necesidad de un marco de gobernanza de datos para una gestión de datos eficaz. (2007)⁹

⁹ Traducción propia de: “It can be concluded that the success of the COBIT framework depends on the quality of the underlying corporate data. [...] This shows that in addition to IT governance there is a need for data governance framework for effective data management”

La norma ISO/IEC 38505-1: 2017 define al gobierno de datos como un subconjunto o subdominio del gobierno de IT, el cual, a su vez, es un subconjunto o subdominio del gobierno corporativo (ISO, 2021).

Según Smallwood:

La Gobernanza de Datos implica procesos y controles para garantizar que la información a nivel de datos sea verdadera, precisa y única. [..]. Se centra en la calidad de la información [..], de manera que los informes, análisis y conclusiones posteriores sean confiables. [..] Su objetivo es garantizar que los controles formales de gestión (sistemas, procesos y empleados responsables, custodios de los datos) se implementen para gobernar los activos de datos críticos para mejorar su calidad y evitar los efectos negativos resultantes de decisiones basadas en datos inexactos. (2014, pág. 191)

Para diferenciarlo del concepto de Gobierno de la información, dicho autor continúa:

La *Association of Records Managers and Administrators*¹⁰ (ARMA), define al Gobierno de la Información como “un marco estratégico compuesto por estándares, procesos, roles y métricas que responsabilizan a las organizaciones y a las personas de crear, organizar, proteger, mantener, usar y eliminar la información de manera que se alinee y contribuya a los objetivos de la organización (2014, pág. 6)¹¹.

Si bien, como se mencionó, éstos están fuertemente relacionados, existe una clara diferencia. Mientras que el Gobierno de Información se centra en “el control de la información generada por los sistemas” (2014, pág. 20)¹², la gobernanza de datos “consiste en los procesos,

¹⁰ Asociación de Administradores y Gestores de Registros.

¹¹ Traducción propia de: “*According to the Association of Records Managers and Administrators (ARMA), IG is a strategic framework composed of standards, processes, roles, and metrics that hold organizations and individuals accountable to create, organize, secure, maintain, use, and dispose of information in ways that align with and contribute to the organization’s goals.*”

¹² Traducción propia de: “*The IG approach to governance focuses not on detailed IT or data capture and quality processes but rather on controlling the information that is generated by IT and office systems*”

métodos y técnicas para garantizar que los datos sean de alta calidad, confiables y únicos (no duplicados), para que su uso posterior sea confiable y preciso” (2014, pág. 21)¹³.

Otra definición de Gobernanza de Datos, propuesta por DAMA International¹⁴, indica que se trata del ejercicio de autoridad y control (planificación, seguimiento y ejecución) sobre la gestión de los activos de datos. (2009, pág. 37).

En otras palabras, “La gobernanza de datos se refiere a quién tiene los derechos de decisión y es responsable de la toma de decisiones de una organización sobre sus activos de datos” (Khatri & Brown, 2010)¹⁵.

Las organizaciones necesitan, además, cumplir con normas de seguridad de la información y marcos legales vigentes que impactan en los procesos específicos de recolección, almacenamiento, procesamiento, actualización, eliminación y distribución de datos. La gestión de la información requiere de la existencia de un marco de trabajo, políticas, normas, procedimientos, estándares y mejores prácticas para procesar y explotar los datos “corporativos” y “estructurados” que se generan y fluyen dentro de la organización, asegurando su calidad, exactitud y no redundancia.

La gobernanza de datos especifica un marco multifuncional para la gestión de datos como un activo empresarial estratégico. Al hacerlo, la gobernanza de datos especifica los derechos de decisión y las responsabilidades para la toma de decisiones de una organización sobre sus datos. Además, la gobernanza de datos formaliza políticas, estándares y procedimientos de datos y supervisa su cumplimiento (Abraham, Schneider, & vom Brocke, 2019).

Por lo expuesto, la implementación de un programa de Gobierno de Datos en una organización debe estar en consonancia con un programa mayor de Gobierno de la Información, y a su vez, éstos no deben estar disociados del Gobierno de TI.

¹³ Traducción propia de: “*Data governance consists of the processes, methods, and techniques to ensure that data is of high quality, reliable, and unique (not duplicated), so that downstream uses in reports and databases are more trusted and accurate*”

¹⁴ Es una asociación internacional de profesionales de Gestión de Datos, que define estándares, mejores prácticas y certificaciones.

¹⁵ Traducción propia de: “*Data governance refers to who holds the decision rights and is held accountable for an organization’s decision-making about its data assets*”

Según Cheong y Chang:

La gobernanza de datos es importante porque define políticas y procedimientos para garantizar una gestión de datos proactiva y eficaz. La adopción de un marco de gobierno de datos también permite la colaboración desde varios niveles de las organizaciones para administrar los datos de toda la empresa y brinda la capacidad de alinear los programas de datos con los objetivos corporativos (2007).¹⁶

“Dado que los datos se han convertido en un activo cada vez más valioso para las organizaciones, también la presión para mejorar la gobernanza de datos dentro de las organizaciones está aumentando” (Nokkala & Salmela, 2019)¹⁷.

La Asociación Internacional DAMA, en su obra, DAMA – DMBOK Framework, define once “áreas de conocimiento” de la gestión de datos, donde distingue y enumera:

- Gobierno de datos: planificación, supervisión y control sobre la gestión de datos, su uso, y los recursos relacionados con éstos.
- Arquitectura de datos: la estructura general de los datos y los recursos relacionados con los mismos, como parte integral de la arquitectura empresarial.
- Modelado y diseño de datos: análisis, diseño, construcción, pruebas y mantenimiento.
- Almacenamiento y operaciones de datos: implementación de almacenamiento de activos de datos físicos estructurados y gestión.
- Seguridad de los datos: garantizar la privacidad, la confidencialidad y el acceso adecuado.
- Integración e interoperabilidad de datos: adquisición, extracción, transformación, movimiento, entrega, replicación, federación, virtualización y soporte operativo.
- Documentos y contenido: almacenar, proteger, indexar y permitir el acceso a los datos que se encuentran en fuentes no estructuradas (archivos electrónicos y registros

¹⁶ Traducción propia de: “*data governance is important because it defines policies and procedures to ensure proactive and effective data management. The adoption of a data governance framework also enables collaboration from various levels of the organizations to manage enterprise-wide data and it provides the ability to align various data related programs with corporate objectives*”

¹⁷ Traducción propia de: “*As data is becoming an increasingly valuable asset for organizations, pressure to improve its governance is increasing*”

físicos), y hacer que éstos estén disponibles para la integración e interoperabilidad con datos estructurados (bases de datos).

- Datos maestros y de referencia: gestión de datos compartidos para reducir la redundancia y garantizar una mejor calidad de los datos a través de la definición estandarizada y el uso de valores de datos.
- Almacenamiento de datos e inteligencia empresarial: gestionar el procesamiento de datos analíticos y permitir el acceso a datos de apoyo a la toma de decisiones para informes y análisis.
- Metadatos: recopilación, categorización, mantenimiento, integración, control, distribución y gestión de los metadatos.
- Calidad de los datos: definir, monitorear, mantener la integridad de los datos y mejorar la calidad de éstos.

Estas “áreas de conocimiento”, que, en conjunto, abarcan todo el espectro de la gestión de datos durante su ciclo de vida completo, se grafican, de acuerdo con este marco de trabajo, en una “rueda”, donde puede observarse que la Gobernanza de Datos, ocupa un lugar central, hacia el cuál convergen el resto de estas, tal como se ilustra en la Figura 2.

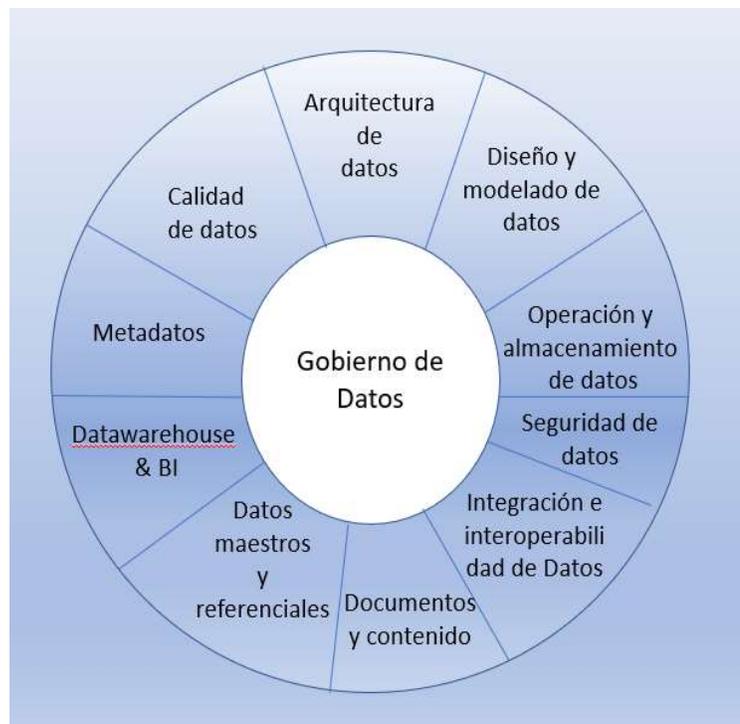


Figura 2. Traducción propia de la rueda del área de conocimiento de la guía DAMA-DMBOK2

La complejidad creciente de la administración de los datos dentro de la organización y la salvaguarda de sus aspectos fundamentales, la integridad, confidencialidad y disponibilidad ha hecho necesario que se complejicen también los procesos, los marcos de trabajo y las herramientas informáticas que acompañan dicha gestión.

En general, los marcos de gobierno de datos están pensados para el gobierno de los datos dentro de la organización; donde se pueden establecer roles claros y simples para la gestión de éstos; su almacenamiento, actualización y eliminación. En el caso de los ecosistemas de empresas, el gobierno de datos abarca o debe abordar relaciones e interacciones complicadas entre múltiples participantes. (Lee, Zhu, & Jeffery, 2019)

La influencia de la autoridad de la organización se limita al entorno interno. Los instrumentos tradicionales para la asignación de derechos de decisión y las responsabilidades en términos de acceso y uso de datos generalmente no van más allá de las fronteras organizacionales. (Lis & Otto, 2020)

Cheong & Chang, postulan que los datos se pueden administrar de manera más efectiva y exitosa mediante la adopción de una estructura y un marco de gobierno de datos.

De acuerdo con diversos autores, citados en su investigación, estos problemas con la gestión de datos en las organizaciones surgen a partir la implementación de proyectos de integración de datos, ya sea por la integración de sistemas legados o por la creación o generación de repositorios de datos centralizados para su explotación (2007).

Con la importancia del crecimiento exponencial de la información en la era de Big Data, la gobernanza de datos ha surgido como un subconjunto importante de la gobernanza de TI y se define como la capacidad de una empresa para orquestar todos los recursos relevantes con el fin de maximizar el valor de la información y la generación de conocimientos para la organización (Mikalef & Krogstie, 2018).

Las prácticas de gobierno de datos de *big data* tienen implicaciones importantes para la disponibilidad, la calidad, el mantenimiento y la seguridad de la variedad de

fuentes de datos estructurados y no estructurados que forman parte de la cadena de valor de la información (Abbasi, Sarke, & Chiang, 2016).

Las demandas de incorporar datos externos, plantea nuevas dificultades con la calidad de los datos. Hoy en día, los datos son de todo tipo de formas, con diferentes frecuencias de actualización, y volúmenes masivos lo cual requiere de una alta calidad y sólida gobernanza de datos. (Berndtsson, Forsberg, Stein, & Svahn, 2018).

Según Otto: “El valor de los datos depende de su calidad. El gobierno de datos proporciona pautas y reglas sobre la calidad de los datos y la gestión de la calidad de éstos. (Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers, 2011)”. De acuerdo con este autor, y tal como explica en el gráfico presentado a continuación, la gestión de los datos es regulada por el gobierno de datos, que se aplica sobre éstos para asegurar su valor, el cual está apoyado en su calidad, en un círculo virtuoso.

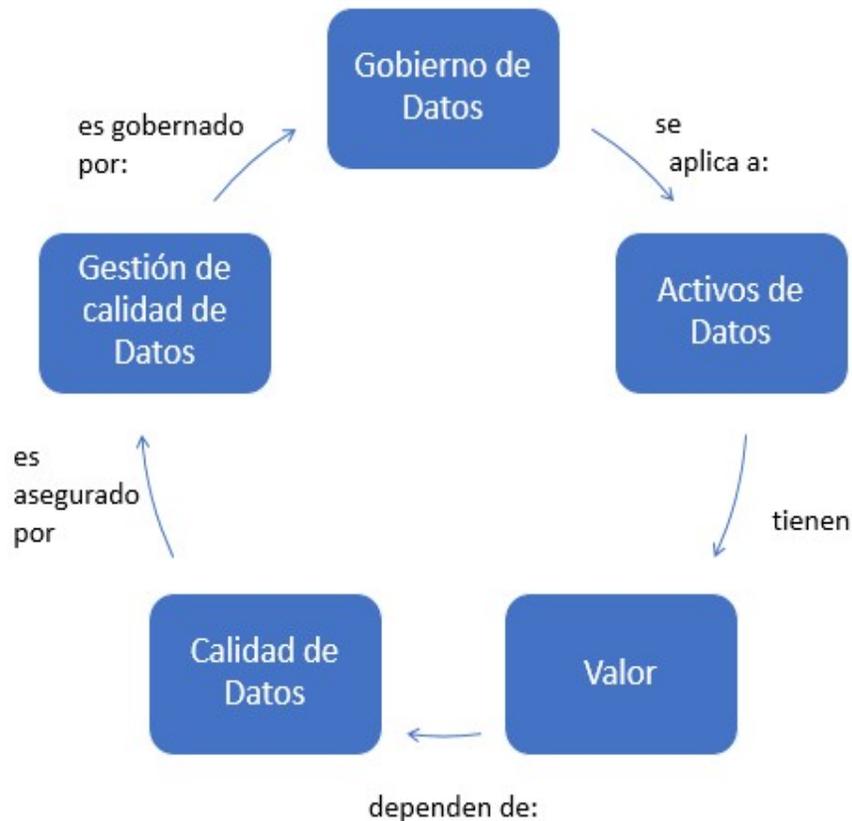


Figura 3 - Traducción propia de Gobierno de Datos y conceptos relacionados (Otto, 2011)

Más precisamente, “la gobernanza de datos representa la función principal de la gestión de datos, ya que especifica qué decisiones deben realizarse en la gestión de datos y quién toma estas decisiones” (Otto, Data Governance, 2011)¹⁸. Su objetivo principal es maximizar el valor de los activos de datos, a través de funciones como la gestión de los datos y la gestión de la calidad de datos:

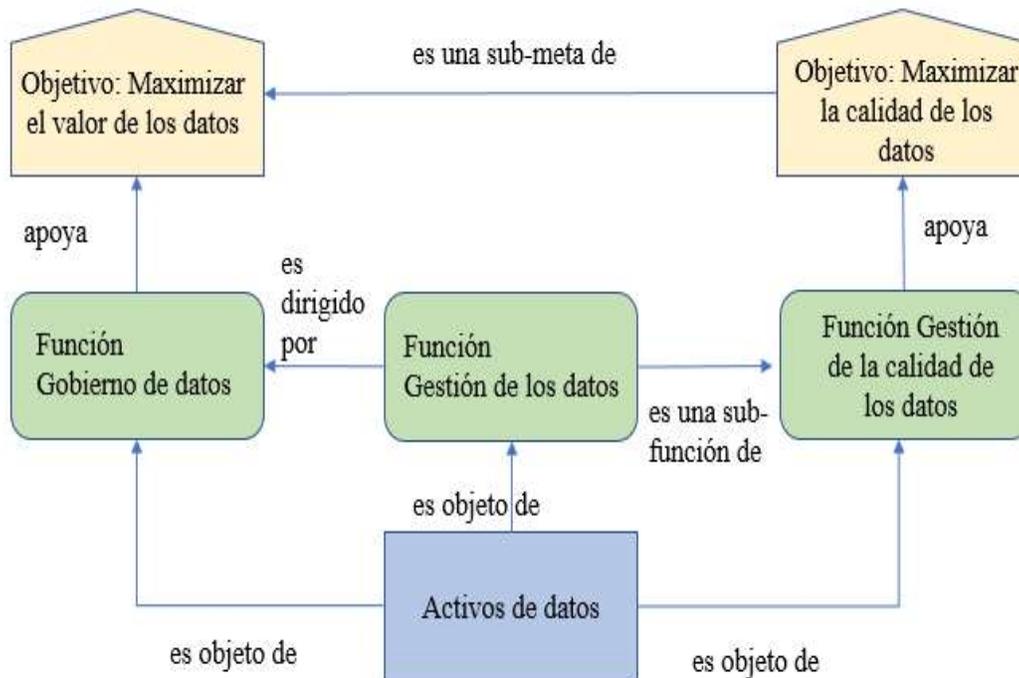


Figura 4 - Traducción propia de Conceptos Fundamentales (Data Governance, Otto, 2011)

Morfología y componentes del Gobierno de Datos

Adicionalmente, Otto, propone una morfología cuyo objetivo es determinar cómo organizar el Gobierno de datos dentro de una organización. Al momento de planificar cómo implementar esta función, distingue entre dos dimensiones para tener en cuenta: los objetivos organizacionales y la estructura. Dentro de los objetivos enuncia dos grandes grupos, los formales y los funcionales. Los objetivos formales, según el autor, se dividen en los objetivos puramente del negocio y los objetivos de las áreas de sistemas y tecnología. Los primeros

¹⁸ Traducción propia de: “Data Governance represents the leading function of data management as it specifies which decisions need to be made in data management and who makes these decisions”

centrados en el “para qué” (por ejemplo, incrementar la eficiencia operativa) y los segundos en el “cómo” (por ejemplo, aumentar la calidad de los datos). Los objetivos funcionales son aquellos propios de un programa de Gobierno de Datos: definir la estrategia de datos de la organización y sus políticas de datos, establecer el control de la calidad de estos, definir la custodia y propiedad de éstos, y definir estándares y establecer la gestión de su ciclo de vida y su arquitectura. En cuanto a la estructura organizativa, identifica tres aspectos sobre los cuáles la organización debe decidir: el locus de control, a partir de la definición de las posiciones dentro de la estructura, tanto a nivel jerárquico, como de perfiles (técnicos o funcionales), las formas de organizar estas funciones (de manera centralizada, descentralizada, etc.) y en tercer lugar, los roles a ejercer y los comités a formar (A morphology of the organisation of data governance, 2011).

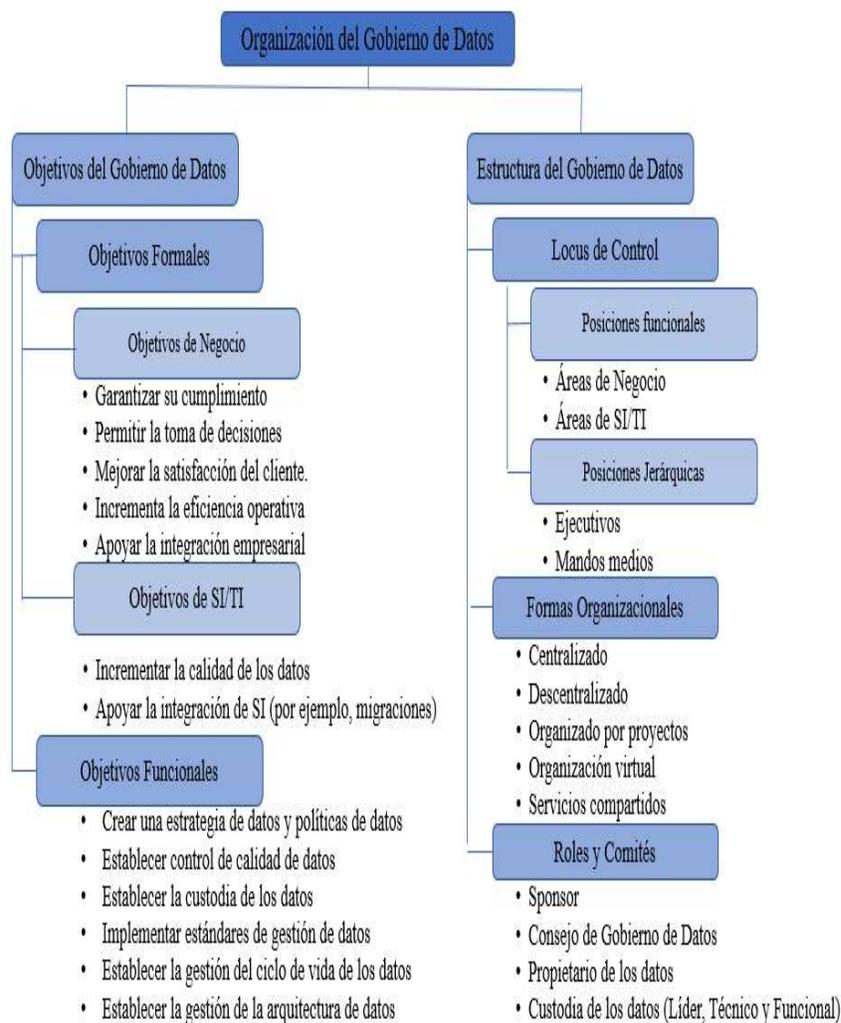


Figura 5 - Traducción propia de Morfología de la organización del Gobierno de Datos (Otto, 2011)

Seguidamente, este mismo autor, propone un marco de trabajo conceptual, donde al analizar al gobierno de datos desde la óptica del diseño organizacional, incorpora un tercer aspecto: la transformación que la organización deberá llevar a cabo para la adecuación de sus procesos y el cambio cultural requerido.

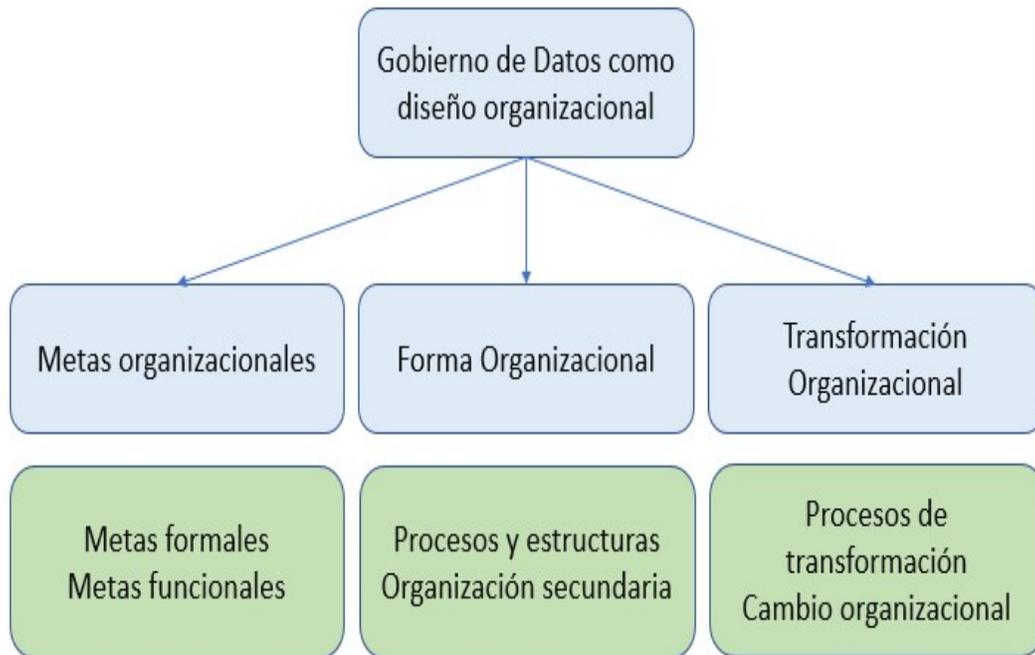


Figura 6 - Traducción propia del Marco de trabajo conceptual de la organización de la gobernanza de datos (Otto, 2011)

Otro modelo, propuesto por Cheong & Chang (2007), y basado en un caso de estudio, detalla los siguientes componentes de un marco de Gobierno de Datos:

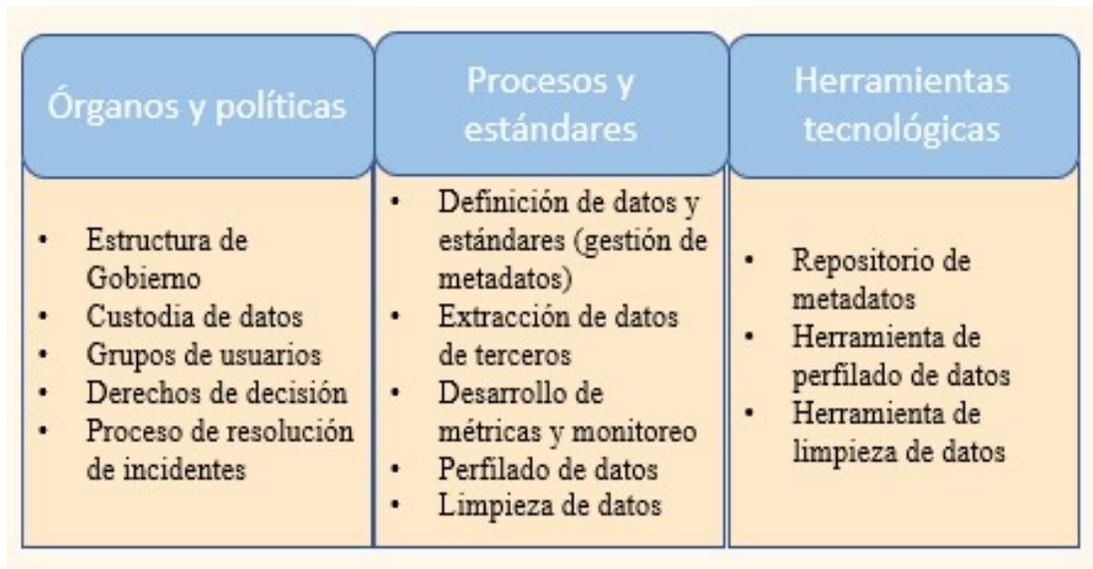


Figura 7 - Traducción propia del Marco de Gobierno de Datos (Cheong & Chang, 2007)

A este modelo, que también contempla la estructura organizacional, roles y procesos se suma la utilización de herramientas informáticas que permitan realizar las distintas tareas necesarias para asegurar la calidad de datos y su gestión, como, por ejemplo, la limpieza, el perfilado, la gestión de metadatos, y su linaje, entre otras.

Mecanismos de Gobernanza

Los componentes del gobierno de datos pueden articularse a través de distintos tipos de mecanismos. Según un estudio realizado por Baijens, Helms y Velstra: “La literatura especializada en gobernanza de datos y TI muestra que la gobernanza se puede implementar a través de varios tipos de mecanismos de gobernanza: estructurales, procedimentales y relacionales” (2020)¹⁹.

¹⁹ Traducción propia de: “Data Governance literature shows that governance can be implemented through several types of governance mechanisms: structural, procedural and relational mechanisms.”

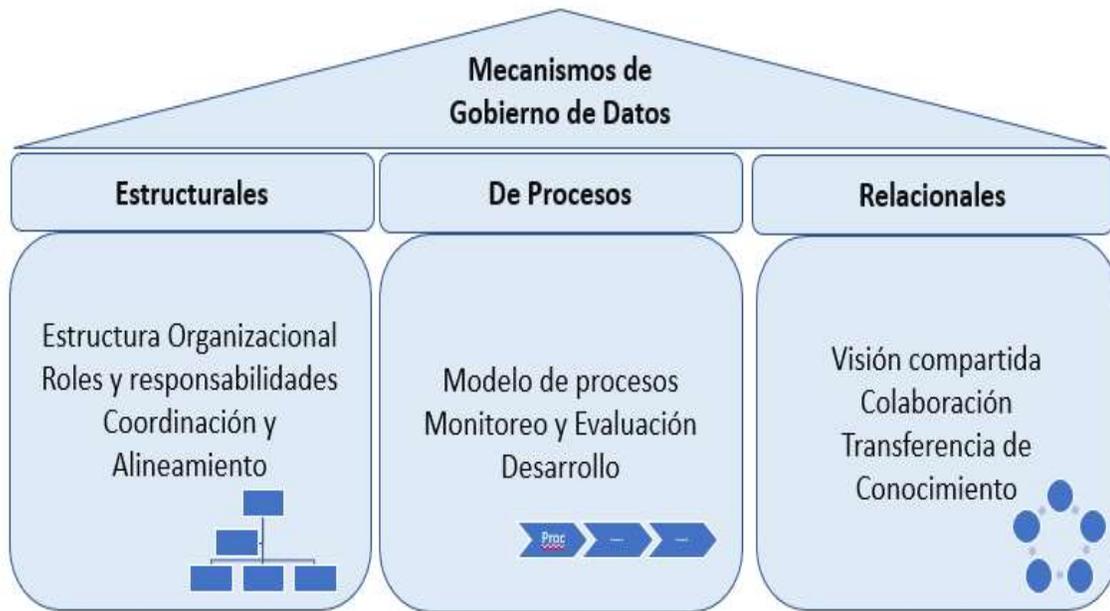


Figura 8 - Traducción propia de los mecanismos de gobiernos de datos propuestos por Baijens, Helms & Velstra

Mecanismos estructurales

Estructura organizacional

Dentro de los mecanismos estructurales, respecto de la estructura organizativa, Grossman y Siegel identifican la existencia de tres alternativas estructurales para la incorporación de la función de análisis de datos en la organización. En algunos casos se opta por un modelo centralizado, donde todas las funciones se colocan en una misma área o unidad. En otros, las funciones de análisis se distribuyen en diferentes unidades o áreas, es decir, de forma descentralizada, y, por último, una combinación de ambas, lo que sería un modelo híbrido. (2014).

Según los autores citados, el modelo centralizado tiene la ventaja de poder lograr en forma más fácil la masa crítica, acceder a los datos necesarios, impulsar una infraestructura integrada y obtener la experiencia necesaria para probar e implementar de manera eficiente varios análisis estadísticos, predictivos y modelos de minería de datos. Sin embargo, en este modelo, los científicos de datos pueden estar muy lejos de las unidades de negocio a las que se supone que deben apoyar. El desafío en este tipo de estructuras es que los científicos de

datos comprendan las distintas unidades de negocio y sus necesidades. Otra cuestión para tener en cuenta es la dependencia jerárquica que el área debería tener.

Por el contrario, señalan que optar por un modelo descentralizado podría solucionar los problemas de colaboración entre los científicos de datos y las áreas de negocio (Grossman & Siegel, 2014).

La desventaja de la descentralización de la función de análisis avanzado de datos es que al depender funcionalmente del área usuaria, los científicos de datos puedan tener una visión parcializada de la organización y su contexto, podrían crearse “silos” de información, complicaciones en el acceso a las fuentes, falta de trabajo colaborativo entre ellos e ineficiencias por trabajar aisladamente, sin aprovechar la sinergia de un equipo centralizado, y dispersión entre los conocimientos y las aplicaciones utilizadas en caso de carecer de un área de arquitectura fuerte que vele por las herramientas y los modelos de datos. Además de quedar la priorización de proyectos supeditada a las necesidades inmediatas del área funcional de la cual dependen, en detrimento de proyectos que supongan mejoras técnicas como, por ejemplo, de infraestructura o performance.

En tercer lugar, se plantea un modelo híbrido, que combine ambas formas organizativas. (2014). Todas estas opciones poseen ventajas y desventajas.

Roles y responsabilidades

Dentro de los mecanismos de gobernanza estructural, con la definición de posiciones y roles se establecen responsabilidades para la toma de decisiones relacionadas con cada dominio perteneciente al área de interés donde se aplica. Como parte del gobierno, una empresa primero define qué decisiones deben tomarse. La literatura sobre gobernanza de datos presenta diferentes áreas de decisión que son más operativas que las áreas de decisión para el gobierno de IT (Fadler, Lefebvre , & Legner , 2021).

En este segundo mecanismo estructural, además de las posiciones ya conocidas como líder de proyecto o usuarios de negocios, por ejemplo, surgieron puestos específicos como científico de datos, arquitecto de datos (Grossman & Siegel, 2014) y analista de datos; todos ellos con algún grado de conocimiento en herramientas de análisis y explotación de datos, estadística y bases de datos.

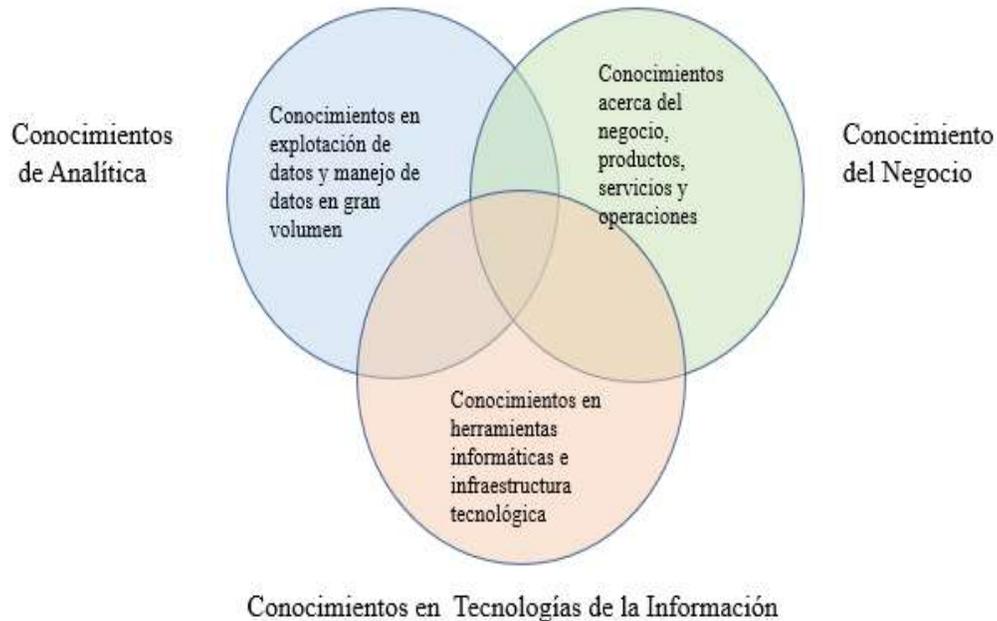


Figura 9. Traducción propia de: El conocimiento requerido por los científicos de datos (Grossman y Siegel, 2014)

Además de los puestos mencionados, que son especialistas de índole más técnico, surgen también dos roles relacionados con los datos: su propiedad y custodia y/o gestión.

Propiedad de los datos

Fadler y Legner realizaron un estudio para analizar las particularidades de la propiedad de los datos en contextos de big data analytics, debido a que en estos casos es más complejo determinar en quiénes debería recaer la responsabilidad sobre los mismos, en comparación a un entorno tradicional de bases y almacenes de datos.

La propiedad de los datos es un pilar fundamental del gobierno de datos, ya que éste especifica los derechos de decisión y las responsabilidades sobre éstos. Este concepto de propiedad comenzó a discutirse desde el comienzo del procesamiento de datos, pero a raíz de la difusión del análisis predictivo aplicado a los datos masivos, adquiere mayor relevancia, y se vuelve más desafiante y complejo, porque en la mayoría de los casos, se

desconoce el futuro del producto de esos datos y su utilización, sobre todo, si además se utiliza inteligencia artificial.

Según los autores citados, el propietario de los datos, - entendido como el propietario de la fuente-, es responsable de éstos como entrada, y el propietario del producto de datos, - entendido como quien los demanda-, es responsable de garantizar el valor comercial del resultado obtenido a partir de éstos. En un entorno de almacenamiento de datos tradicional, el otorgamiento de accesos y los requisitos de calidad son conocidos y pueden establecerse desde un principio. Sin embargo, en el entorno del análisis predictivo de datos masivos, el propósito de los datos se desconoce intencionalmente para permitir la exploración y experimentación de datos, por lo tanto, esto puede ser diferente; es el propietario de los datos quien debe administrar el acceso a los datos y debe reaccionar continuamente a los requisitos de datos cambiantes. Esto requiere, además, de conocimiento técnico específico de ambas partes, por lo tanto, sugieren también la participación de IT con una función mediadora entre ambas (2020).

Custodia de los datos

Según Cheong & Chang, los activos de datos son administrados por el custodio de datos. Es responsable también de su calidad. Su función consiste en resolver los problemas planteados por los usuarios, en respaldar los planes de gestión de datos, de limpieza de datos, de garantizar que los datos sean adecuados para su propósito, y de convertir los planes estratégicos en planes tácticos, además de la gestión de cambios y la gestión de partes interesadas.

Instrumentos para la designación de roles y responsabilidades

Matriz RACI

Según Otto, “varios autores (Loshin 2008, págs. 33 y sigs.; Weber et Alabama. 2009) proponen utilizar la matriz RACI”, para la asignación de roles y actividades dentro del Gobierno de Datos. Esta matriz define cuatro roles principales para cada conjunto de actividades: “aprobador”, “responsable”, “consultado”, “informado”. A continuación, se muestra un ejemplo de este instrumento:

Áreas de decisión	Roles	Dueño del proceso de negocio	Experto en el proceso de negocio	Equipo de estándares de datos	Usuario clave de los datos	Usuario final
Gestión de datos		Aprobador	Informado	Informado	Informado	Informado
Definición y estandarización de datos maestros		Aprobador	Responsable	Informado		
Principios y pautas		Aprobador	Responsable	Informado	Informado	Informado
Mantenimiento de datos		Consultado	Aprobador	Responsable	Responsable	Informado
Gestión de metadatos y documentación de datos maestros			Aprobador	Responsable	Responsable	Informado
Medición de calidad de datos			Consultado	Aprobador/Responsable	Aprobador/Responsable	

Figura 10. Traducción propia de un ejemplo de matriz RACI citado por Boris Otto en Data Governance, 2011

Coordinación y alineamiento

Como tercer mecanismo estructural, para la coordinación y alineamiento del área de Analítica con los objetivos corporativos, Grossman y Siegel (2014) sugieren la conformación de diversos comités, a saber:

- un comité de gobierno de datos que incluya a la alta dirección, a representantes de la organización de TI y diversas partes interesadas del negocio, con la función de priorizar los proyectos; y obtener los recursos y objetivos necesarios.
- otro comité de políticas técnicas que determine qué datos, aplicaciones, procesos, mejores prácticas y los estándares a utilizar en toda la organización.

- un comité de cumplimiento y seguridad
- un comité de gestión y calidad de datos que garantice que los datos y metadatos de la organización sean precisos, completos y coherentes.

Cheong & Chang señalan como mecanismo adicional de coordinación, la realización de reuniones de grupos de usuarios clave y personal de TI, donde se puedan plantear y resolver temas operativos urgentes relacionados con la gestión de los datos y solicitud de nuevos requerimientos que faciliten la gestión de estos. La figura a continuación muestra un esquema donde se espera que las áreas de negocios y de tecnología se mantengan informados mutuamente y alineados con los objetivos de gobierno de datos en los tres niveles jerárquicos: estratégico, táctico y operativo. (2007)

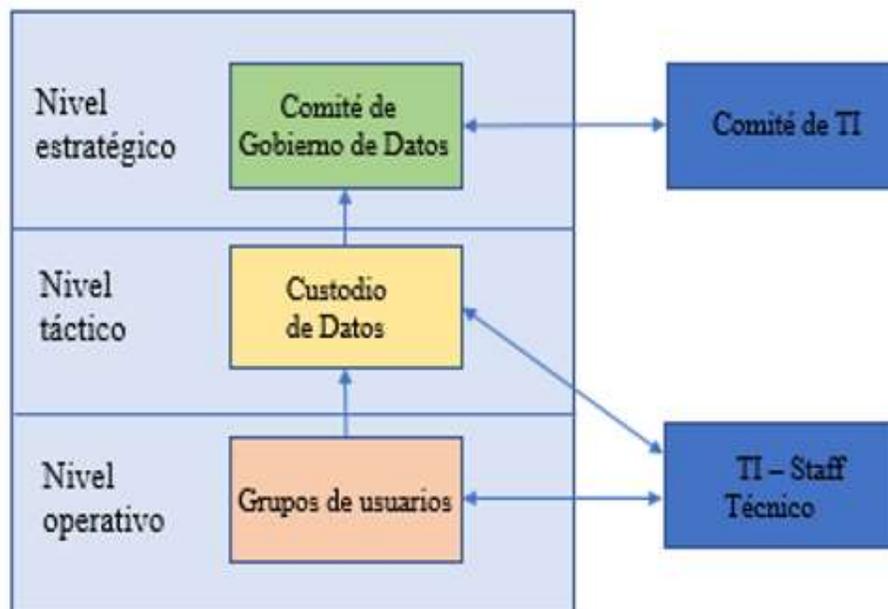


Figura 11. Traducción propia de la estructura de Gobierno de Datos, Cheong & Chang, 2007.

En la medida en que esta forma de organizar el trabajo conjunto madure puede escalar e incluir a otras áreas, designado custodios de datos y grupos de usuario de acuerdo a los distintos dominios de información, lo que resultaría en un modelo de gobierno de datos “federado”, tal como se muestra en la siguiente figura:

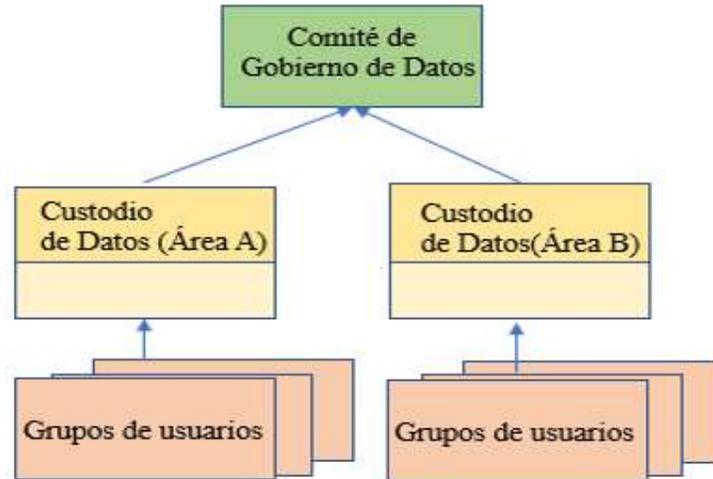


Figura 12. Traducción propia de la estructura escalable de Gobierno de Datos, Cheong & Chang, 2007.

Estos autores también señalan que debe existir una estrecha relación entre el Comité de Gobierno de datos y el Gobierno de TI para garantizar la confiabilidad de la información de la organización, y ambos deben informar y estar alineados con el Comité de Gobierno Corporativo (2007).

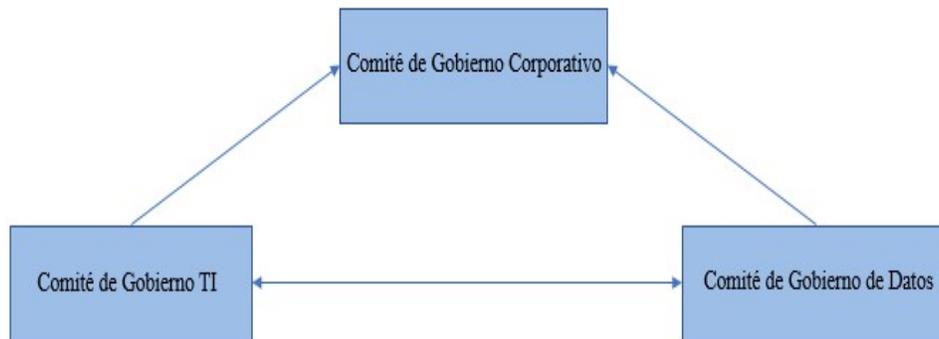


Figura 13. Traducción propia de la Relación entre el Gobierno Corporativo, de IT y de Datos, Cheong & Chang, 2007.

Mecanismos de Procesos

“Los procesos de alineación son técnicas de gestión para asegurar una participación amplia y eficaz en decisiones de gobernanza y su implementación”. (Weill & Ross, 2004)²⁰. “Los mecanismos de procedimiento complementan las estructuras organizativas y los roles al definir cómo se toman decisiones.” (Fadler, Lefebvre , & Legner , 2021) ²¹

“La gestión por procesos proporciona una capacidad de gobierno poderosa en todos los niveles” (Maiante, 2016). Esto es así porque este mecanismo permite determinar cuáles son las acciones por realizar por cada participante del proceso, y quién es responsable de estas; lo cual supone una estrecha relación con el mecanismo de gobierno de roles y responsabilidades.

Las estrategias de modelado representan un esfuerzo por la descomposición de cada uno de los procesos de negocio, organizados y/o priorizados en función de las necesidades de negocio.

La necesidad de “descomposición y/o fragmentación de procesos” se refiere al establecimiento de un nivel de agrupación y/o descomposición de los procesos de negocio para facilitar su entendimiento y su tratamiento.

²⁰ Traducción propia de: “*Alignment processes are management techniques for securing widespread and effective involvement in governance decisions and their implementation*”

²¹ Traducción propia de: “*procedural mechanisms complement organizational structures and roles in defining how decisions are made*”

Se comienza por el nivel uno, que representa los Macroprocesos, el cual grafica los procesos centrales de la organización, a grandes rasgos y a muy alto nivel, mencionando “acciones” genéricas. A continuación, el siguiente nivel, el dos, representa los grupos de procesos. De esta forma se agrupan los principales procesos que contiene cada macroproceso. Seguidamente, en el nivel tres, se representan los distintos escenarios de procesos, e implica distinguir las distintas asignaciones generales que son realizadas en la organización. En el nivel cuatro, se determinan las distintas variantes de procesos necesarios. El último nivel, el número 5, corresponde al procedimiento detallado a efectuar para cada nivel cuatro.



Figura 14., Modelado de Procesos tomado del material provisto por el profesor Hernán Maiane en las clases 2016

El modelo del grado de madurez de procesos de negocio, CMMI, (por sus siglas en inglés, Capability Maturity Model Integration) establece un grado de madurez, en base al estado de situación de cada proceso; contemplando desde la inexistencia de éste, lo que sería el grado 1 o inicial, el grado 2, cuando éste es repetitivo, el grado 3 cuando el proceso está formalmente definido, grado 4, cuando administrado y el grado más alto, cuando el proceso está optimizado. De esta manera, se pueden analizar los procesos existentes en la organización para determinar su estadio y los pasos a seguir para su mejora y evolución.

(Maiane, 2016)

Otro método de gestión que puede utilizarse para la definición de procesos es el conocido como el método del círculo P-D-C-A (Plan - Do - Check - Act) popularizado por el Dr. W. Edwards Deming y ampliamente adoptado en varios países, o el ciclo DMAIC (Define, Measure, Analyze, Improve y Control) en la metodología de Seis Sigma. (Maiane, 2016).

Dentro de este conjunto de mecanismos, Baijens, Helms y Velstra distinguen tres grupos: los modelos de proceso, seguimiento y evaluación, y desarrollo (2020).

En primer lugar, los modelos de proceso respaldan una forma estructurada y controlada de llevar a cabo proyectos de análisis de datos. Se encuentran disponibles diversos conjuntos de modelos de procesos que conducen a la implementación de resultados de análisis de datos, el más conocido es el modelo CRISP-DM, que proporciona un conjunto de pasos y tareas que se deben realizar para ofrecer conocimientos analíticos de datos (Mariscal et al., 2010). [...] En segundo lugar, comprender y mejorar el nivel de consistencia de los procesos requiere monitoreo y evaluar la eficiencia y eficacia de la analítica de datos. [...] Finalmente, debe existir una hoja de ruta de desarrollo para garantizar que el análisis de datos se desarrolle hacia los objetivos que se persiguen como organización. (2020)

Mecanismos Relacionales

Dentro de los mecanismos de gobernanza relacional, Baijens, Helms y Velstra identifican tres ejes principales: las percepciones compartidas, la colaboración y la transferencia de conocimientos y experiencia. Los resultados del análisis avanzado de datos, especialmente de los primeros, pueden ser inciertos, y de difícil interpretación y/o aplicación, por lo que resulta importante que los miembros de la organización compartan esa percepción acerca de las actividades de análisis de datos. A su vez, estas actividades se llevan a cabo con trabajo multidisciplinario, lo que requiere altos niveles de coordinación y comunicación entre los distintos participantes del proceso (Baijens, Helms, & Velstra , 2020). Respecto de las habilidades requeridas, para las organizaciones es importante prestar especial atención a la gestión del conocimiento, y contar con herramientas informáticas que permitan el resguardo de al menos, parte del conocimiento explícito generado durante los procesos de gestión de datos.

Arquetipos de Gobernanza

Los componentes y mecanismos explicados permiten la elaboración de arquetipos sobre su implementación en las organizaciones.

Fadler, Lefebvre , & Legner realizaron un estudio a través del cual identifican tres arquetipos de gobernanza de datos, los cuales caracterizan las formas típicas de gobernar este activo de acuerdo con los mecanismos de gobernanza estructural, procedimental y relacional implementados. El arquetipo I se concentra en mejorar la calidad de los datos maestros, el arquetipo II se caracteriza por habilitar la gestión de datos en toda la empresa y el arquetipo III enfatiza la coordinación de la red para habilitar la monetización de los datos. (2021)

En el cuadro que se presenta a continuación, se resumen las características de cada arquetipo enunciadas por los autores citados:

Mecanismos de Gobierno de Datos		Arquetipos		
		I - Calidad	II - Gestión	III - Monetización
Estructurales	Organización y roles	Equipo central con foco en pocos dominios	Equipo central con más miembros, con un alcance mayor y roles descentralizados.	Equipo centralizado con foco estratégico y amplia red de roles descentralizados.
	Responsabilidades centralizadas	Requerimientos, métodos, estándares, pautas, calidad de datos. Gestión del ciclo de vida.	Estrategia, Métodos, Estándares y pautas de datos, Calidad de datos. Arquitectura de datos.	Estrategia organizacional, métodos, estándares y pautas de datos, coordinación de comunidad de datos.
	Responsabilidades descentralizadas	Gestión del ciclo de vida del dato.	Requerimientos del Negocio. Gestión del ciclo de vida del dato.	Estrategia de dominio específico. Requerimientos de negocio. Métodos, estándares y pautas. Gestión del ciclo de vida de datos.
De Procesos	Estrategia y planes	Estrategia de datos: foco en datos maestros. Alcance: calidad de datos para mejorar costos	Estrategia emergente en toda la empresa. Alcance más amplio con foco en la disponibilidad y acceso a los datos	Estrategia de datos en toda la empresa, con foco en dominios específicos para habilitar la monetización de datos
	Inversiones	En mejorar la calidad de datos y en infraestructura	En mejorar la calidad de datos y en infraestructura	En oportunidades de monetización de datos e infraestructura
	Oportunidades de negocios	Sin proceso explícito	Para nuevos dominios de datos	Identificación proactiva de oportunidades de monetización de datos
	Monitoreo y Evaluación	Monitoreo de Calidad de datos maestros	Monitoreo de la Calidad de un mayor rango de datos	Monitoreo de Calidad de datos. Evaluación de su valor y uso.
	Gestión del ciclo de vida	Establecido	Establecido, centralizado y descentralizado	Gestionado por las unidades de negocio
	Comunicación	Estándares de datos. Normativa	Estándares de datos. Normativa. Pautas y Roles.	Estándares de datos. Roles y responsabilidades
Relacionales	Alineamiento y coordinación	Principalmente con IT, mediante reuniones. Con el negocio mediante funciones	Con IT. Principalmente, alineamiento operacional con las áreas de negocio.	Alineamiento estratégico y operativo con el negocio. Gestión comunitaria
	Transferencia de conocimiento	Uso de datos	Entrenamiento en métodos y datos	Entrenamiento en métodos, datos y conocimiento

Figura 15. Arquetipos de Gobierno de Datos. Fuente: Fadler, Lefebvre, & Legner (2021)

Factores críticos de éxito del Gobierno de Datos

La implementación del gobierno de datos en una organización conlleva un gran esfuerzo de sus participantes, ya que quienes habitualmente trabajan en la gestión de datos a lo largo de su ciclo de vida, ven incrementadas sus tareas, modificados sus niveles de autorización, permisos y crecen los controles sobre los mismos, además de un alto grado de coordinación y alineación entre los participantes del proceso.

En base a una investigación de Cheong & Chang, estos autores identifican y describen diez factores críticos de éxito para implementar exitosamente el gobierno de datos en una organización:

1. Definición clara de los roles y responsabilidades de las personas involucradas en el proceso de gobernanza de datos.
2. Definición de estándares de datos corporativos para asegurar que sean "adecuados para su propósito".
3. Alineación de la tecnología, los procesos y los órganos corporativos con los objetivos de la organización.
4. Gestión adecuada de los interesados (generadores y consumidores de datos).
5. Diseño de una estructura de gobierno de datos que incluya la participación de todos los niveles de la organización.
6. Definición de métricas de calidad de datos específicas para medir el éxito de la gobernanza de datos.
7. Exigencia de calidad a socios. Si la organización comparte datos con otras organizaciones (socios), existe un requisito indispensable de que su(s) socio(s) sea(n) responsable(s) de la calidad de sus datos.
8. Implementación de controles de la calidad de los datos.
9. Evaluación y seguimiento periódico del cumplimiento de las políticas y los procedimientos de gestión de datos.
10. Formación, comunicación y concientización del valor de la gobernanza de los datos, y la importancia y beneficios de la calidad de los datos.

Herramientas informáticas para Gobierno de Datos

Las actividades del gobierno de datos están actualmente soportadas por herramientas informáticas que permiten realizar limpieza de datos, administración (“stewardship”), linaje, y gestión de metadatos, entre otras funciones.

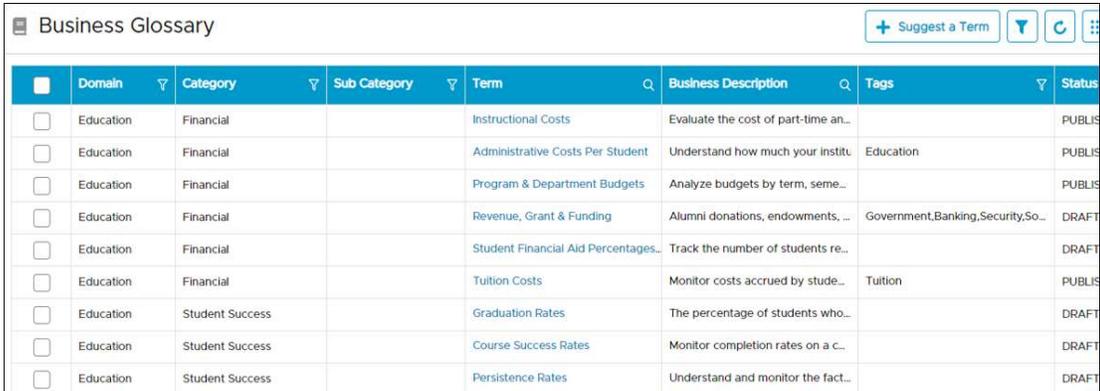
Funcionalidades principales

Linaje

La necesidad de contar con información acerca de los datos hace necesario contar con herramientas informáticas para administrar este flujo. La trazabilidad del dato desde su origen hasta su consumo se conoce con el nombre de “linaje”. Esta funcionalidad permite visualizar rápidamente el trayecto que estos recorren desde su fuente de origen, los sistemas sobre los cuales circula, los repositorios donde es almacenado, y finalmente, su destino. En otras palabras, el linaje muestra el recorrido del dato, su ciclo de vida a lo largo de la organización. Durante éste, los datos pueden ir modificando sus nombres de acuerdo con los sistemas donde son recolectados, procesados, almacenados y distribuidos, pueden modificarse los tipos de datos, e incluso su contenido debido a las reglas de negocio que se aplican en cada sistema, y a las reglas de transformación de cada repositorio integrado en el que son almacenados, lo cual hace muy complejo su rastreo. Por ejemplo, supongamos un campo “STATUS” de la tabla “CUSTOMER” del CRM, que contiene los valores numéricos 0, que significa “inactivo”, y 1, que significa “activo”. En el ERP, el campo que contiene la indicación del estado del cliente se llama “ESTADO”, y contiene los campos de tipo alfanuméricos, cuyos valores son “H”, “habilitado” y “D”, “deshabilitado”. Luego de la correspondiente integración, transformación y carga, dentro del almacén de datos corporativo, el campo que indica el estado de los clientes se llama “ID_ESTADO”, cuyos códigos son 1, para “Activo”, y 2 “No activo”. En todos los casos, el dato refleja la misma realidad, pero con diferentes nombres, diferentes tipos de datos, fuentes, longitud, contenido, etc. Conocer y mantener actualizada esta equivalencia entre las distintas fuentes es muy importante para la integración de la información dentro de la organización, pero su mantenimiento “manual” es altamente costoso y es susceptible de una alta tasa de error. Las herramientas informáticas que dan soporte a las actividades de gobierno de datos cuentan con la funcionalidad de identificar automáticamente, o con alguna participación del usuario, de todas las fuentes donde se encuentra el mismo dato. Estas herramientas tienen

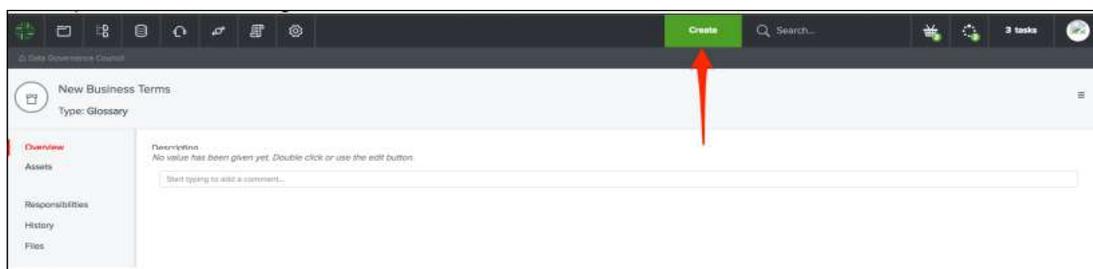
Glosario de Negocio

El glosario de negocio es una funcionalidad que le permite a los usuarios registrar las definiciones acordadas a nivel corporativo acerca de todos los términos y objetos utilizados en relación con los datos, en un repositorio centralizado y accesible a todos los participantes del ciclo de vida de estos. Por ejemplo, permite cargar la definición de las entidades, las reglas de negocio que se aplican, quienes son sus responsables, cuáles son las fuentes físicas, es decir, permite registrar la relación entre las definiciones conceptuales y los metadatos técnicos (bases de datos, campos, tablas, modelos analíticos o reportes). Por ejemplo, con el ejemplo anterior, tendríamos que la posibilidad de crear dentro del glosario la entidad “ESTADO DEL CLIENTE”, en su definición, se registraría: “la situación del cliente en cuanto a sus operaciones con la organización, un CLIENTE ACTIVO, es aquel que realiza periódicamente transacciones, por ejemplo, tiene productos o servicios activos contratados actualmente. Un CLIENTE INACTIVO es aquel que realizó operaciones con la organización en el pasado, pero por diversos motivos, fue dado de BAJA”; y en los datos relativos a los metadatos técnicos se indicarían los nombres de las bases de datos, tablas y campos donde se almacenan, tal como puede visualizarse en la imagen a continuación:



	Domain	Category	Sub Category	Term	Business Description	Tags	Status
<input type="checkbox"/>	Education	Financial		Instructional Costs	Evaluate the cost of part-time an...		PUBLIS
<input type="checkbox"/>	Education	Financial		Administrative Costs Per Student	Understand how much your institu	Education	PUBLIS
<input type="checkbox"/>	Education	Financial		Program & Department Budgets	Analyze budgets by term, seme...		PUBLIS
<input type="checkbox"/>	Education	Financial		Revenue, Grant & Funding	Alumni donations, endowments, ...	Government,Banking,Security,So...	DRAFT
<input type="checkbox"/>	Education	Financial		Student Financial Aid Percentages...	Track the number of students re...		DRAFT
<input type="checkbox"/>	Education	Financial		Tuition Costs	Monitor costs accrued by stude...	Tuition	PUBLIS
<input type="checkbox"/>	Education	Student Success		Graduation Rates	The percentage of students who...		DRAFT
<input type="checkbox"/>	Education	Student Success		Course Success Rates	Monitor completion rates on a c...		DRAFT
<input type="checkbox"/>	Education	Student Success		Persistence Rates	Understand and monitor the fact...		DRAFT

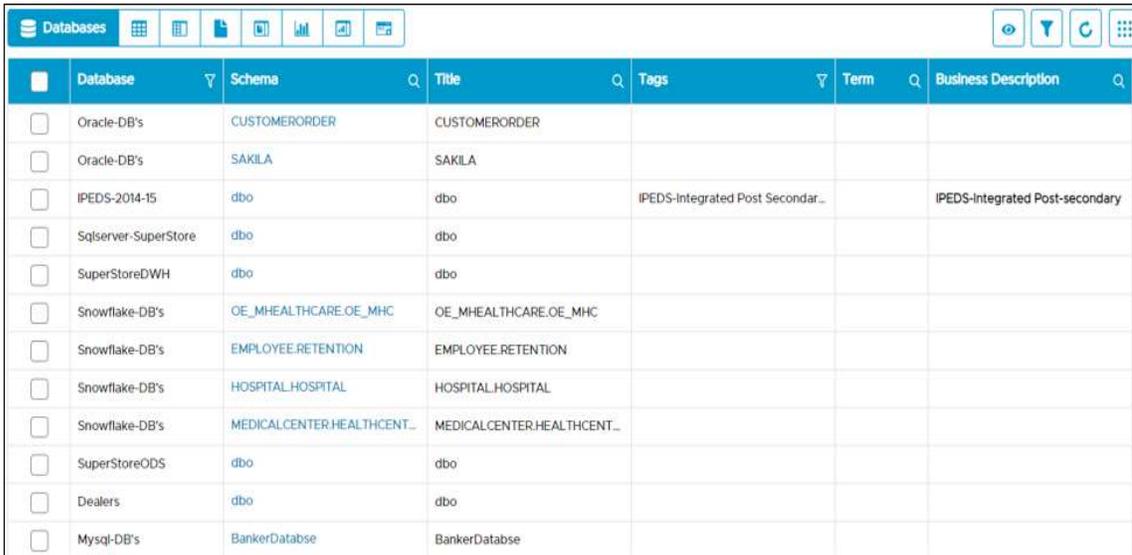
Fuente: Proveedor OvalEdge®



Fuente: Proveedor Collibra®

Catálogo de Datos

El catálogo de datos es una funcionalidad que permite visualizar de manera única y centralizada la colección ordenada o listado, de los objetos que componen la arquitectura de datos de la organización: bases de datos, tablas, campos, procesos automatizados de extracción, transformación y carga de datos, repositorios, informes, tableros, métricas, con su descripción individual. En otras palabras, se trata del inventario de los objetos de datos, lo cual constituye la materia prima para la determinación de la custodia, administración y propiedad de los datos.



The screenshot shows a data catalog interface with a table of databases. The table has columns for Database, Schema, Title, Tags, Term, and Business Description. Each row represents a different database or schema, with checkboxes in the first column.

<input type="checkbox"/>	Database	Schema	Title	Tags	Term	Business Description
<input type="checkbox"/>	Oracle-DB's	CUSTOMERORDER	CUSTOMERORDER			
<input type="checkbox"/>	Oracle-DB's	SAKILA	SAKILA			
<input type="checkbox"/>	IPEDS-2014-15	dbo	dbo	IPEDS-Integrated Post Secondar...		IPEDS-Integrated Post-secondary
<input type="checkbox"/>	Sqlserver-SuperStore	dbo	dbo			
<input type="checkbox"/>	SuperStoreDWH	dbo	dbo			
<input type="checkbox"/>	Snowflake-DB's	OE_MHEALTHCARE.OE_MHC	OE_MHEALTHCARE.OE_MHC			
<input type="checkbox"/>	Snowflake-DB's	EMPLOYEE.RETENTION	EMPLOYEE.RETENTION			
<input type="checkbox"/>	Snowflake-DB's	HOSPITAL.HOSPITAL	HOSPITAL.HOSPITAL			
<input type="checkbox"/>	Snowflake-DB's	MEDICAL.CENTER.HEALTHCENT...	MEDICAL.CENTER.HEALTHCENT...			
<input type="checkbox"/>	SuperStoreODS	dbo	dbo			
<input type="checkbox"/>	Dealers	dbo	dbo			
<input type="checkbox"/>	Mysql-DB's	BankerDatabase	BankerDatabase			

Fuente: Proveedor OvalEdge®

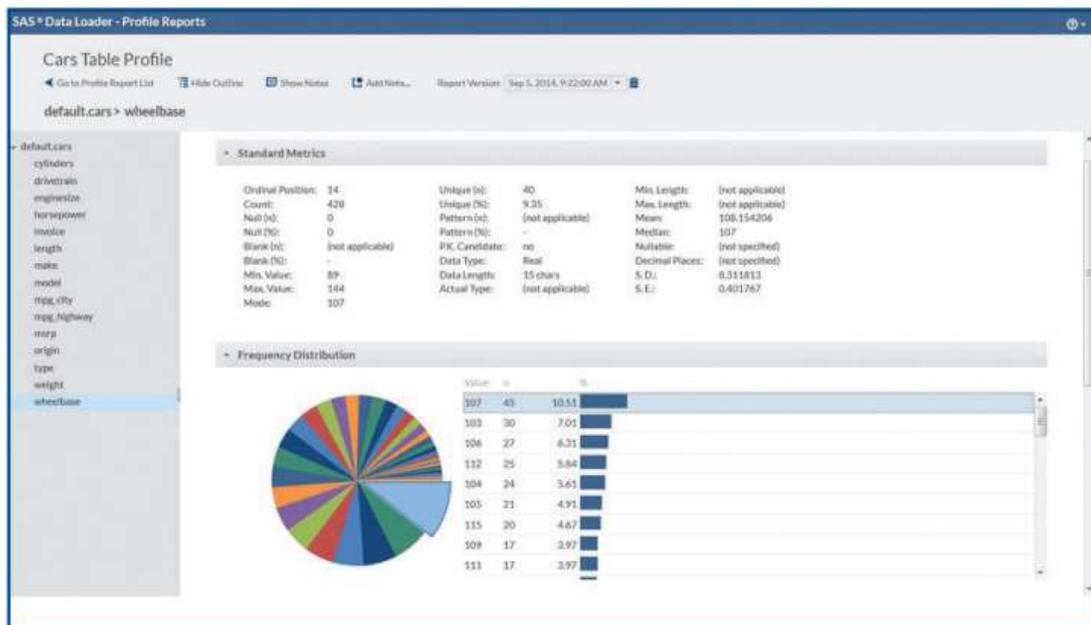
Custodia, propiedad y administración de los objetos de datos

Uno de los puntos centrales de la gobernanza de datos es determinar los roles que cada participante del ciclo de vida de los datos asumirá en relación con éstos. Las herramientas informáticas proveen la posibilidad de armar y diseñar el “listado” de los usuarios, sus roles, responsabilidades y dominios de datos de su incumbencia, de manera similar a una matriz RACI. De esta manera, queda en conocimiento de toda la organización, especialmente de los administradores de seguridad de la información, quienes ejercen cada función. Es decir, para cada dominio de datos quien es el aprobador de cambios, el responsable, los

involucrados a consultar y/o informar. Al estar automatizado, su administración y mantenimiento es muy sencilla.

Calidad

Las herramientas de calidad de datos permiten realizar el perfilado de los datos, es decir, analizar la calidad de los datos en base a los criterios previamente definidos, y luego del análisis y estandarización de estos, proceder a su limpieza (eliminación de datos inconsistentes, de duplicados, etc.), en sus diversas fuentes de origen. Permiten la aplicación de reglas de validación y la realización de métricas, reportes y alertas automáticas sobre incidencias de calidad.



Fuente: Proveedor SAS®

Gestión de metadatos

Según el grupo Gartner®, los metadatos son la "información que describe varias facetas de un activo de información para mejorar su usabilidad a lo largo de su ciclo de vida". Esta funcionalidad permite la administración de los datos técnicos acerca de los diferentes objetos de información, y permite también la comparación de esquemas, como, por ejemplo, visualizar características técnicas entre repositorios similares, pero de distintos ambientes, como pueden ser desarrollo y producción; a la vez que vincula esos datos técnicos con los términos de negocio, es decir, con los objetos de datos del glosario de negocio.

Soluciones disponibles en el mercado

De acuerdo con Gartner® (2021), el mercado de las soluciones de calidad de datos está integrado con los mercados de soluciones de integración de datos, gestión de metadatos y gestión de datos maestros; que en conjunto son utilizados por las empresas para brindar confiabilidad a sus activos de datos. “El mercado de herramientas de integración de datos incluye proveedores que ofrecen productos de software para permitir la construcción e implementación de infraestructura de acceso y entrega de datos para una variedad de escenarios de integración de datos”²². Por su parte, las soluciones de calidad de datos “incluyen una variedad de funciones críticas, como como perfilado, análisis, estandarización, limpieza, comparación, monitoreo, creación de reglas y análisis, así como flujo de trabajo incorporado, bases de conocimiento y colaboración”²³. En cuanto al mercado de soluciones de gestión de metadatos, Gartner® lo define como el conjunto de aplicaciones que incluyen las funcionalidades de documentación y administración de metadatos, glosarios de negocios, linaje de datos, análisis de impacto, administración de reglas de negocios y de conectores.

Para la elección de las herramientas para el análisis de las propuestas comerciales disponibles en el mercado actual y presentadas a continuación, se tomó en cuenta la información brindada por los cuadrantes mágicos de la consultora mencionada, de los años 2019 y 2020 para las soluciones de calidad de datos, e integración de datos, donde se observa un claro liderazgo de los siguientes proveedores: IBM, Informática, SAP, SAS y Talend, para ambos mercados, y adicionalmente se consultó el mismo informe, para el mercado de soluciones de gestión de metadatos para el mismo período, donde se observa que, tanto Informática como IBM se mantienen en el cuadrante de los líderes, junto con otros proveedores. Dichos cuadrantes se pueden consultar en el anexo I del presente trabajo.

²² Traducción propia de: “*The market for data integration tools includes vendors that offer software products to enable the construction and implementation of data access and data delivery infrastructure for a variety of data integration scenarios*”

²³ Traducción propia de: “*include a range of critical functions, such as profiling, parsing, standardization, cleansing, matching, monitoring, rule creation and analytics, as well as built-in workflow, knowledge bases and collaboration*”

En base a la información recabada de los proveedores mencionados, y estudios de mercado, se presenta a continuación una breve reseña de algunas de las principales las herramientas informáticas disponibles, con foco en las prestaciones generales que brindan. Para la elección de las herramientas a presentar, se tomó como criterio el rango de funcionalidades que abarcan, puntualmente, gestión de la calidad de datos, gestión de metadatos, integración de datos y gobierno de datos propiamente dicho. Las soluciones seleccionadas son las siguientes:

Suite Proveedor IBM ®

Solución InfoSphere® Information Server, compuesta, entre otros, por:

IBM InfoSphere® Information Governance Catalog

IBM InfoSphere® Metadata Asset Manager

IBM InfoSphere® Information Server for Data Integration

IBM InfoSphere® Information Server for Data Quality

Suite Proveedor Informatica ©

Solución Informática Governance and Data Quality®, compuesta por:

Informatica Enterprise Data Catalog

Informatica Axon™ Data Governance

Suite Proveedor SAP®

Gestión de Datos, compuesta, entre otros, por:

Solución SAP® Information Steward

SAP® PowerDesigner

Suite Proveedor SAS®

Solución SAS® Data Governance, compuesta por:

SAS® Repositorio de Relaciones

SAS® Lineage

SAS® Business Data Network

SAS® Metadata Bridges

Suite Proveedor Talend ®

Solución Talend ® Data Fabric, compuesta, entre otras, por:

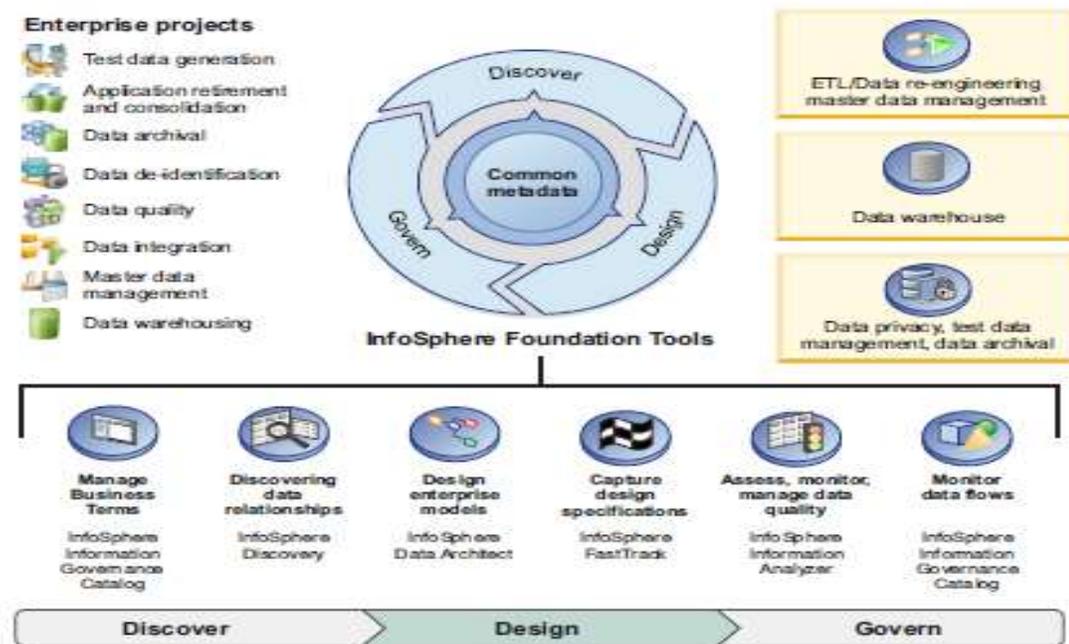
Talend ® Data Catalog

Talend ® Data Preparation



Suite Proveedor IBM ® – Solución InfoSphere® Information Server

La suite de InfoSphere® Information Server consta de diversos componentes, que proporcionan cada uno distintas funciones para la integración de datos, calidad y gobierno. Para este análisis se consideraron los siguientes productos pertenecientes a la Suite InfoSphere® Information Server: InfoSphere® Information Governance Catalog, InfoSphere® Metadata Asset Manager, InfoSphere® Information Server for Data Integration, InfoSphere® Information Server for Data Quality, los cuales se combinan de diferentes maneras y constituyen diferentes ofertas, que incluso pueden ejecutarse con Hadoop (BigInsights BigQuality y BigInsights BigIntegrate) o integrarse con herramientas de inteligencia artificial como Watson.

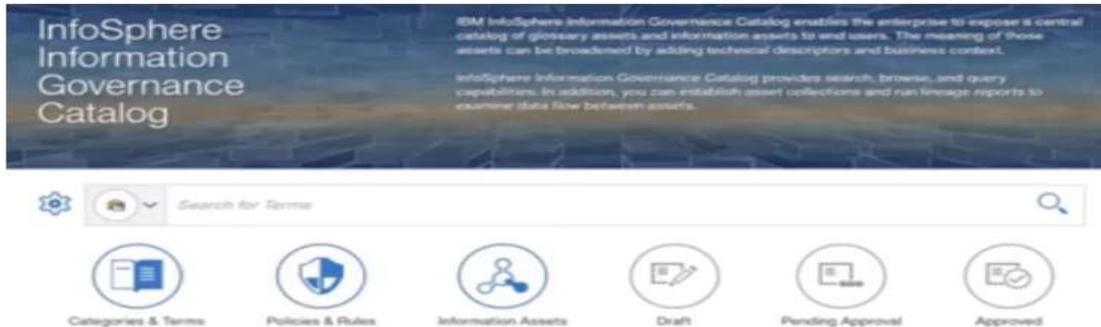


Fuente: Proveedor IBM ®

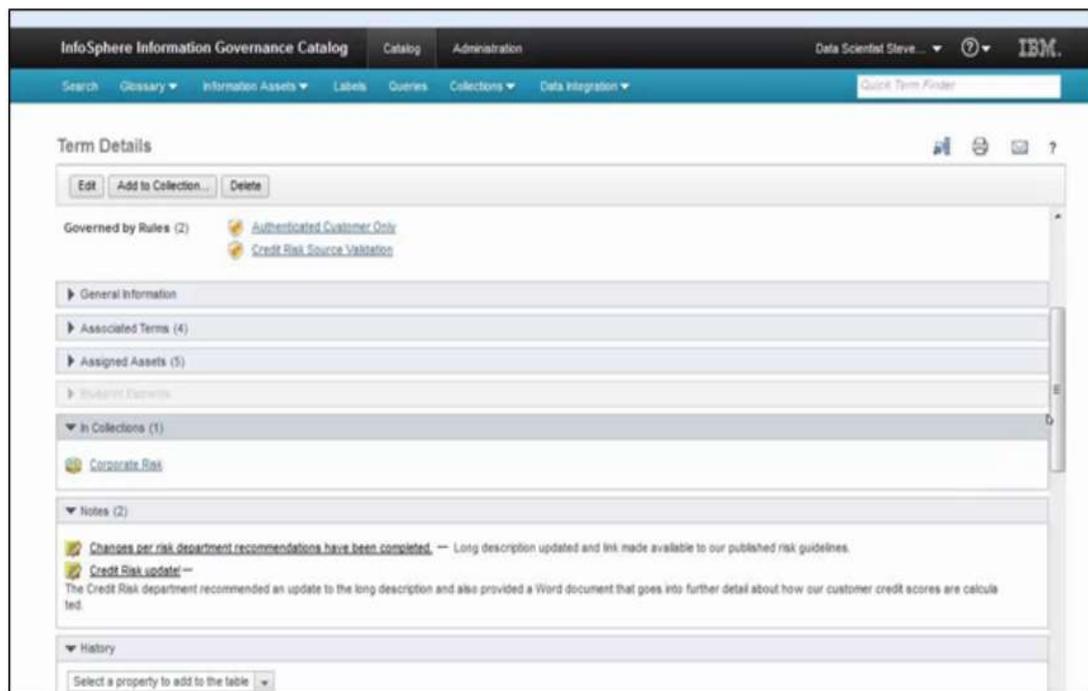
IBM InfoSphere® Information Governance Catalog

Es el producto de IBM para crear y mantener el catálogo de datos; permite a los usuarios explorar, organizar, categorizar y compartir entre sí la información relativa a los activos de datos, conjuntos de datos, modelos de analítica y sus relaciones. Dentro de esta herramienta

se pueden crear categorías y términos que le agregan el significado comercial a los elementos. Permite obtener el linaje de datos de las tablas y columnas de la base de datos; desde su origen hasta su destino. Se combina con IBM Watson® Knowledge Catalog.

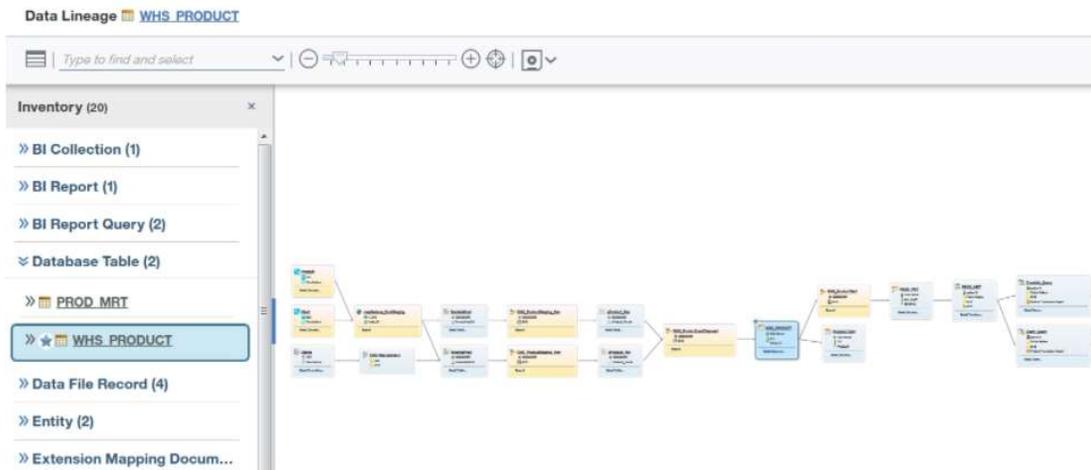


Fuente: Proveedor IBM ®



Fuente: Proveedor IBM ®

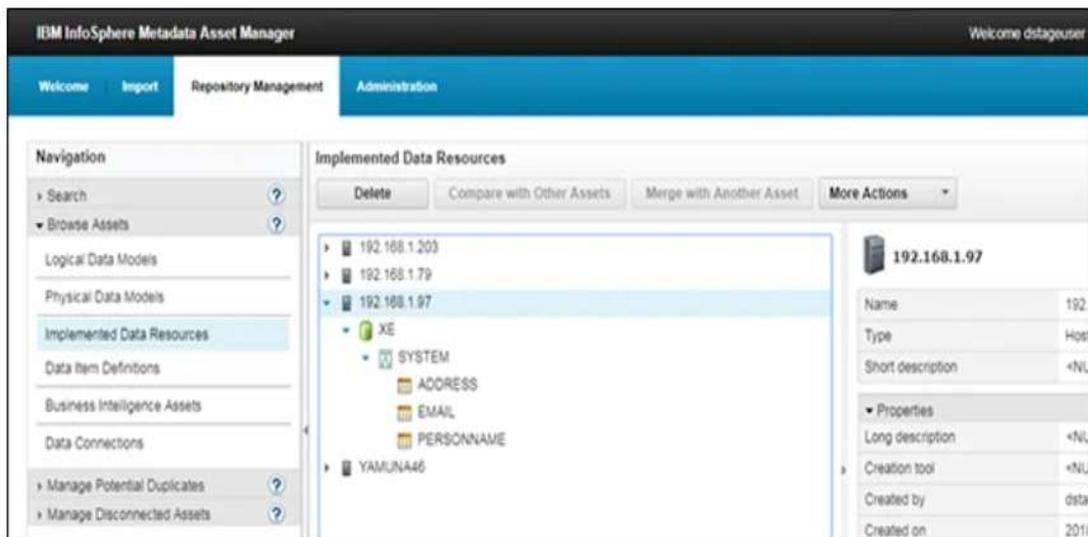
El linaje de datos se puede analizar en forma gráfica:



Fuente: Proveedor IBM ®

IBM InfoSphere® Metadata Asset Manager

InfoSphere® Metadata Asset Manager permite importar, exportar y administrar activos de metadatos comunes, necesarios para el funcionamiento de todos los componentes de la suite InfoSphere® Information Server. Puede importar activos a un área de preparación antes de compartirlos con el repositorio de metadatos. En este repositorio, puede exportar activos, explorar y buscar activos de metadatos comunes, establecer relaciones de implementación entre ellos y fusionar duplicados. Se trata de un cliente web que se ejecuta en distintos navegadores. Importa metadatos de herramientas, bases de datos y archivos de terceras partes mediante conectores.



Fuente: Proveedor IBM ®

IBM InfoSphere® Information Server for Data Integration

IBM InfoSphere® Information Server for Data Integration extrae y transforma datos de diversas fuentes heterogéneas y realiza su carga en destino, a través de diferentes reglas de transformación.

IBM InfoSphere® Information Server for Data Quality

InfoSphere® Information Server para calidad de datos es el producto cuya función es limpiar los datos y monitorear su calidad en forma continua, mediante el chequeo del cumplimiento de las reglas y políticas predefinidas. Permite la creación de métricas y opciones de remediación.



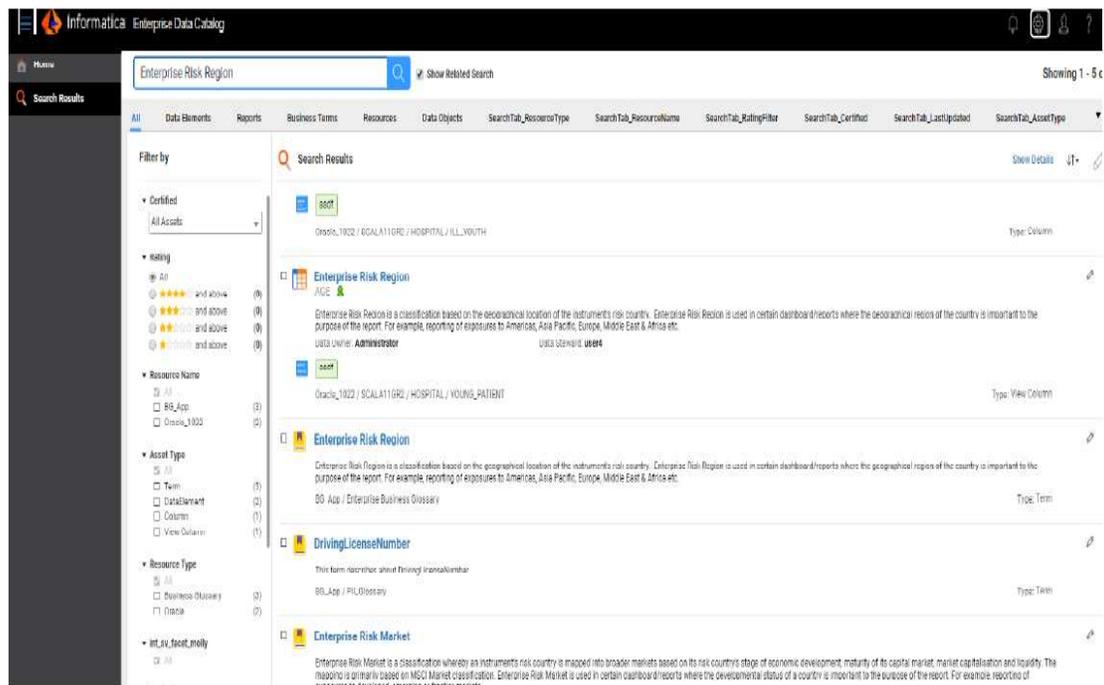
Fuente: Proveedor IBM ®



La solución propuesta por Informatica © para el Gobierno de Datos corporativo consta de tres herramientas complementarias: Informatica Enterprise Data Catalog, Informatica Axon™ Data Governance, e Informatica Data Quality.

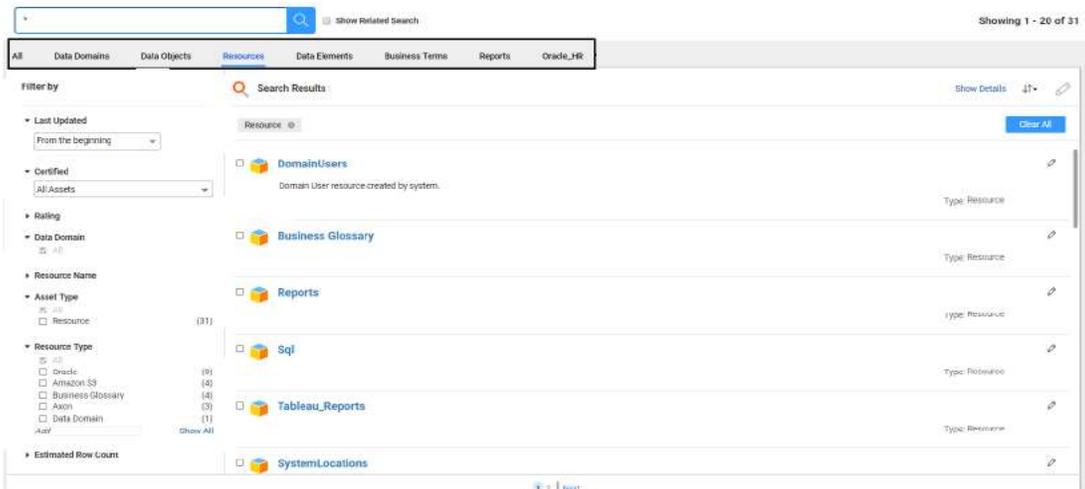
Informatica Enterprise Data Catalog

Es la herramienta a través de la cual se pueden extraer los metadatos técnicos de los objetos, organizarlos de acuerdo con las necesidades del negocio y generar automáticamente el linaje de datos y las relaciones entre éstos. Extrae metadatos de fuentes externas, como bases de datos, almacenes de datos, glosarios empresariales, recursos de integración de datos o informes de inteligencia empresarial.



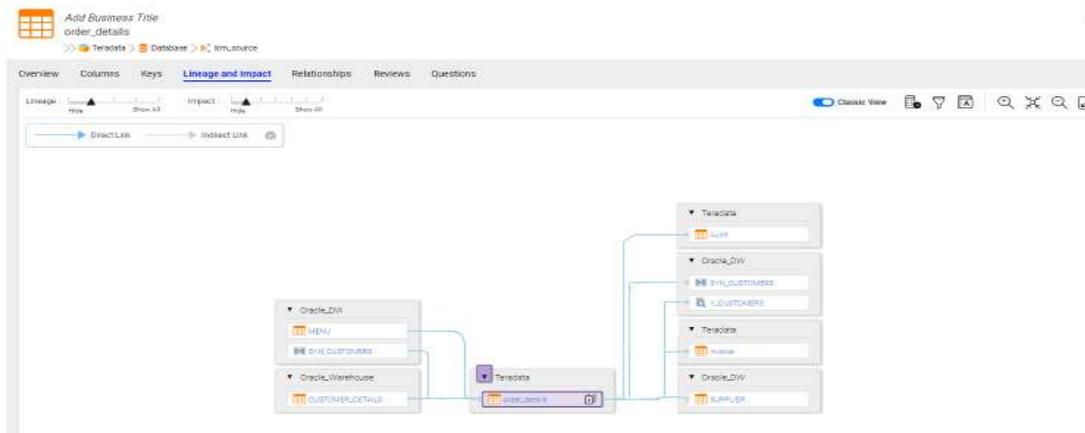
Fuente: Proveedor Informatica ©

Los activos representan objetos de datos como tablas, columnas, informes, vistas y esquemas.



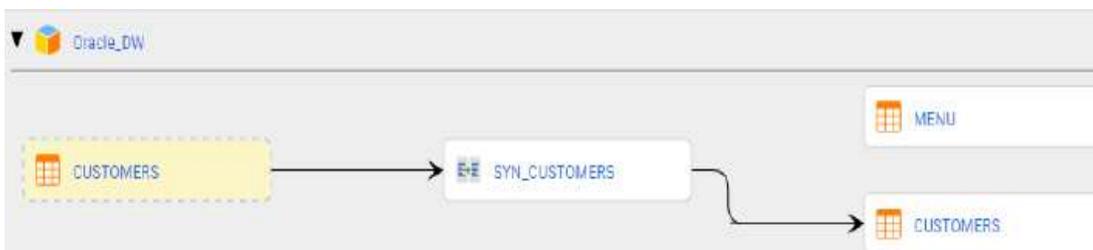
Fuente: Proveedor Informatica ©

Permite visualizar el linaje en forma de diagramas (de relación), donde se visualizan los detalles de cada objeto, tal como se muestra a continuación:



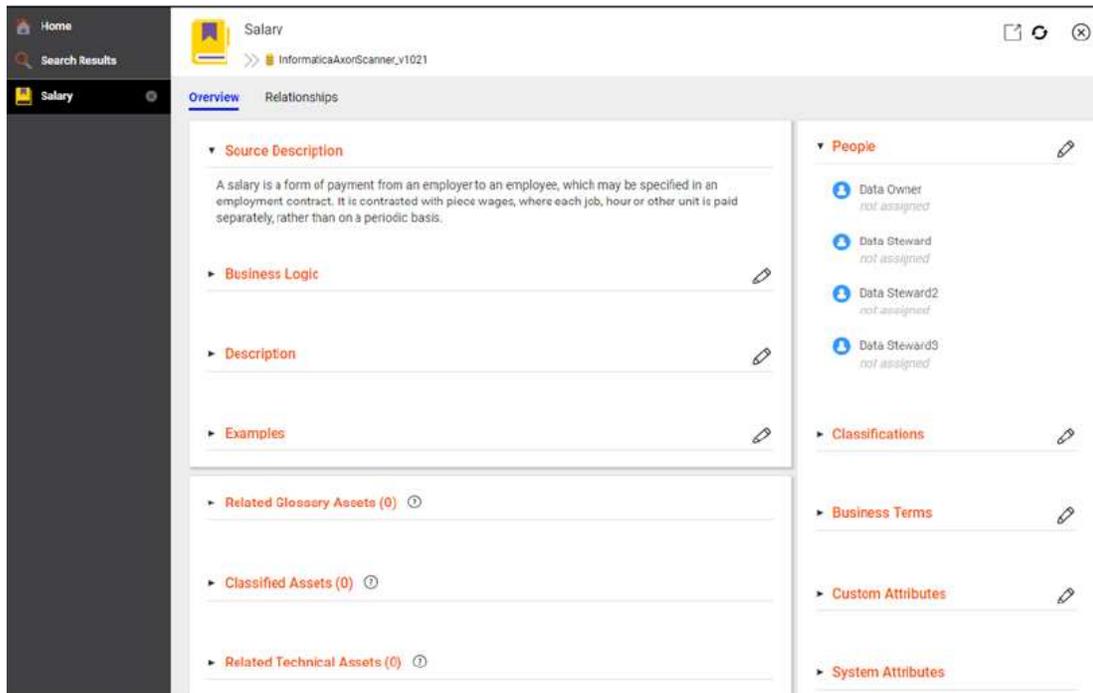
Fuente: Proveedor Informatica ©

O en modo: “vista compacta”, donde sólo se visualizan los objetos a alto nivel.



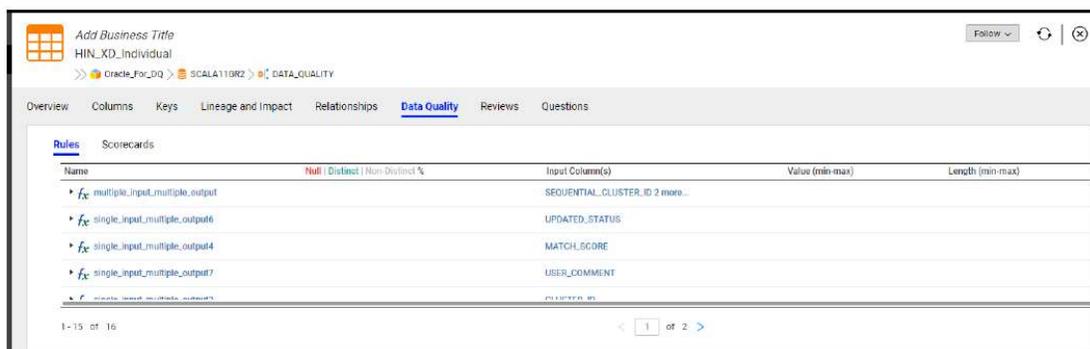
Fuente: Proveedor Informatica ©

Esta herramienta también cuenta con la funcionalidad de crear un glosario con los términos de negocio utilizados por la compañía, en el cual se puede registrar su definición y su relación con el resto de los objetos.



Fuente: Proveedor Informatica ©

Permite también la creación de reglas de calidad a aplicar sobre dichos objetos:



Fuente: Proveedor Informatica ©

Y el monitoreo de éstas, mediante distintos controles y métricas, con diferentes visualizaciones:

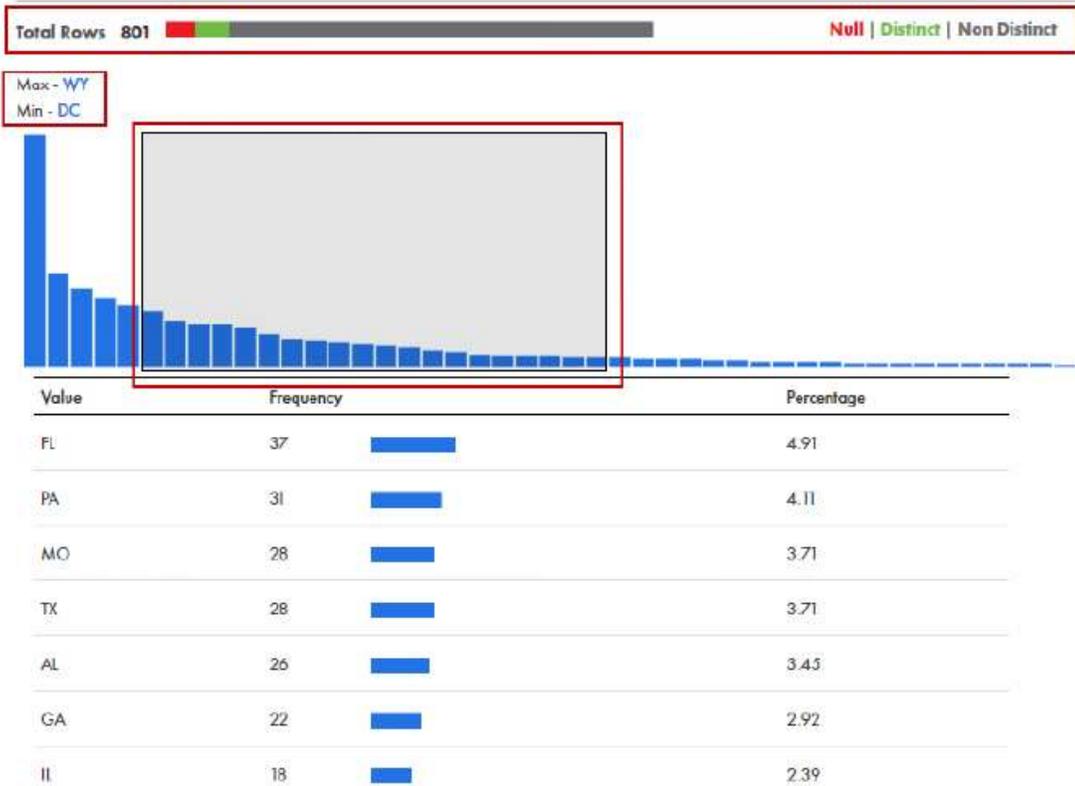
Name	Business Title	Data Domains (0)	Null Distinct Non-Distinct %	Source Data Type Inferred Data Types
1 CITY		Address(81.82%) +9 more	0 22.72 77.28	char (100) String(18) 100.00%
2 COMPANY		Address(0.09%) +4 more	0 13.63 86.37	char (100) Fixed Length String(0) 0.90% String(16) 100.00%
3 COUNTRY		Address(100%) +4 more	0 4.54 95.46	char (100) Fixed Length String(2) 100.00% String(2) 100.00%
4 CREATEDATE		Date_MM_DD_YYYY(100%)	0 27.27 72.73	date (10) Date Time(mm/dd/year HH24:miss) 100.00% +2 m...
5 CURRENCY		Currency	100 0 0	char (100)
6 CUSTOMERID		Age(100%) +9 more	0 45.45 54.55	decimal (10) Decimal(2) 100.00% +2 more
7 FIRSTNAME		Address(0.09%) +9 more	0 43.45 56.55	char (100) String(0) 100.00%
8 LASTNAME		Address(0.09%) +10 more	0 45.45 54.55	char (100) String(8) 100.00%
9 MISCDATE		Date_MM_DD_YYYY(100%)	0 13.63 86.37	timestamp (10) Date Time(mm/dd/year HH24:miss) 100.00% +2 m...
10 ORDERAMOUNT		AlphaNumeric_SpecialCharacter... +3 more	0 43.45 56.55	decimal (10) Decimal(94) 100.00% +2 more

Fuente: Proveedor Informatica ©



Fuente: Proveedor Informatica ©

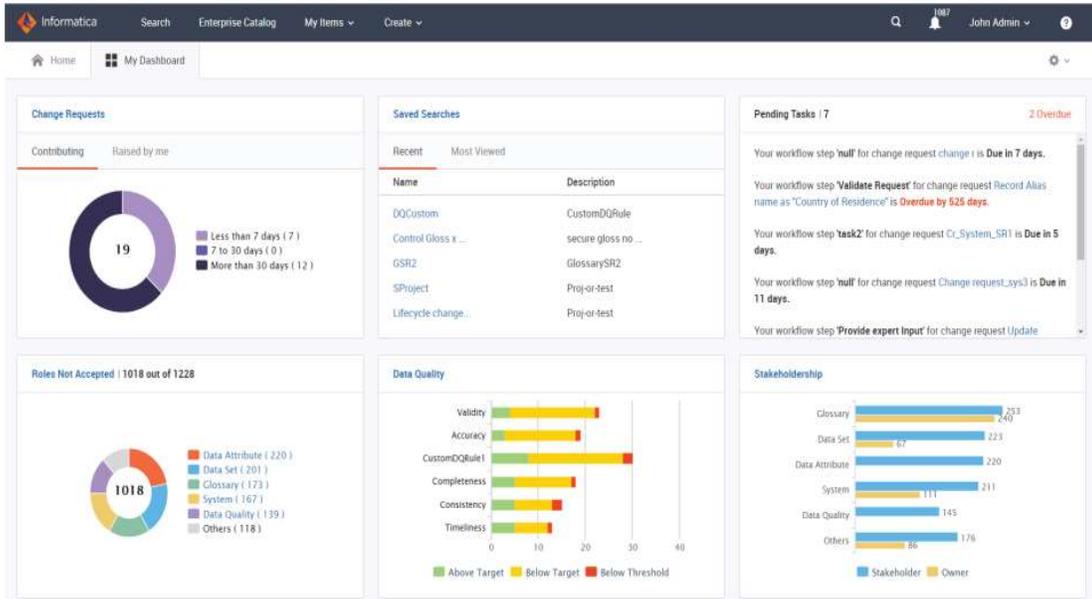
Value Frequency



Fuente: Proveedor Informatica ©

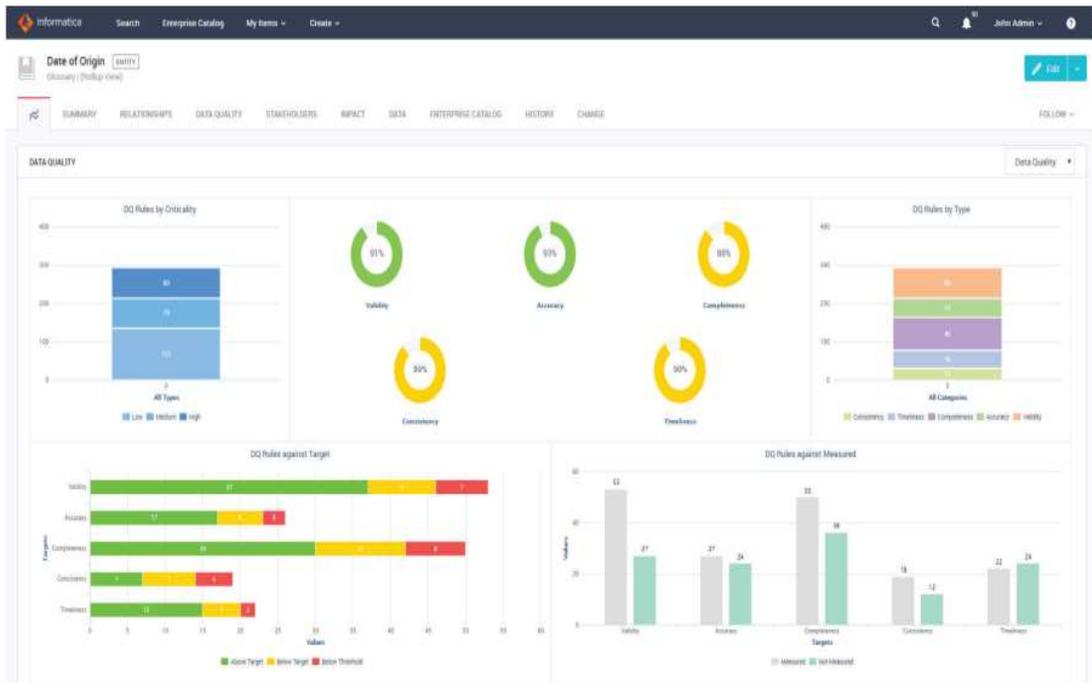
Informatica Axon™ Data Governance

Informatica Axon™ Data Governance es una herramienta para documentar los activos de datos que requieren gobierno, junto con la información relativa a sus propietarios, administradores y custodios. Administra el inventario de activos de datos de la organización. A partir de éste, se pueden asignar los roles que cada involucrado ejercerá con relación a éstos. Tiene incorporadas funcionalidades de calidad de datos, que permiten monitorearlos, realizar su perfilado y la limpieza de éstos.



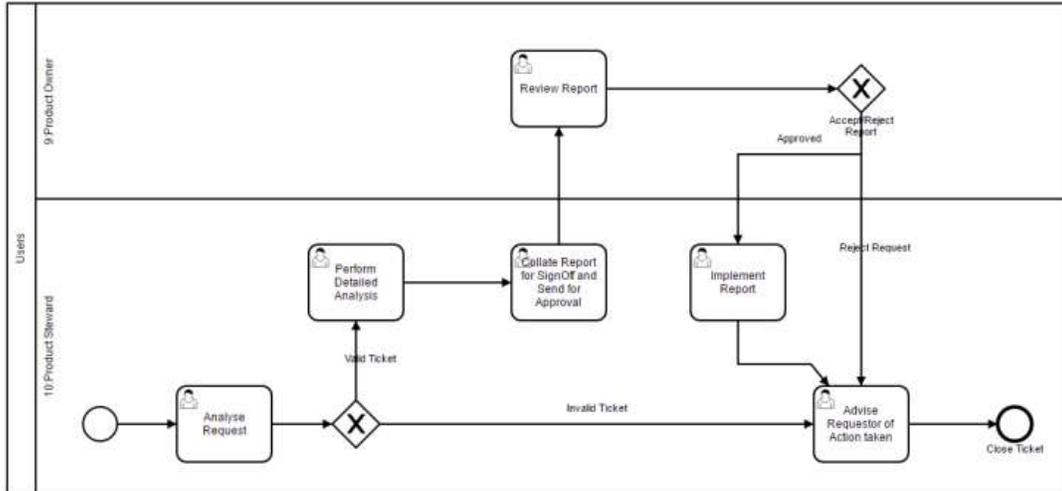
Fuente: Proveedor Informatica ©

Cuenta con tableros de control predeterminados sobre las siguientes temáticas: Calidad de datos, Custodia, Procesos y Privacidad.



Fuente: Proveedor Informatica ©

Permite la creación, diseño y visualización de procesos, con diferentes vistas, como, por ejemplo, la presentada a continuación:



Fuente: Proveedor Informatica ©



Dentro de las soluciones que ofrece el proveedor mencionado, para el mercado de soluciones de gestión de metadatos y calidad de datos analizados por Gartner®, se encuentran: SAP® Information Steward y SAP® PowerDesigner, que además de funciones de gestión de metadatos tiene funciones que permiten gestionar la arquitectura de datos. Para el mercado de calidad de datos, también se consideran SAP® Data Services y SAP® Data Quality Management microservices for location data.

Solución SAP® Information Steward

Esta herramienta permite la realización de las siguientes funciones:

Perfilado de datos: analiza automáticamente la información y muestra opciones de estandarización de valores, formatos y agrupaciones.

Table	Median	Null %	Blank %	Zero %	Value	Pattern	Word
ADDRESS1	17	0.0	1.9		245	190	400
ADDRESS2	0	0.0	76.1		42	37	91
CITY	0	0.0	2.8		187	43	204
FACILITY_NAME	10	0.0	0.0		18	11	23
NAME	13	0.0	0.0		526	123	690
PATIENT_ID	3	0.0	0.0		530	3	530
POSTALCODE	5	0.0	6.7		217	7	218
PRIMARY_DIAGNOSES	23	0.0	0.0		84	22	35

Value	Percent (#)
Bose General	8.10% (44)
Eastern MC	7.25% (39)
Mercy Hospital	6.69% (36)
Portland MC	6.32% (34)
St. Vincent MC	6.12% (33)
Milwaukee Hospital	5.02% (27)
Memorial Hospital	4.46% (24)
60007	0.19% (1)
55420	0.19% (1)

ADDRESS1	ADDRESS2	CITY	FACILITY_NAME	NAME	PATIENT_ID	POSTALCODE	PRIMARY_DIAGNOSES	REGION
3130 BROADWAY STREET	SUITE 50	KANSAS CITY	Mercy Hospital	CAMERON HARRIS	25	64111	Orthopedics Diagnosis 3	MO
303 NORTH GLENDALES BLVD	SUITE 800	BURBANK	Mercy Hospital	ANGELA MORGAN	39	91502	Gynecology Diagnosis 3	CA
17944 NORTH EAST 65TH STREET	<Blank>	SEATTLE	Mercy Hospital	MARCO LOPEZ	74	98052	Urology Diagnosis 1	WA
106 FRISCO STREET	SUITE 104	MARKED TREE	Mercy Hospital	GABRIELLE MURPHY	80	<Blank>	Vascular Services Diagnosis 5	AR
4501 INTELCO LOOP SE	<Blank>	OLYMPIA	Mercy Hospital	MEGAN LONG	94	98503	Neurosurgery Diagnosis 5	WA
1899 PALOMBO AVE	<Blank>	CASPER	Mercy Hospital	JOY RUIZ	95	82601	Rehabilitation Diagnosis 2	WY
13000 W CLUSTER AVE	<Blank>	BLAIRSTOWN	Mercy Hospital	TERESA ALVAREZ	124	83007	Spine Diagnosis 5	UT

Fuente: Proveedor SAP®

Creación de reglas de negocios y limpieza de datos: en base al análisis de datos efectuado, realiza una recomendación automática de reglas posibles a aplicar, donde el usuario puede aceptar, denegar o crear otras reglas, que serán aplicadas en los datos para mejorar su calidad. Permite realizar la limpieza de datos en forma automática.

Name	Type	Reason	Actions
COUNTRY			<input type="checkbox"/> Remove
COUNTRY allowed	Proposed Rule	Possible bad value(s): US	<input checked="" type="checkbox"/> Accept
POSTALCODE			<input type="checkbox"/> Remove
PostalCode Rule	Proposed Binding	Rule associated with content	<input checked="" type="checkbox"/> Accept
SSN			<input type="checkbox"/> Remove
SSN allowed pattern	Proposed Rule	Possible bad pattern(s): 999-*	<input checked="" type="checkbox"/> Accept

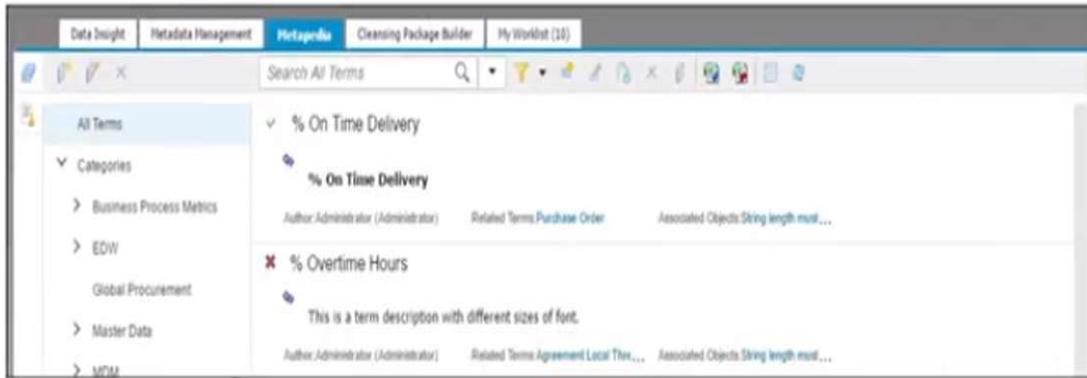
Value	Percent (#)	Exc
US	0.9% (3)	
CAN	2.9% (10)	
USA	96.2% (332)	

Fuente: Proveedor SAP ®

Before	After
ADDRESS1 ADDRESS2	Address
3130 BRDWAY STREET SUITE 50	3130 Broadway St Ste 50
6625 THE CORNERS PARKA... APT 260	6625 The Corners Plwy Ste 260
PO BOX 30096	PO Box 30096
2323 GRAND BLVD FLOOR 6	2323 Grand Blvd Fl 6
17944 NE 65TH STREET	17944 NE 65th St
6625 THE CORNERS PARK... # 260	6625 The Corners Plwy Ste 260
30 VOICE ROAD	30 Voice Rd
3495 PIEDMONT ROAD NO... # 520	3495 Piedmont Rd NE Ste 520

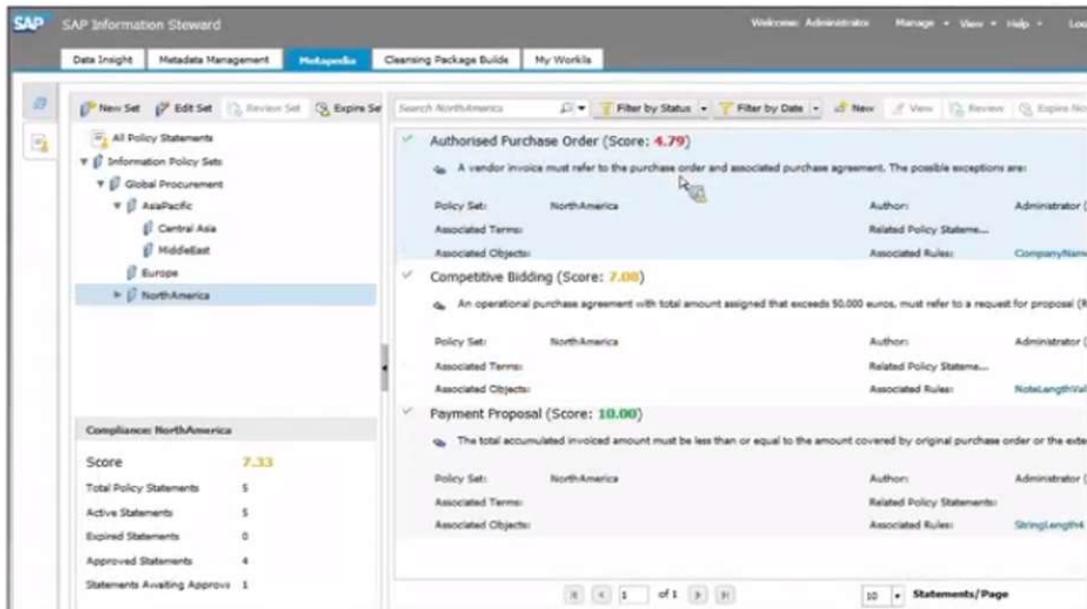
Fuente: Proveedor SAP ®

Glosario de términos de negocio: bajo el nombre “metapedia”, la herramienta permite desplegar el glosario explicativo de los activos de datos en una forma amena y accesible para todos los usuarios.



Fuente: Proveedor SAP ®

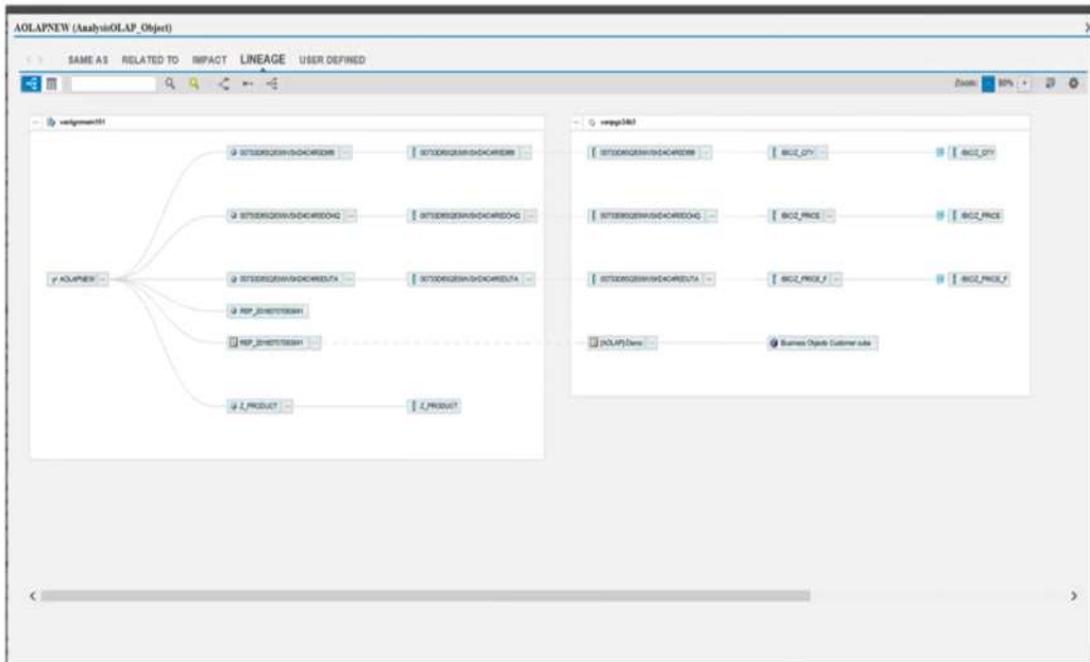
Gestión de políticas de información: permite definir políticas en forma centralizada y compartirlas con todos los usuarios.



Fuente: Proveedor SAP ®

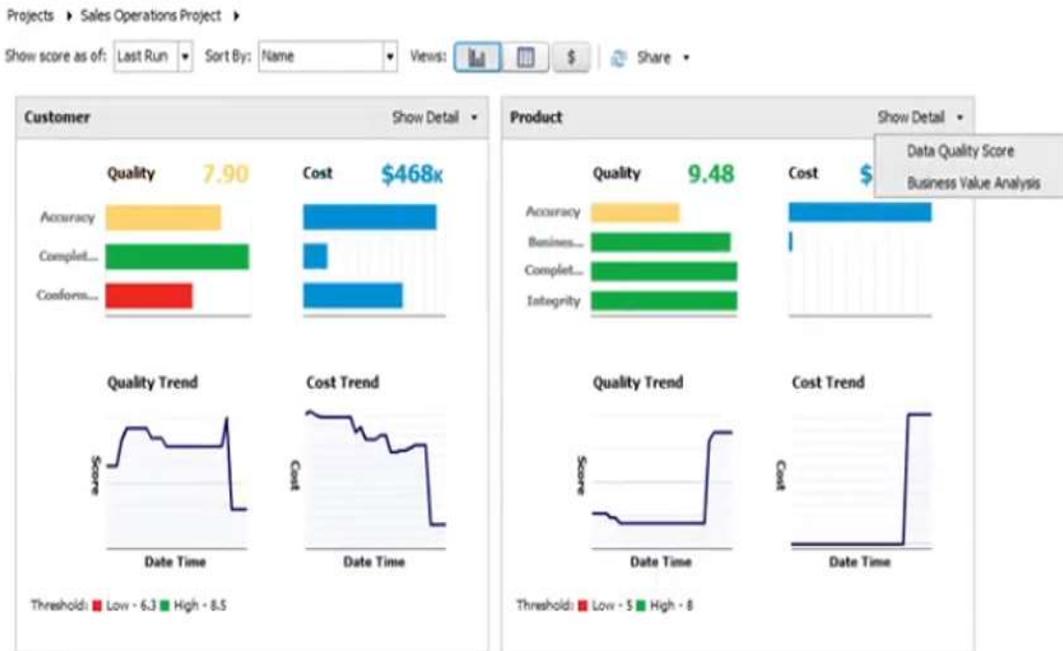
Linaje de Datos:

Esta funcionalidad permite visualizar en forma gráfica el recorrido de los datos desde su origen en las bases de datos transaccionales hacia su destino, en forma automática.



Fuente: Proveedor SAP ®

Cuenta con tableros de comando predeterminados que muestran indicadores sobre la calidad de los datos:



Fuente: Proveedor SAP ®

SAP® PowerDesigner

Sus funcionalidades principales son la creación de modelos de datos, enlaces y sincronización, gestión de metadatos, y detección automática de capas y requisitos de arquitectura de datos. Permite la creación de un repositorio de metadatos único y centralizado.



El conjunto de aplicaciones de este proveedor abarca a los tres grupos: herramientas para la integración de datos, es decir, aquellas que se utilizan para la creación de repositorios unificados a partir de fuentes heterogéneas, utilizadas para la generación de procesos de extracción, transformación y carga, conocidos como “ETLs” o “ingestores”; herramientas de calidad de datos, que contemplan las funciones de de-duplicación (eliminación de registros duplicados), identificación, estandarización, enriquecimiento, monitoreo y perfilado de datos, y por último, herramientas específicas de Gobierno de Datos, las cuales se componen de: el glosario o diccionario de datos, que permite compartir un lenguaje común para los datos técnicos y de negocios, y asignar las distintas responsabilidades en la gestión de los dominios de datos; el linaje o trazabilidad de los datos, que muestra gráficamente, la secuencia completa de los procesos que sufren los datos desde el origen hacia su destino final. Los productos comerciales de esta solución son: SAS® Repositorio de Relaciones, SAS® Lineage, SAS® Metadata Bridges y SAS® Business Data Network.

SAS® Repositorio de Relaciones

Este es el componente principal de la solución. Se trata de un repositorio donde se registran los datos técnicos y de negocio, y sus relaciones; en otras palabras, los metadatos y las relaciones entre éstos.

SAS® Lineage

Es un componente basado en Web que permite mostrar de manera gráfica dónde están almacenados los datos y sus metadatos, a lo largo de su ciclo de vida dentro de la organización. Proporciona un almacén compartido para toda la información de las relaciones entre los objetos, denominado “servicio de relaciones SAS”. También posee la capacidad de importar contenido de fuentes de terceros. Sus principales ventajas son: permitir la visualización de manera gráfica y automática de la trazabilidad de los datos, de las reglas y transformaciones en cada flujo, e identificar las fuentes correctas y los flujos de datos óptimos para cada nuevo requerimiento.

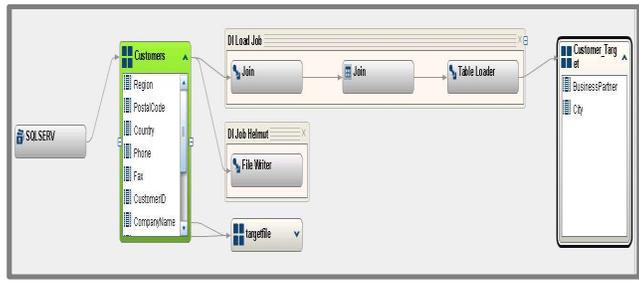
Esta información se puede visualizar de dos maneras:

Gráfico de red



Fuente: Proveedor SAS ®

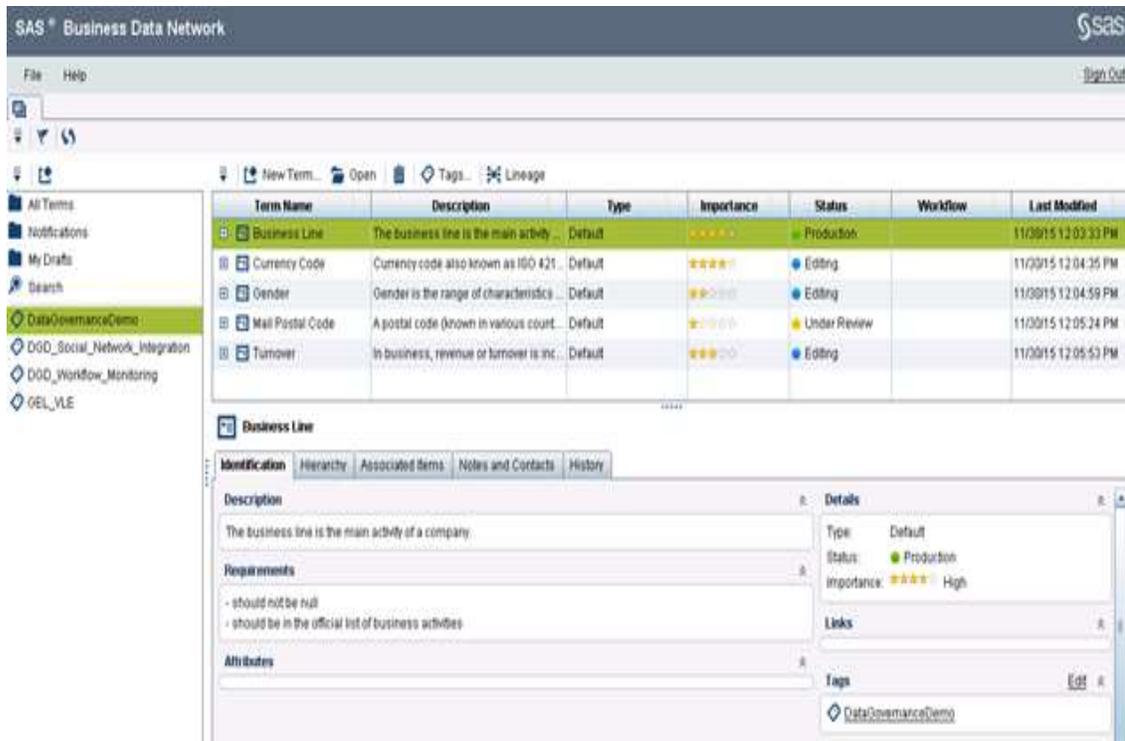
Gráfico de dependencias



Fuente: Proveedor SAS ®

SAS® Business Data Network

Es la aplicación Web donde se registran las definiciones de negocio en el Repositorio de Relaciones. Permite trabajar de manera colaborativa, y definir, además, las fuentes de datos, las personas involucradas, gestores, custodios y propietarios. Se integra con SAS Lineage y permite definir permisos de seguridad, roles y usuarios.



Fuente: Proveedor SAS ®

SAS® Metadata Bridges

Se trata de los conectores a terceras partes que permiten extraer los datos técnicos requeridos e incorporarlos al repositorio SAS. Funcionan de dos formas diferentes, puede realizarse una conexión directa con la tercera parte e importar y extraer de su repositorio los metadatos, o de forma indirecta, estos pueden ser exportados por el equipo técnico y luego ser importados por el conector de SAS para ser registrados en el repositorio de relaciones.

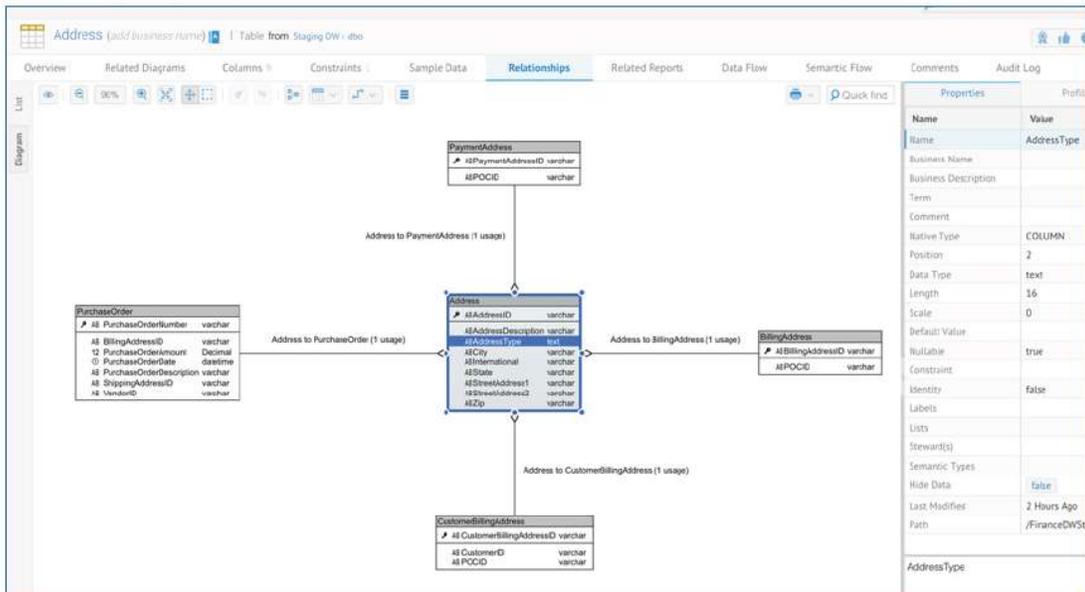
La modalidad de conexión directa como su nombre lo indica exige que el conector acceda al servidor de la tercera parte a importar y tenga permisos para realizar esta operación. Esta modalidad es más invasiva, pero permite la detección automática de las relaciones entre los objetos. La modalidad importación no exige conexión ni permisos sobre la tercera parte, pero exige mayor trabajo para la generación de relaciones en forma manual por el equipo técnico.



La solución brindada por este proveedor también abarca los tres aspectos mencionados anteriormente: calidad de datos, integración de datos y funciones específicas relacionadas con el gobierno de datos propiamente dicho: linaje o trazabilidad y glosario de negocios.

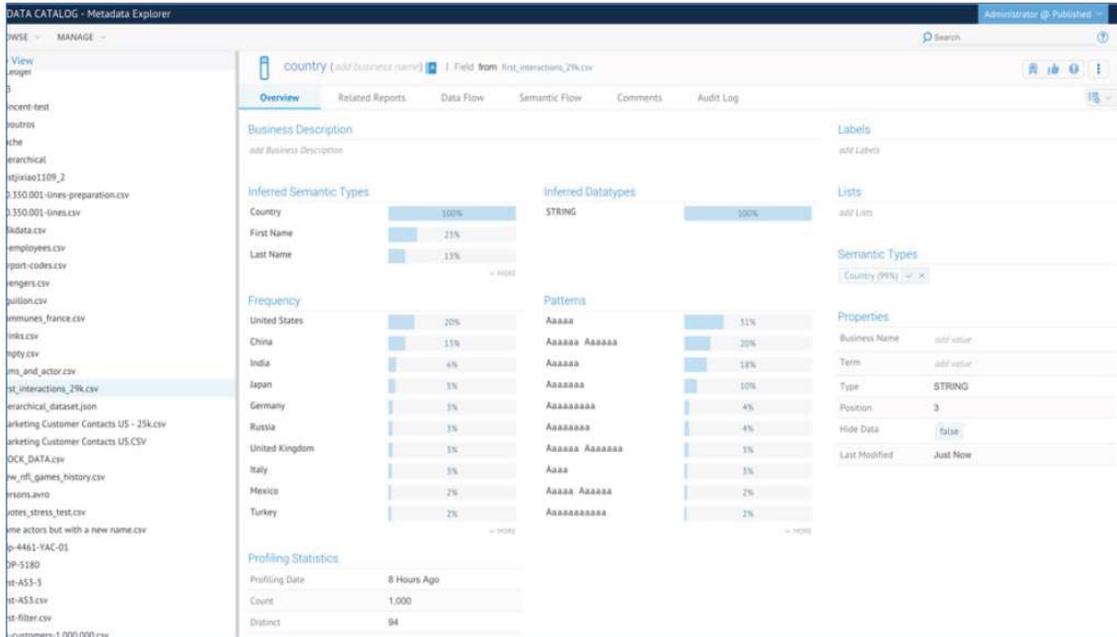
Talend® Data Catalog

Este módulo permite crear de manera automatizada un catálogo central y controlado de datos enriquecidos en el que se pueden explorar, organizar y documentar automáticamente los metadatos. Permite identificar las fuentes de origen, los almacenes de datos, los procesos de extracción, transformación y carga, los repositorios de las aplicaciones de inteligencia de negocios, y modelos conceptuales, de manera automática, a través de la utilización de conectores a aplicaciones de terceros. Traza automáticamente los vínculos entre conjuntos de datos y los conecta con el glosario de negocios.



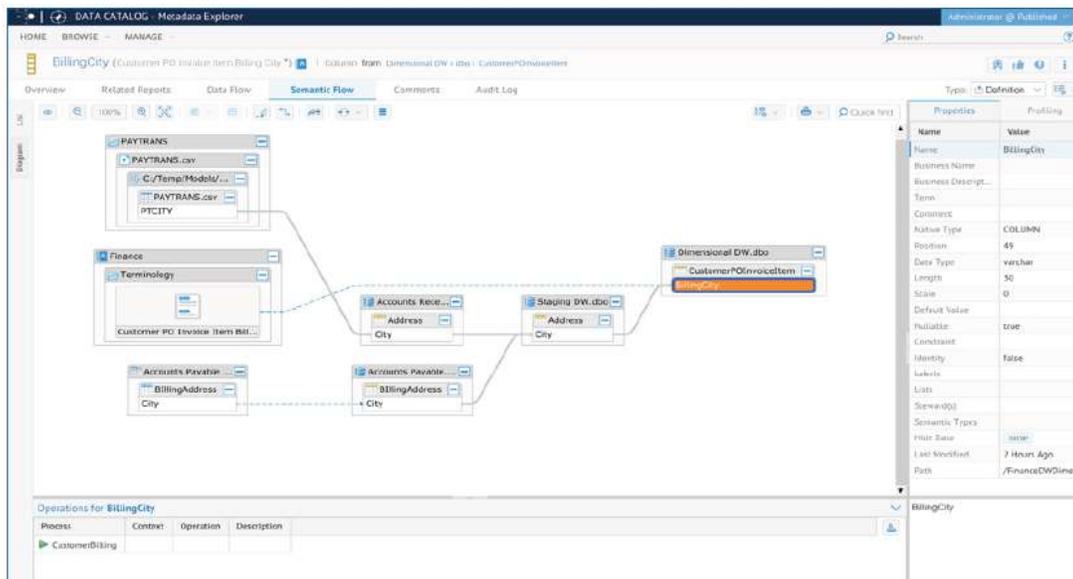
Fuente: Proveedor Talend®

A su vez, puede exponer los posibles problemas de calidad de los datos que requieren corrección. Permiten visualizar una vista previa del contenido de los datos, al ver muestras de estos, tal como se muestra a continuación:



Fuente: Proveedor Talend®

Este módulo permite también la creación del glosario de negocios:



Fuente: Proveedor Talend®

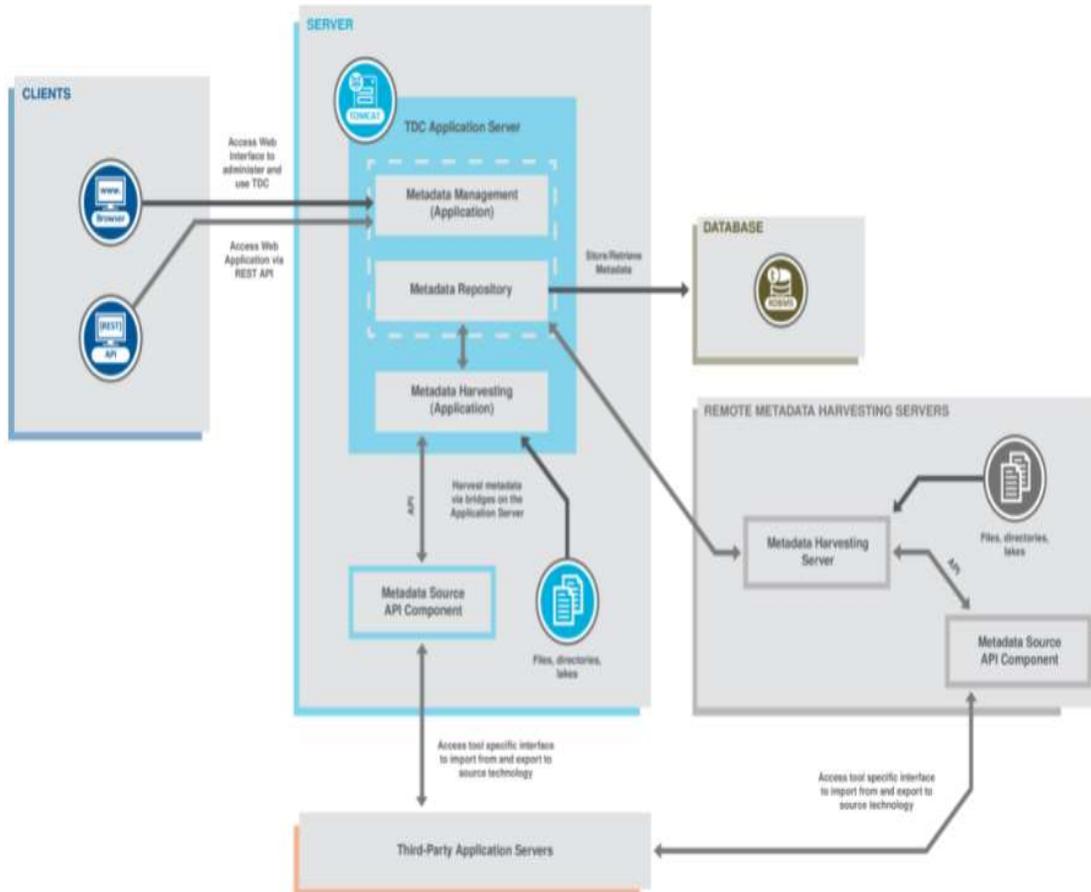
Y la realización de la matriz de asignación de roles:

The screenshot shows the Talend Management Console interface. On the left is a navigation menu with categories: OPERATIONS, MANAGEMENT, PROJECTS, ENGINES, ENVIRONMENTS, PROMOTIONS, CONFIGURATIONS, USERS, GROUPS, ROLES, and SUBSCRIPTIONS. The 'ROLES' category is highlighted. The main area displays a table with the following data:

NAME	APPLICATIONS	NUMBER OF USERS
API Designer	API Designer	4
API Tester	API Tester	4
Campaign Owner	Data Stewardship	4
Data Preparation Administrator	Data Preparation	3
Data Preparation Manager	Data Preparation	1
Data Preparator	Data Preparation	2
Data Steward	Data Stewardship	5
Environment Administrator	Management Console	3
Infrastructure Administrator	Management Console	3
Integration Developer	Studio	3

Fuente: Proveedor Talend ®

Su arquitectura es la siguiente:



Fuente: Proveedor Talend ®

Talend ® Data Preparation

Talend Data Preparation es la aplicación que permite la realización de las operaciones de preparación de datos destinadas a aumentar la calidad de éstos, ya que permite la definición y aplicación de reglas y políticas para la accesibilidad y el gobierno de los datos; la exploración y el perfilamiento de datos, su limpieza, normalización, modelado y enriquecimiento para su posterior explotación.

Talend Data Fabric (7.0.1.20190411_1414) | DQ_Demos_project (Connection: Local)

Job A_matchCustomers 0.1

incoming_customers → fuzzyMatching_1 → filterColumns_1 → filterColumns_2 → new_customers

filterColumns_1: row2 (Matches order: 1)
 filterColumns_2: row4 (Matches order: 2)
 filterColumns_3: row6 (Matches order: 3)

Metadata for Job A_matchCustomers 0.1:

- Name: A_matchCustomers
- Author: user@talend.com
- Version: 0.1
- Creation: 8/13/10 0:41 PM
- Modification: 11/14/18 6:31 PM
- Job Type: Standard
- Framework: []
- Purpose: example of fuzzy matching
- Description: This job uses the fuzzyMatching component in order to split incoming data into three sets of data: matches, non-matches, possible matches.

Fuente: Proveedor Talend®

DATA PREPARATION

DATA PREPARATION

1. Change data type on column price
 2. Change data type on column price
 3. Change to upper case on column car_name
 4. Reorder columns on column car_name_upper
 5. Remove column on column car_name_upper
 6. Fill cells with value on column cluster_label
 7. Replace value on cell
 8. Replace value on cell
 9. Replace value on cell
 10. Change data type on column cluster_label
 11. Update column on column price
 12. Rename column on column price_copy

mpg	cylinders	displacement	horsepower	weight	acceleration	model	origin	car_name	mpg	cluster_label	PP_classification
1	16	151	109	2645	17.0	chevrolet	usa	chevrolet chevelle malibu	16	regular	0.100000
2	18	161	102	2414	15.2	buick	usa	buick wildcat	18	regular	0.100000
3	13	351	159	3433	10.4	ford	usa	ford torino	13	regular	0.100000
4	16	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	16	regular	0.100000
5	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
6	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000
7	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
8	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000
9	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
10	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000
11	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
12	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000
13	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
14	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000
15	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
16	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000
17	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
18	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000
19	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
20	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000
21	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
22	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000
23	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
24	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000
25	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
26	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000
27	15	181	122	2791	12.7	ford	usa	ford mustang	15	regular	0.100000
28	18	161	102	2414	15.2	chevrolet	usa	chevrolet monte carlo	18	regular	0.100000

Fuente: Proveedor Talend®

Capítulo 2. Explicación del caso elegido

La organización seleccionada para realizar el presente estudio de caso es una compañía operadora de telefonía, que además brinda servicios de conectividad y entretenimiento. Cuenta con aproximadamente 24.000 colaboradores en todo el país, y con alrededor de 29 millones de clientes, a los cuales provee diferentes prestaciones, en todos sus dispositivos, con conexiones fijas y móviles de alta velocidad, y una plataforma de contenidos en vivo y a demanda que incluye series, películas, juegos en línea, música y programas de TV. Brinda soluciones de telefonía fija y móvil, transmisión de datos, televisión paga e Internet, para individuos, empresas e instituciones en todo el país. Abarca esos tres mercados: masivo, corporativo y gubernamental.

Esta empresa se constituyó como consecuencia de la privatización de la Empresa Nacional de Telecomunicaciones (ENTEL) que tenía a cargo la prestación del servicio público de telecomunicaciones en la República Argentina. Inició sus operaciones a fines del año 1990 y prestó el servicio público de telecomunicaciones con un régimen de exclusividad por un período de siete años a partir de la fecha de la transferencia, teniendo derecho a la extensión de la exclusividad por tres años. Luego de ese período inicial, el Gobierno Nacional estableció las pautas para una transición ordenada hacia la libre competencia en telecomunicaciones hasta el 10 de octubre de 1999, fecha a partir de la cual finalizó el período de exclusividad y la Sociedad quedó habilitada para prestar el Servicio Básico Telefónico en todo el país.

En los últimos años el sector de las telecomunicaciones se ha transformado significativamente. La aparición de nuevos servicios y un creciente número de compañías que los ofrecen son consecuencia del proceso de apertura que enfrenta la industria. Respondiendo a las necesidades del mercado y teniendo en cuenta el marco regulatorio vigente, dicha sociedad ha creado en los últimos años distintas empresas que conforman el Grupo. Estas actúan en forma coordinada haciendo más competitivo al conjunto. La compañía también opera en el servicio de telefonía celular móvil, de televisión por cable y televisión digital, mercados a los cuales se expandió mediante la adquisición y fusión con

otras empresas. Esta compañía se encuentra regulada por el ENACOM, Ente Nacional de Telecomunicaciones.

Mercado de Telecomunicaciones

Actualmente, la organización analizada es uno de los tres grandes participantes del sector de telecomunicaciones en Argentina, Paraguay y Uruguay.

De acuerdo con los informes periódicos publicados por el ENACOM, el mercado de telecomunicaciones presenta los siguientes indicadores a marzo del año 2021.

Principales Indicadores a Marzo de 2021

ACCESOS	
Telefonía Fija	7,20 Millones
Acceso a Internet Fija	9,65 Millones
TV por Suscripción	9,61 Millones
Comunicaciones Móviles	56,96 Millones
Telefonía Pública	0,035 Millones

MISCELÁNEAS	
Velocidad media de descarga	43,11 mbps
Portaciones Móviles	14,36 Millones de números
Radiobases 4G instaladas	32.330 Radiobases
Hogares con Internet Fija (%)	67,95 %

PENETRACIÓN	
Telefonía Fija	15,76 Accesos c/100 hab.
Acceso a Internet Fija	21,13 Accesos c/100 hab.
TV por Suscripción	21,04 Accesos c/100 hab.
Comunicaciones Móviles	124,73 Accesos c/100 hab.

FACTURACIÓN	
Telefonía Fija	19.447 Millones de \$
Acceso a Internet Fija	38.257 Millones de \$
TV por Suscripción	46.831 Millones de \$
Comunicaciones Móviles	90.217 Millones de \$

MERCADO POSTAL	
Facturación	18.125 Millones de \$
Producción	199 Millones de unidades

[Para mas Indicadores Nacionales](#)

[Para mas Indicadores por Provincia](#)

Contexto político, económico y social

Respecto de los ingresos, el aislamiento preventivo obligatorio provocó -al menos en los dos primeros meses de cuarentena- aproximadamente, una disminución de los ingresos de la empresa, debido a que, en los inicios de la pandemia, y por disposición del gobierno nacional, se impedía realizar el corte de los servicios por mora en el pago.

A partir de octubre de 2020, el servicio de telecomunicaciones fue declarado servicio esencial en nuestro país, lo que les impidió a las empresas del sector realizar ajustes de precios. Por este motivo, las empresas del sector recurrieron a la justicia, y la organización tuvo que realizar una agresiva política de ajuste interno para reducir sus costos de operación, junto con el fortalecimiento del sector comercial para aumentar su participación en el mercado.

En relación con los productos y consumos, debido a la digitalización profundizada por la pandemia, aumentó la demanda de los abonos de internet de mayor velocidad; y de transmisión de video. En los próximos años, organización planea invertir en el desarrollo de la red 5G en la región, para sumar a los servicios que brinda, servicios de seguimiento, de realidad virtual, de realidad aumentada.

Marco Regulatorio

Como todo ente económico, la organización seleccionada está sujeta a la normativa vigente. Dentro de las normas más relevantes para el presente análisis, podemos destacar: el Reglamento de clientes de los servicios de tecnologías de la información y las comunicaciones, la ley de defensa del consumidor, número 24240 y sus modificatorias, y la ley 25.326, de protección de datos personales, conocida como “habeas data”, entre otras.

Leyes de Protección de Datos

En la Argentina, la protección de los datos personales está regulada por la ley 25.326, sancionada el 4 de octubre del año 2000, cuyo objeto es: “[...] la protección integral de los datos personales asentados en archivos, registros, bancos de datos, u otros medios técnicos de tratamiento de datos, sean estos públicos, o privados destinados a dar informes [...]”. Esta normativa se sustenta en los principios jurídicos de: licitud, calidad de los datos, consentimiento, información, seguridad, confidencialidad y finalidad. El organismo a cargo de velar por su cumplimiento es la Dirección Nacional de Protección de Datos Personales. En virtud de los avances tecnológicos ocurridos desde su sanción, el creciente aumento de las amenazas a la privacidad debido a la explotación masiva de datos, y la actualización de las normas internacionales relativas al tema, existen, al momento del presente análisis, un proyecto de ley del año 2018 tendiente a actualizarla.

A nivel internacional, desde el 25 de mayo de 2018 está en vigencia el Reglamento General de Protección de Datos de la Unión Europea. “Los gobiernos regulan los derechos de propiedad de los datos individuales con políticas de protección de datos como el Reglamento General de Protección de Datos (RGPD) de la Unión Europea” (Fadler & Legner, 2020).²⁴ Este reglamento establece normas relativas a la protección de los datos personales de las personas físicas y la libre circulación de sus datos dentro de la Unión Europea (Reglamento General de Protección de Datos, 2021). Otra norma vigente es la Ley de Privacidad del Consumidor de California (CCPA), la cual otorga a los consumidores residentes en California, derechos sobre la gestión de sus datos personales.

El deber de cumplimiento de las normativas relativas a la protección de datos enfatiza la necesidad de contar con mecanismos de control y gobierno de los datos, en pos de la seguridad de la información que cada organización debe garantizar.

Organigrama de la empresa. Nivel estratégico

En su nivel superior, la estructura de la empresa está compuesta por un directorio conformado por ocho directores titulares, entre ellos el presidente y su vicepresidente, y ocho directores suplentes. A su vez, cuenta con un consejo de síndicos, conformado por tres

²⁴ Traducción propia de: “Governments react to these developments by enforcing individual data ownership rights with data protection policies such as the General Data Protection Regulation (GDPR) in the European Union”

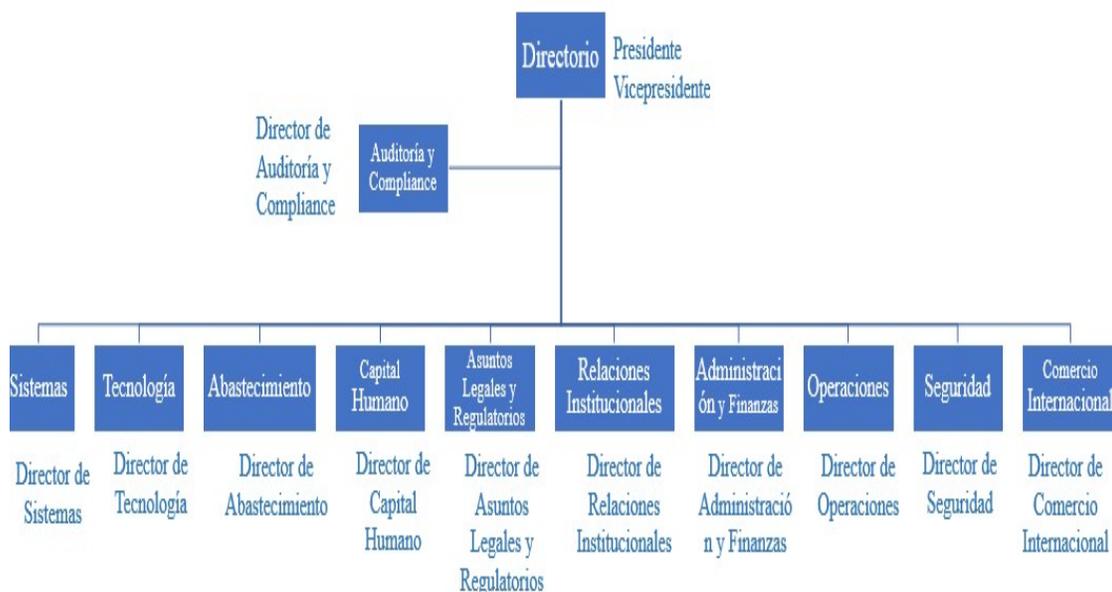
miembros titulares y tres miembros suplentes. Como auditores externos, están contratados dos prestigiosos estudios contables internacionales. El comité de auditoría se compone de tres miembros. La designación de los directores titulares y suplentes y la distribución de los cargos es dispuesta periódicamente por la Asamblea General Ordinaria y Extraordinaria de Accionistas.

El nivel siguiente, las gerencias de primer nivel se dividen en: Sistemas, Tecnología, Abastecimiento, Capital Humano, Asuntos Legales y Regulatorios, Relaciones Institucionales, Administración y Finanzas, Operaciones, Seguridad y Comercio Internacional.

Cada una de estas gerencias, se encuentra a su vez, dividida en distintas gerencias de segundo nivel, las cuales, a su vez, también se dividen en supervisiones.

La empresa cuenta con una estructura matricial, por la cual, cada colaborador reporta tanto a su nivel inmediato superior, como al gerente del o los proyectos en los cuáles está involucrado.

En total, cuenta con aproximadamente 23000 colaboradores internos en la región, más recursos externos tercerizados en diversas áreas.



Cultura basada en datos

Desde su privatización, la empresa realizó diversos cambios progresivos para instalar una cultura basada en datos. Este cambio fue promovido y apoyado por el nivel estratégico de la organización a través de distintas acciones concretas, como la incorporación de las inversiones en tecnología en el plan estratégico. Como parte de este proceso, se inició la creación de distintos repositorios de información o “data marts” específico para cada área interesada, como facturación y marketing, por ejemplo, y posteriormente, se realizó la construcción de un datawarehouse corporativo; con la participación de las áreas usuarios. Se creó un área de inteligencia de negocios, compuesta por dos grandes equipos, uno encargado de los aspectos técnicos relacionados con las bases de datos y los procesos de extracción, transformación y carga de datos, y otro, más orientado a los aspectos funcionales, encargado de interpretar las necesidades de información y más relacionado con el negocio. Este equipo contaba con aproximadamente sesenta colaboradores. En cuanto a las herramientas informáticas, se adquirieron diversas herramientas para la visualización y explotación de datos, tales como Microstrategy® y Qlik View®. Con el correr del tiempo y con el objetivo de ordenar y priorizar los proyectos se creó otra área encargada de la gestión de requerimientos, y un comité de usuarios para su priorización. A medida que los reportes provistos por el área de negocios demostraron ser fiables y de calidad, y entregados en tiempo y forma, más áreas usuarias se incorporaron como consumidoras del datawarehouse corporativo. Algunas áreas también utilizaron las herramientas de explotación disponibles para explotar directamente la información de los sistemas aplicativos. Este aumento de usuarios, objetos, aplicaciones, repositorios, etc. dio lugar a la creación e intervención del área de arquitectura tecnológica. La información “corporativa” pasó a ser considerada “oficial” y “crítica” para la operación del negocio, y se transformó en el sustento de todos los procesos decisorios dentro de la organización. A raíz de esto, se creó luego, el área de “data mining” o “analítica” con la finalidad de aplicar técnicas más avanzadas de explotación de datos, y aplicación de modelos estadísticos que permitan descubrir patrones y realizar predicciones. Esta área, era consumidora del datawarehouse corporativo, utilizaba la herramienta SaS®. Con el correr del tiempo, y el aumento de la volumetría de datos que se generó, la organización decidió crear una nueva área para explorar el uso de BigData. Se crearon “datalakes” o “lagos de datos” y se utilizó la herramienta Hadoop.

De acuerdo con el modelo de Berndtsson, Forsberg, Stein, & Svahn, podría decirse que la organización se encuentra entre el nivel 3 y 4 de madurez de la cultura basada en datos.

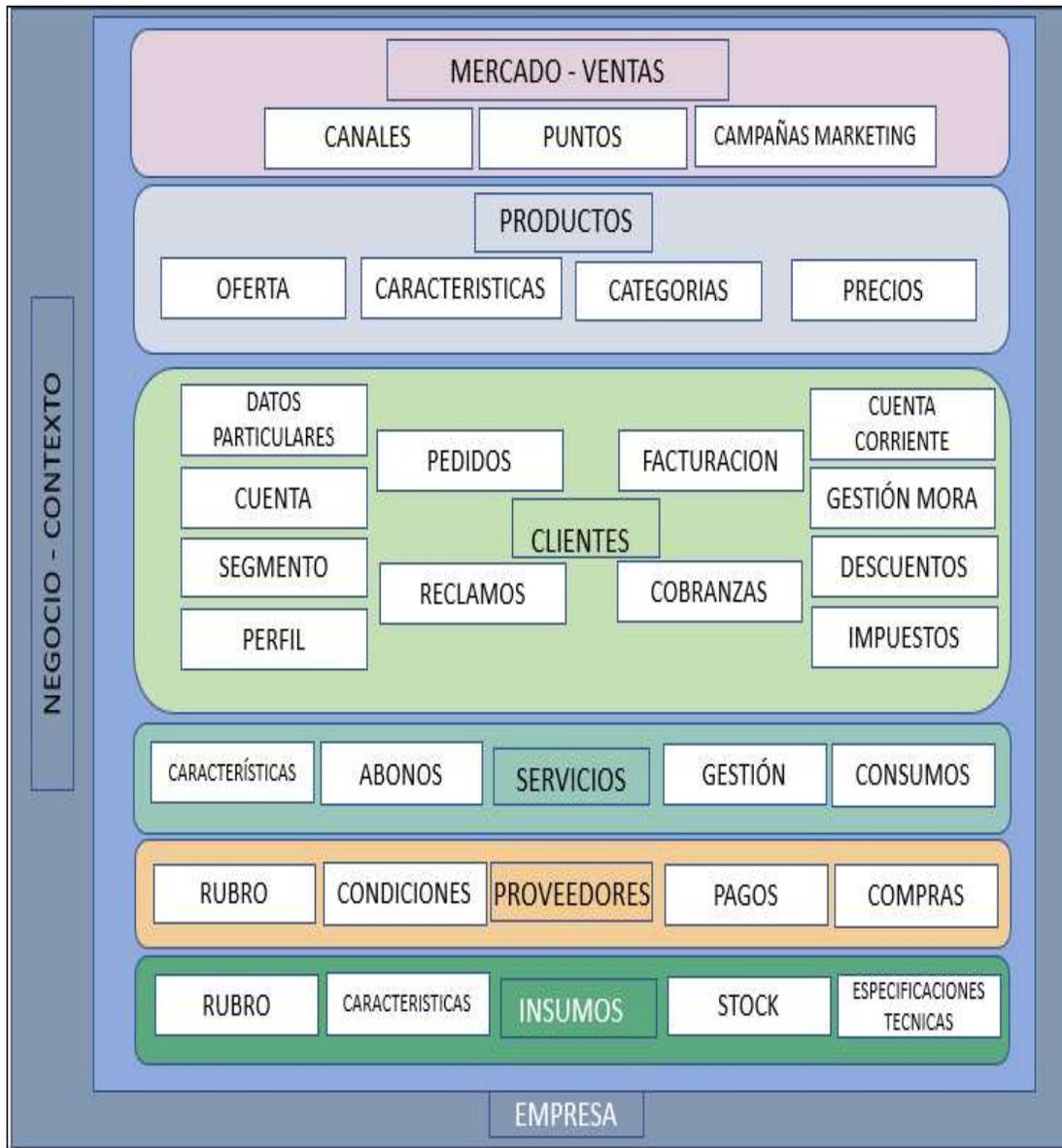
Arquitectura Empresarial

En virtud del proceso de transformación digital llevado a cabo por la empresa, el área de Arquitectura e Infraestructura TI, comenzó a trabajar desde hace aproximadamente tres años en el ordenamiento y la unificación de su arquitectura empresarial. Como marco de trabajo, se optó por la aplicación de TOGAF®. Este marco distingue cuatro tipos de arquitectura, a saber: arquitectura de negocio, de datos, de aplicación y tecnológica. Estos cuatro tipos, en menor o en mayor medida, tienen estrecha relación con el gobierno de datos y su dominio de aplicación. En primer lugar, la arquitectura de negocios está compuesta “por la estrategia de negocio, gobierno, organización y procesos clave de la organización (The Open Group)”, por lo tanto, la estrategia de datos, el gobierno de datos y todos los procesos asociados a la gestión de datos, la gestión de su calidad y seguridad son componentes de este tipo de arquitectura. La arquitectura de datos es el tipo de arquitectura más estrechamente relacionado con el programa de gobierno de datos, ya que se define como: “la estructura de datos lógicos y físicos que posee una organización y sus recursos de gestión de datos (The Open Group)”. El trabajo del área en este dominio fue el primer paso para posibilitar la declaración de políticas de datos y sus principios, los cuales dieron origen al diseño del programa de gobierno, ya que es fundamental tener claramente definidos y relevados los activos de datos. En cuanto a la arquitectura de aplicaciones, definida como “un plano de las aplicaciones individuales a implementar, sus interacciones y sus relaciones con los procesos de negocio principales de la organización (The Open Group)” es importante su identificación, relevamiento y documentación ya que esta información sirve para determinar el linaje de los datos. Específicamente, en cuanto a la arquitectura de datos el área trabajó en la documentación y definición de los artefactos propuestos por el marco de trabajo citado, a saber:



Fuente: TOGAF®

Del relevamiento realizado, surge que la organización identificó alrededor de 1000 entidades de negocio y cerca de 2000 atributos. Las entidades de negocio son aquellas unidades lógicas de análisis relevantes para una organización, como pueden ser, por ejemplo, sus clientes, sus productos, proveedores, etc. Estas entidades son transversales a los distintos sistemas y procesos, y engloban diversos atributos. A su vez, están agrupadas en “dominios”, es decir conjunto de entidades de negocio asociadas a una misma gerencia o proceso. Como paso siguiente se diagramó un modelo de datos único, conteniendo las entidades de negocio identificadas y sus relaciones y semántica; con el objetivo de contar con una visión integral y completa de los datos para facilitar la gestión de la información. El diagrama de datos conceptual, a muy alto nivel, quedó definido de la siguiente manera:



Las entidades de negocio son aquellas representadas por rectángulos, y se encuentran agrupadas en dominios (asociaciones lógicas), graficadas por rectángulos con bordes redondeadas. La identificación de las entidades de negocio claves, cuyo registro en forma física es compartido por diversas aplicaciones, es fundamental ya que supone un gran esfuerzo en la sincronización e integración de los datos, debido a que deben ser consistentes y estar actualizados cualquiera sea la aplicación que los utilice durante la operatoria de la organización.

Principios arquitectónicos de los datos:

La aplicación del marco de trabajo mencionado implicó también la definición de los principios rectores de la gestión de datos:

- Los datos son un activo estratégico y tienen valor económico.
- Los datos deben ser el soporte de las decisiones.
- Los datos deben tener calidad, ser confiables, accesibles y únicos para toda la organización.
- En todo paso del ciclo de vida de los datos se debe velar por su seguridad.

Estos lineamientos deben ser tenidos en cuenta no sólo en la operatoria, sino también en el diseño de los procesos, y elección de metodologías, herramientas y artefactos organizacionales.

Capítulo 3. Implementación del programa de Gobierno de Datos

Situación inicial

Ante el creciente valor de los datos como activo económico y la necesidad de velar por el cumplimiento de políticas de calidad asociadas a éstos, sumado a la complejidad de la gestión de la información, el nivel estratégico detectó la necesidad de implementar un programa de Gobierno de Datos. En un principio, decidió encargar la tarea de llevar adelante dicho programa al área de Analítica Avanzada de la organización. Sin embargo, la misma, no se encontraba en condiciones de absorber las horas de trabajo necesarias para realizarlo. Debido a esta situación, se tomó la decisión de destinar más recursos a este objetivo.

Como primer paso, el nivel estratégico de la organización definió por escrito las políticas del Gobierno de Datos de la empresa, y como consecuencia de éstas, se creó la gerencia de Gobierno de Datos, y se diseñó el programa de Gobierno de Datos, entre otros mecanismos de gobierno. El gerente designado de la flamante área realizó una investigación acerca de los marcos de trabajo más utilizados en la actualidad para la aplicación de estos programas. Como resultado del análisis, se optó por elegir los componentes de cada marco evaluado que resultaran más apropiados para la organización de acuerdo con su cultura, con base principal, en el marco de trabajo DAMA®.

Análisis de los mecanismos utilizados

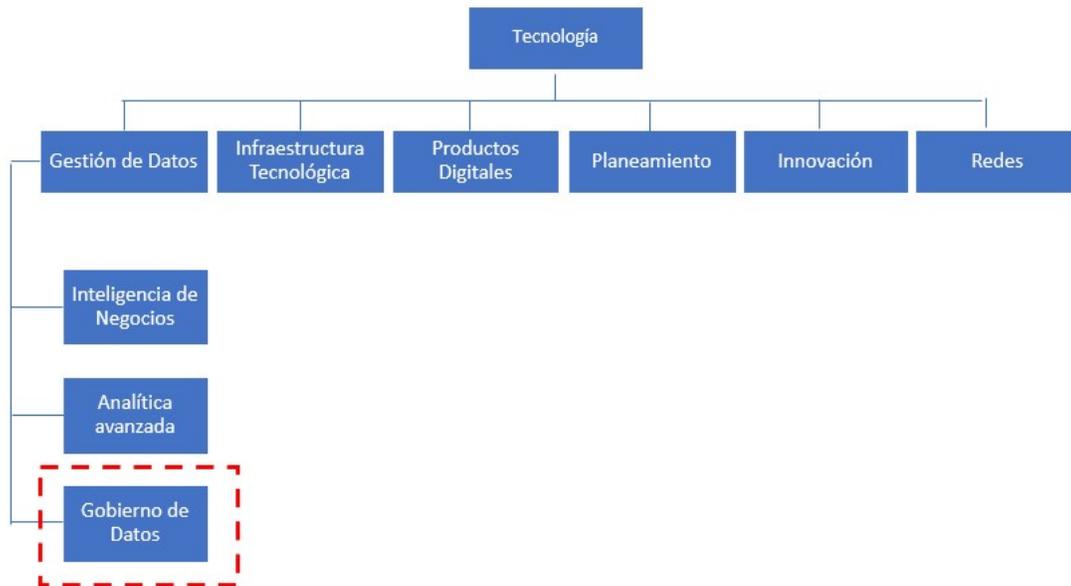
Mecanismos estructurales.

Estructura organizacional

La compañía optó por realizar una transformación organizacional reordenando su estructura jerárquica, y transformándola en una estructura matricial, para otorgar un papel preponderante a los datos. Se creó el área de “Datos IT”, que depende funcionalmente del área de Desarrollo IT de la dirección de Tecnología, pero paralelamente, tiene dependencia matricial de las áreas de Comercialización y Gestión de Datos.

Dentro del área de Gestión de Datos, se creó una sub-área de “Gobierno de Datos”, con el mismo nivel jerárquico que el área de inteligencia de negocios y el área de “analítica

avanzada”. El área existente de inteligencia operativa, que anteriormente dependía íntegramente del área de desarrollo de TI, fue absorbida por el área de inteligencia de negocios, y el área de data-mining se transformó en el área de analítica avanzada, y se creó una unidad que engloba a ambas y al área de gobierno, dependiente del “Chief Information Officer” y ya no del “Chief Technology Officer”.



A su vez, a esta estructura jerárquica formal se superpone una estructura matricial para proyectos de datos que requieren de la participación de las áreas usuarias.

Roles y responsabilidades

Al realizar los cambios estructurales descriptos, se reubicaron colaboradores internos y externos, se realizó una búsqueda específica de profesionales de la ciencia de datos, y se designó un responsable del área en formación de gobierno de datos. El resto del equipo del equipo está en proceso de formación y crecimiento al momento del presente análisis.

Esta nueva área de Gobierno de datos tiene rango de gerencia, y sus responsabilidades son fomentar la implementación de las mejores prácticas de gobierno de datos, a partir de la aplicación de procesos y estándares, a fin de optimizar la utilización de datos de forma alineada con la estrategia organizacional; y llevar a cabo la implementación del programa de

Gobierno de Datos de la empresa, el cual se define como un programa continuo de proyectos relacionados con el gobierno de datos.

Las responsabilidades específicas que se asignaron al área son:

- Dar apoyo y seguimiento a todas las actividades del gobierno de datos, implementando las herramientas necesarias y promoviendo su correcta documentación y difusión dentro de la organización.
- Asesorar en aspectos de seguridad de la información, privacidad y cumplimiento regulatorio a todas las partes involucradas en el ciclo de vida de la información.
- Participar y acompañar la definición y asignación de los diferentes roles de la comunidad de datos: dueños, custodios y responsables.
- Programar y liderar y/o acompañar la realización de los comités de datos definidos (gobierno, calidad, etc.) y acompañar la ejecución de los proyectos estratégicos definidos en éstos.
- Diseñar, definir e implementar métricas de cumplimiento del gobierno de datos.
- Garantizar que se cumplan todos los parámetros de calidad, seguridad y cumplimiento en la compra de datos externos o monetización de datos internos.
- Promover y difundir la definición de términos de negocio corporativos y únicos y la debida documentación de los ambientes de datos, modelos, procesos, etc.
- Brindar enlace o coordinación entre los distintos programas de calidad de datos, cumplimiento, arquitectura de datos y aplicaciones, seguridad de la información y monetización de datos.
- Definir estrategias de implementación de procesos y estándares, concentrar y alinear políticas, normas y procedimientos de las actividades relacionadas con el gobierno de datos.

En la política de Gobierno de Datos establecida por la organización se establecieron los roles esperados dentro de la comunidad de datos:

- Dueño de los datos
- Gestor de los datos
- Custodio de los datos
- Áreas de apoyo

y se los definió de la siguiente manera:

Dueño de Datos:

El “dueño de los datos” es “el responsable de la gestión de los ambientes de datos donde el dato se genera conoce el negocio y tiene la potestad para decidir sobre las definiciones, los estándares de seguridad y privacidad, calidad, cumplimiento de las normativas y regulaciones. Interviene en la compra de datos externos y en la monetización de los datos de los cuales tiene la propiedad”. Dentro de sus responsabilidades, que están circunscriptas a un dominio de datos particular, se encuentra:

- Dar seguimiento al cumplimiento de las políticas, estándares, procesos y procedimientos de su dominio de datos.
- Participar en los comités a los que son convocados.
- Asegurar la calidad de datos de su dominio.
- Definir glosarios de datos.
- Aprobar los accesos a sus dominios de datos.
- Administrar los riesgos y niveles de criticidad y confidencialidad de su dominio.
- Determinar el origen mandatorio de los datos (sistema o repositorio fuente).

Gestor de Datos:

Los gestores de datos tienen a cargo la ejecución de las políticas, estándares, procesos y procedimientos definidos en base a la estrategia de gobierno, sobre los ámbitos de datos que han sido asignados como referentes. Sus principales funciones son:

- Asegurar que los datos se administren adecuadamente, sean consistentes y cumplan con los estándares de calidad.
- Gestionar los metadatos.
- Participar en la definición de procedimientos y estándares para mejorar el gobierno de los datos.
- Recomendar mejoras en los procesos de gestión de datos.

Custodio de Datos:

Está a cargo de las actividades de gobierno de datos sobre el dominio que tiene asignado. Debe poseer un amplio conocimiento de los procesos de captura, almacenamiento, distribución y consumo de datos en la arquitectura de datos de la organización. Dentro de sus funciones, se encuentra, principalmente, la custodia de los datos, el proporcionar el acceso a éstos de acuerdo con las reglas definidas por el dueño de datos, y la disposición de estos para su consumo y monetización.

Áreas de apoyo:

Se trata de aquellas áreas que deberán dar soporte al gobierno de datos como equipo extendido, con foco en una temática específica; como, por ejemplo: seguridad de la información, servicios legales, arquitectura TI y sistemas.

Estas definiciones se encuentran detalladas en las políticas, pero aún no se ha llevado a cabo el proceso de relevamiento, identificación y asignación de responsabilidades sobre los activos de datos a nivel de dominio de datos en forma abstracta, sin embargo, si están definidos los propietarios, gestores y responsables a nivel de aplicación y de repositorios físicos de datos. Tampoco se cuenta con herramientas informáticas para apoyar este proceso ni su gestión posterior.

Coordinación y alineamiento

Junto con la creación del área de Gobierno de Datos, se definieron políticas por escrito, que fueron difundidas a través del portal interno de empleados, promocionadas a través de los canales de chat corporativos, y a través de videos de difusión, además de comunicaciones formales internas a través de canales tradicionales como el correo electrónico, con el objetivo de informar a todos los colaboradores de la importancia del gobierno de datos y de su responsabilidad en sus puestos de trabajo específicos como participantes de la “comunidad de datos”. A su vez, se dispuso la creación de diversos comités, a saber:

- Un comité estratégico de Gobierno de Datos
- Un comité operativo de Gobierno de Datos

Comité estratégico de Gobierno de Datos

Este comité está compuesto por el área de gobierno de datos y los dueños de datos; más precisamente por el gerente del área de Gestión de datos, el gerente del área de Gobierno de Datos, los miembros de las áreas de soporte y los dueños de datos citados específicamente. La periodicidad de realización es bimestral. Su objetivo es coordinar y administrar los equipos encargados de velar por el cumplimiento de políticas, estándares y lineamientos de gobierno de datos, y la implementación y operación de las funciones del gobierno, para garantizar los estándares de calidad deseados y el acceso, uso y monetización de datos.

Entre sus funciones se encuentra:

- Priorizar y definir proyectos estratégicos de datos alineados a la estrategia de datos de la organización.
- Definir, seleccionar, actualizar y/o modificar estándares, procesos, procedimientos y métricas e indicadores relativos a la calidad de los datos.
- Asegurar el cumplimiento de normativas relacionadas con la gestión de metadatos, calidad, seguridad, privacidad de los datos.

Comité operativo de Gobierno de Datos

Se compone por el gerente del área de Gobierno de Datos, los miembros de las áreas de Soporte y los gestores/custodios de datos citados particularmente en la convocatoria. Su frecuencia es mensual, y pueden ser invitados otros miembros en función de los temas a tratar. Su objetivo es tratar los proyectos e incidencias relacionados a la gestión de datos.

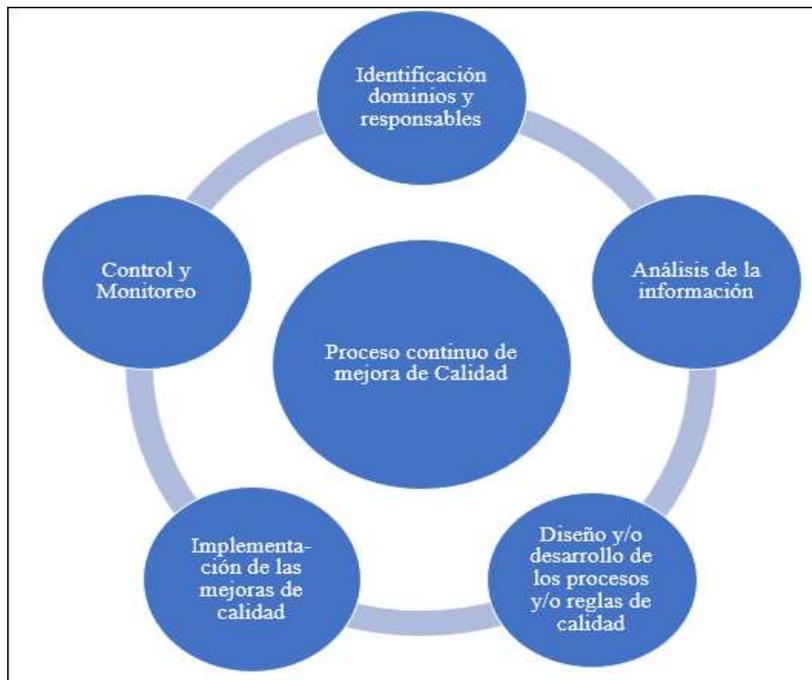
Los comités se han realizado desde la aprobación de las políticas de gobierno hace aproximadamente un año.

Mecanismos de Procesos

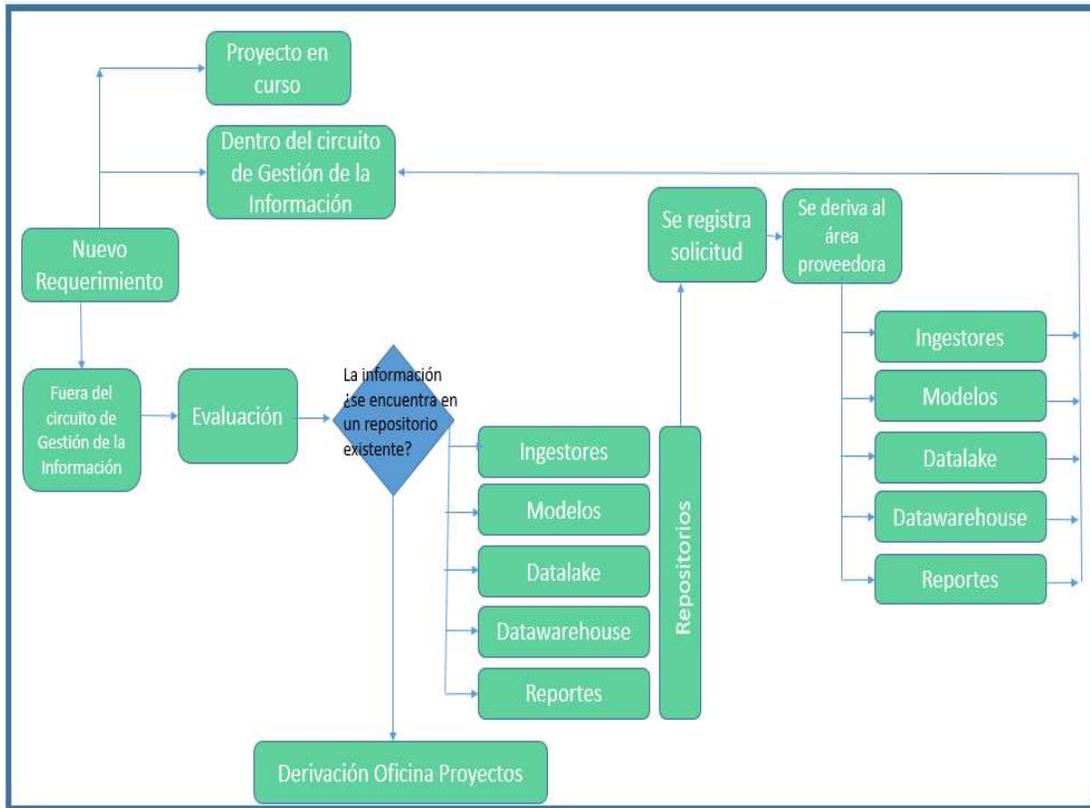
Modelo de procesos

Respecto de los procesos formalizados por el área de Gobierno de Datos al momento del presente análisis, sólo se definieron tres macroprocesos (nivel 1), de forma muy general, a saber:

- El macroproceso de calidad de los datos: consiste en identificar posibles mejoras a las reglas de calidad existentes e implementarlas en cada dominio de datos de acuerdo con sus particularidades, en acuerdo con los propietarios y responsables de dicho dominio.



- El macroproceso de gestión la demanda:



- El macroproceso de gestión de servicios de información:



El área también trabajó fuertemente en la definición de sus estándares de datos internos.

Se definieron tres: el estándar de nombres, que abarca a todos los objetos de datos (campos, tablas, reportes, repositorios, etc.), el estándar de calidad de datos y el “perfilado” de datos, es decir, el conjunto de reglas que estos deben cumplir.

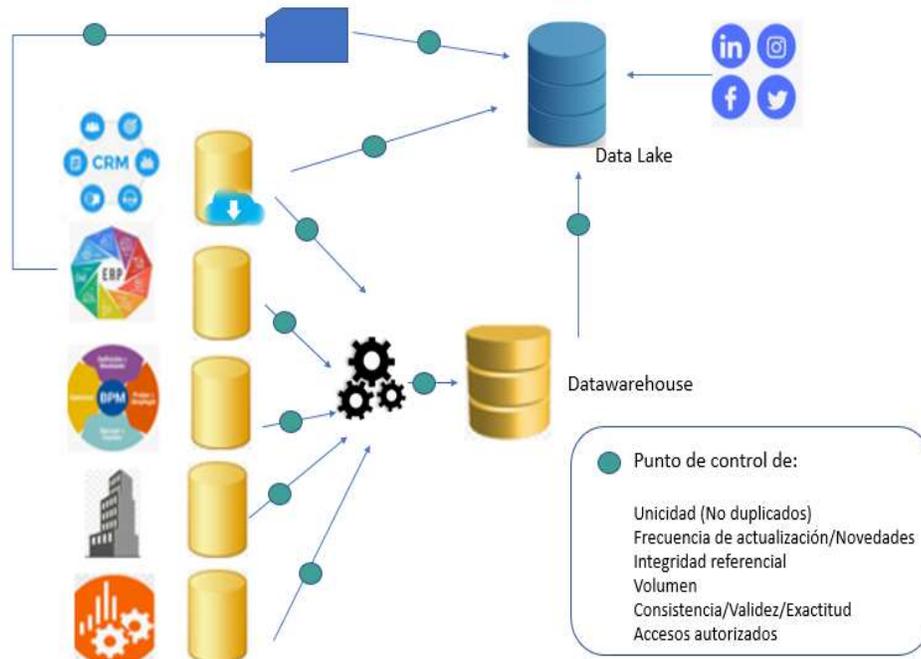
Para ello se basó en los estándares propuestos por organismos oficiales tales como: el Instituto Geográfico Nacional, el Instituto Nacional de Estadística y Censos de Argentina (INDEC) y la Organización Internacional de Normalización (ISO), entre otros. La definición de estándares internos tiene por objetivo el aseguramiento de su confiabilidad, y facilitar la integración y comparación entre fuentes de origen diversas, y a su vez, permitir la posibilidad de relacionarlas con datos externos a la organización.

Monitoreo y evaluación

La definición de estándares internos constituye el punto de partida para el monitoreo y evaluación de su cumplimiento. Como se mencionó en el punto anterior, se establecieron criterios para el perfilado de datos. Estos criterios deben ser tenidos en cuenta al momento de evaluar la calidad de los datos. En cada uno de los puntos de control, se debe verificar que los datos:

- sean consistentes entre los distintos repositorios (los datos en destino deben ser iguales a los datos del origen, con las transformaciones adecuadas si corresponde)
- sean únicos, es decir, no deben estar repetidos
- cumplan las características requeridas (tipos de datos, longitud del campo, definición de claves primarias, foráneas, entre otras)
- sean válidos (pertenezcan al dominio que le corresponda y contengan valores pertenecientes dentro del conjunto de datos admitidos o esperados)
- estén completos (control de nulos y/o vacíos)
- sean congruentes entre sí
- y sean actualizados con la frecuencia requerida de acuerdo con sus características

Estos controles deben estar automatizados y realizarse en los distintos puntos de control identificados para el ciclo de vida de los datos:



Fuente: Elaboración propia basada en relevamiento

Desarrollo

Respecto de los modelos de desarrollo, la organización, en todas las sub-áreas de IT, utiliza, preferentemente metodologías ágiles. Los procesos formales definidos sobre cómo abordar los requerimientos, no tienen ninguna particularidad específica para las áreas relacionadas con la gestión de la información.

Mecanismos relacionales

Visión compartida

Los equipos de trabajo son multidisciplinarios, cada miembro tiene un perfil diferente, por lo tanto, para facilitar la cohesión de estos, y alcanzar una visión compartida, se enfatiza el trabajo en equipo, y las reuniones periódicas entre las diferentes áreas participantes de los procesos de gestión de la información.

Colaboración

Para el trabajo colaborativo en los equipos de desarrollo relacionados con la gestión de la información, la organización cuenta con herramientas para compartir archivos, realizar video llamadas en línea, como Microsoft Teams® y para gestionar los proyectos, incidencias y los proyectos, se utiliza Jira® Software. Otra herramienta de comunicación interna disponible es Yammer.

Transferencia del conocimiento

Para la gestión del conocimiento explícito se cuenta con herramientas informáticas para consultar procesos y procedimientos, repositorios de archivos compartidos y un portal interno para colaboradores donde se encuentra resguardada toda la documentación funcional y técnica de los proyectos de tecnologías de la información finalizados y en curso. En el mismo repositorio, cada área puede publicar toda la información relativa a su ámbito de incumbencia que necesite difundir. Esta base de conocimiento está generada con la herramienta Confluence® y se encuentra integrada con Jira®. En esta última herramienta se almacena información acerca de la forma de resolución de cada incidencia, la cual puede ser consultada cada vez que sea necesario. Existe también un repositorio donde se almacenan todas las reuniones grabadas a través de la herramienta de videoconferencias Microsoft Teams®, las cuales sólo pueden ser consultadas por usuarios autorizados. En el mismo directorio se guarda una grabación de todos los cursos dictados en línea.

Según los arquetipos propuestos por Fadler, Lefebvre , & Legner, y dado que la organización se encuentra focalizada en asegurar y mejorar la calidad de los datos maestros, podría decirse que su arquetipo de Gobierno de Datos es de tipo 1, porque a nivel estructural, tiene un área centralizada para dichas tareas, sus recursos y procesos están enfocados en la obtención de calidad y sus mecanismos de alineación y coordinación, el área de TI tiene un papel predominante.

Conclusiones generales

El gobierno de datos es un complemento necesario e importante del gobierno corporativo y del gobierno de TI dentro de la organización. A su vez, es fundamental para resguardar el valor de los datos como activo estratégico. La importancia de estos como un bien económico, sumado a las normas vigentes sobre protección de datos hace necesaria su implementación. Determinar en quiénes recae la responsabilidad de su gestión y custodia tiene, no sólo efectos prácticos en la operatoria cotidiana, sino que también permite disuadir conflictos y comunicar a nivel organizacional que éstos son tratados con seriedad y se vela por su calidad, a la vez que dirime discrepancias sobre su validez.

De acuerdo con el análisis efectuado sobre el caso de estudio elegido, basado en el relevamiento y análisis de los mecanismos implementados, entrevistas con involucrados, y observación directa, se concluye que la implementación de un programa de Gobierno de Datos no se trata ni de un proceso rápido ni mucho menos sencillo. Tampoco se obtienen resultados visibles en corto y mediano plazo. El programa de Gobierno de Datos abarca a todas las áreas partícipes en el ciclo de vida de la información, e implica una transformación y/o adecuación de políticas, procesos, procedimientos y estándares que afectan a la organización en su conjunto, pero especialmente, requiere cambios en la cultura organizacional, principalmente en todo lo relacionado al uso, distribución y democratización de la información. Adicionalmente, exige no sólo de conocimiento profundo de su ciclo de vida dentro de la organización, sino que también, requiere conocimiento del negocio, de los procesos organizacionales y de tecnologías informáticas de gestión, transformación, exploración, explotación y distribución de datos. Necesita sostenerse a partir del correcto funcionamiento y colaboración de diversas áreas clave, como tecnología, desarrollo de aplicaciones, seguridad de la información y arquitectura, además de las áreas de negocio.

Al mismo tiempo, conlleva un gran esfuerzo organizacional, que se ve afectado por las mismas resistencias e incertidumbres que se generan al momento de la implementación de una nueva tecnología o un cambio sustancial en la estructura jerárquica. Esto implica que la adecuada gestión del cambio es un factor crítico para el éxito que no se puede descuidar, resulta necesario acompañar a los colaboradores en la transición, dado que se generan responsabilidades extras, nuevas tareas administrativas, se suman controles que demandan tiempo, aparecen nuevos actores en los procesos decisivos relacionados con los datos,

nuevos puestos, más reuniones, lo que supone trabajo adicional para todos los involucrados, lo que lleva a la consecuente resistencia al cambio.

En el caso presentado, el directorio, al vislumbrar la necesidad del gobierno de datos, originalmente, asignó las responsabilidades sobre la gestión de estos y su calidad, de forma descentralizada y distribuida entre las áreas relacionadas con la gestión de datos.

Durante esta etapa, y debido a la poca disponibilidad de dichas áreas para asumir las tareas necesarias y llevar a cabo los cambios requeridos, se optó por crear un área específica que las centralizara y asumiera la responsabilidad de llevar a cabo el programa.

El primer paso fue definir por escrito la política de datos de la compañía. A partir de ésta, se implementó el primer mecanismo estructural: la creación de una nueva gerencia específica: Gobierno de Datos. Luego, el responsable de la nueva área llevó a cabo un relevamiento de los marcos de trabajo difundidos en la actualidad, por proveedores de soluciones de explotación de datos, organismos internacionales, asociaciones profesionales y estándares, y en base a su análisis, optó por diseñar un marco de trabajo propio para la organización, a partir de la selección de los componentes más relevantes de los marcos estudiados, realizó adaptaciones de acuerdo a las particularidades y la cultura organizacional. Principalmente, se basó en las prácticas recomendadas por el marco de trabajo DAMA®.

Del análisis realizado sobre los mecanismos utilizados, surge que se otorgó la misma importancia a los tres tipos: estructurales y de procesos y de coordinación.

En relación con la creación de una nueva gerencia con el propósito de llevar a cabo el programa de gobierno, se puede señalar que dicha área cuenta sólo poder formal, su equipo se encuentra subdimensionado y dada la complejidad y la dimensión de la arquitectura de aplicación y de datos de la empresa, sumado a la falta de herramientas informáticas que faciliten la realización de sus tareas, (actualmente son manuales) aún no se visualizan los resultados esperados.

En cuanto los mecanismos de procesos, se definieron las políticas de calidad y se aprobó el estándar de calidad de datos con el objetivo de asegurar su confiabilidad y posibilitar la integración y comparación entre las diferentes fuentes de origen, y la posibilidad de relacionarlas con datos externos. Para la definición de sus estándares internos, la

organización recurrió a estándares de calidad de la industria y de organismos oficiales como modelo. Sin embargo, se puede mencionar que el relevamiento y unificación de los responsables, gestores y custodios existentes o potenciales de las entidades corporativas, no se ha finalizado. Esto tiene como consecuencia que, si bien las políticas, macroprocesos y estándares definidas son conocidas por toda la organización, desde el punto de vista teórico, no tienen una aplicación práctica y real, al menos por el momento. Para acompañar estos procesos y agilizar las actividades, se han evaluado diversas herramientas disponibles en el mercado, pero aún no se ha seleccionado ni implementado ninguna, lo cual dificulta visiblemente el avance de las tareas operativas del área.

Con respecto a los factores que apalancan el éxito de estos programas, puede decirse que se ha iniciado un trabajo de difusión y concientización de la importancia del gobierno de datos, sin embargo, en muchas oportunidades aún no se convoca al área para participar en nuevos proyectos en donde se tiene la oportunidad incorporar las buenas prácticas de gobierno desde el inicio. Desde la creación de la unidad mencionada, sólo se la convocó a participar como actor en un proyecto incipiente y por iniciativa del proveedor externo a cargo del proyecto, que tiene incorporado participar al área de gobierno como una buena práctica; no así en proyectos llevados a cabo por personal interno.

La arquitectura de aplicaciones y de datos se encuentra en constante cambio debido a la transformación digital de la empresa, y al no contar con la información sobre quienes ejercen los roles para cada fuente digitalizada de manera centralizada, se dificulta mucho su actualización de manera manual en planillas de cálculo; y también dificulta la convocatoria y realización de los comités establecidos por la política de datos. También se dificulta la creación del diccionario de términos de negocio y de datos, herramientas que también son importantes para las actividades de gestión, custodia y aseguramiento de la calidad.

Se considera importante avanzar en la incorporación de una herramienta informática que permite agilizar y automatizar las tareas operativas del área, aumentar la participación y el involucramiento de los miembros de la gerencia en las decisiones que impactan a los sistemas de la organización, para que puedan aportar la visión del gobierno de datos.

Por lo expuesto, y a partir de la aplicación del modelo CMMI, se desprende que la organización se encuentra en un estadio de madurez de grado 4 de dicho modelo, ya que todos los componentes del programa de gobierno de datos están implementados o próximos

a implementarse, lo cual implica que están “administrados”. A su vez, se observa que están dadas las condiciones para que pueda evolucionar, en el mediano plazo, al estadio 5, previo afianzamiento de estos. Para alcanzar dicho objetivo, será necesaria la implementación de métricas de cumplimiento que permitan conocer el estado de situación y realizar mejoras de forma continua para poder optimizar los procesos relativos al gobierno de datos, continuar fomentando la participación del área en todas las decisiones que impacten la arquitectura de los datos, y acompañar el proceso con adecuadas técnicas de gestión del cambio, para lograr la aceptación de los nuevos roles, desafíos, controles y procesos.

Bibliografía

- Abbasi, A., Sarke, S., & Chiang, R. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, 17, 32.
- Abraham, R., Schneider, J., & vom Brocke, J. (2019, Diciembre). Data governance: A conceptual framework, structured review, and research. (Elsevier, Ed.) *International Journal of Information Management*, 49, 424-438. doi:<https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- Argentina.gob.ar. (2021, 5). Retrieved from <https://www.argentina.gob.ar/aaip/datospersonales/proyecto-ley-datos-personales>
- Baijens, J., Helms, R., & Velstra, T. (2020). Towards a Framework for Data Analytics Governance Mechanisms. *Proceedings of the 28th European Conference on Information Systems (ECIS), An Online AIS Conference*. 81. Marrakech: ECIS 2020 Research Papers. Retrieved from https://aisel.aisnet.org/ecis2020_rp/81
- Berndtsson, M., Forsberg, D., Stein, D., & Svahn, T. (2018). Becoming a Data-Driven Organisation. (E. 2. (AISEL), Ed.) 43. Retrieved from https://aisel.aisnet.org/ecis2018_rip/43/
- Briano, J. C., Freijedo, C. F., Rota, P., Tricoci, G., & Waldbott de Bassenheim, C. (2011). *Sistemas de Información Gerencial. Tecnología para agregar valor a las organizaciones*. Prentice Hall - Pearson Education.
- Brynjolfsson, E., Hitt, L., & Kim, H. H. (2011, abril 22). <https://ssrn.com>. doi:1819486
- Cheong, L., & Chang, V. (2007). The Need for Data Governance: A Case Study. In A. f. (AISEL) (Ed.), *18th Australasian Conference on Information System*. 100. Toowoomba: Association for Information Systems AIS Electronic Library (AISEL).
- Chinkes, E. (2008). *Business Intelligence para mejores decisiones de negocio*. Buenos Aires: EDICON-Fondo Editorial Consejo. Consejo Profesional de Ciencias Económicas de la Ciudad Autónoma de Buenos Aires.
- DAMA International. (2009). *The DAMA Guide to Data Management Body of Knowledge (DAMA-DMBOK GUIDE)*.
- ENACOM - Ente Nacional de Comunicaciones. (2021, 5). Retrieved from <https://enacom.gob.ar/>
- Fadler, M., & Legner, C. (2020). Who owns the data in the enterprise? Rethinking data ownership in times of big data analytics. In A. f. (AISEL) (Ed.), *European Conference on Information Systems*. 207. Marrakech: ECIS 2020 Proceedings Research Papers. Retrieved from https://aisel.aisnet.org/ecis2020_rp/207/
- Fadler, M., Lefebvre, H., & Legner, C. (2021). Data governance: from master data quality to data monetization. In A. f. (AISEL) (Ed.), *Twenty-Ninth European Conference on Information Systems*. 155. Marrakech: ECIS 2021 Research Papers. Retrieved from https://aisel.aisnet.org/ecis2021_rp/155
- Gartner. (2021). Retrieved from <https://www.gartner.com/en>
- Grossman, R., & Siegel, K. (2014). Organizational Models for Big Data and Analytics. (O. D. Community, Ed.) *Journal of Organization Design*, 20-25. doi:10.7146/jod.9799
- IBM. (2021, 12). Retrieved from <https://www.ibm.com/ar-es/analytics/data-governance>
- Infoleg.gob.ar. (2021, 5). Retrieved from <http://servicios.infoleg.gob.ar/infolegInternet/anexos/60000-64999/64790/norma.htm>
- Informatica. (2021, 12). Retrieved from <https://docs.informatica.com>
- ISO. (2021, 5). Retrieved from <https://www.iso.org/standard/56639.html>

- Khatri, V., & Brown, C. (2010, 01). Designing data governance. *Communications of the ACM*, 53(1), 148-152. doi:10.1145/1629175.1629210
- Larose, D., & Larose, C. (2014). *Discovering knowledge in Data. An introduction to Data Mining*. (I. C. Society, Ed.) New Jersey, USA: Wiley.
- Lee, S., Zhu, L., & Jeffery, R. (2019). Data Governance Decisions for Platform Ecosystems. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, (p. 10). doi:10.24251/HICSS.2019.766
- Lis, D., & Otto, B. (2020, 08). Data Governance in Data Ecosystems – Insights from Organizations. (A. f. (AISeL), Ed.) *Amcis 2020 Proceedings, 12*. Retrieved from https://aisel.aisnet.org/amcis2020/strategic_uses_it/strategic_uses_it/12/
- Maiante, H. A. (2016, 05). Gestión por procesos. *Material preparado por el Profesor A. Regular (FCE/UBA) Hernán A. Maiante*.
- Mayer-Schönberger, V., Cukier, K., & Traducción de Iriarte, A. (2013). *Big Data. La revolución de los datos masivos*.
- Mikalef, P., & Krogstie, J. (2018). Big Data Governance and Dynamic Capabilities: The Moderating effect of Environmental Uncertainty. In P. 2. Proceedings (Ed.), *Pacific Asia Conference on Information Systemas (PACIS). 206*. Association for Information Systems AIS Electronic Library (AISeL). Retrieved 2018, from <https://aisel.aisnet.org/pacis2018/206/>
- Mohapatra, B., Parisa, V., & Banerjee, J. (2014, 5). The value of Big Data Analytics to the Business. *ISACA Journal*, 4.
- Nokkala, T., & Salmela, H. (2019). Data Governance in Digital Platforms. In A. f. (AISeL) (Ed.), *Twenty-fifth Americas Conference on Information Systems*. Cancún: Association for Information Systems AIS Electronic Library (AISeL). Retrieved from <https://aisel.aisnet.org/amcis2019/ebusiness/ebusiness/12/>
- Otto, B. (2011). A morphology of the organisation of data governance. In E. 2. Proceedings (Ed.), *European Conference on Information Systems (ECIS). 272*. ECIS 2011 Proceedings. Retrieved from <https://aisel.aisnet.org/ecis2011/272/>
- Otto, B. (2011, 8 6). Data Governance. *BISE Journal - Business & Information Systems Engineering*, 4. doi:10.1007/s11576-011-0275-1
- Otto, B. (2011, 8). Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. (A. J. (AISeL), Ed.) *Communications of the Association for Information Systems*, 29(3). doi:DOI: 10.17705/1CAIS.02903
- Piorún, D. (2016). Normas y Marcos relacionados con la implementación de esquemas de Gobierno de la Tecnología de la Información. Buenos Aires.
- Ramachandran, R. (2017, 6). Big Data Analytics and Its Transformational Effect on Global Business. *ISACA Journal*, 2.
- Reglamento General de Protección de Datos*. (2021, 5 27). Retrieved from <https://rgpd.es/SAP>. (2021, 12). Retrieved from <https://www.sap.com/latinamerica/products/data-profiling-steward.html>
- Saroka, R. H. (2002). *Sistemas de Información en la era digital*. Fundación OSDE.
- SAS*. (2021, 12). Retrieved from https://www.sas.com/es_ar/solutions/data-management.html
- Smallwood, R. (2014). *Information Governance. Concepts, Strategies and Best Practices*. Wiley.
- Talend*. (2021, 12). Retrieved from <https://help.talend.com/r/en-US/7.3/data-catalog-user-guide>
- The Open Group. (n.d.). The TOGAF® Standard, Version 9.2. Retrieved from <https://www.opengroup.org/togaf>

Weill, P., & Ross, J. (2004). IT Governance: How Top Performers Manage IT Decision Rights for Superior Results.

Anexo I

A continuación, se presentan los cuadrantes mágicos de Gartner® para los tres mercados que componen la oferta de herramientas para el Gobierno de Datos.

Soluciones de Calidad de Datos

Año 2019

Figure 1. Magic Quadrant for Data Quality Tools



Año 2020

Figure 1. Magic Quadrant for Data Quality Solutions



Soluciones de Integración de Datos

Año 2019



Año 2020



Soluciones de gestión de metadatos (data management solutions)

Año 2019

Figure 1. Magic Quadrant for Metadata Management Solutions



Source: Gartner (October 2019)

Año 2020

Figure 1. Magic Quadrant for Metadata Management Solutions



Source: Gartner (November 2020)