

**Universidad ORT Uruguay**  
**Facultad de Ingeniería**

# **Modelo de madurez para gobernanza de Big Data**

Entregado como requisito para la obtención del título de Licenciado en Sistemas

**Martín Armellino – 108814**

**Maurizio Paolino – 180080**

**Tutor: Helena Garbarino**

**2019**

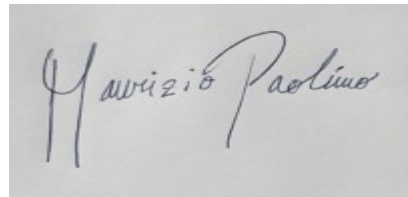
## Declaración de autoría

Nosotros, Martín Armellino y Maurizio Paolino , declaramos que el trabajo que se presenta en esa obra es de nuestra propia mano. Podemos asegurar que:

- La obra fue producida en su totalidad mientras realizábamos el proyecto de final de carrera;
- Cuando hemos consultado el trabajo publicado por otros, lo hemos atribuido con claridad;
- Cuando hemos citado obras de otros, hemos indicado las fuentes. Con excepción de estas citas, la obra es enteramente nuestra;
- En la obra, hemos acusado recibo de las ayudas recibidas;
- Cuando la obra se basa en trabajo realizado conjuntamente con otros, hemos explicado claramente qué fue contribuido por otros, y qué fue contribuido por nosotros;
- Ninguna parte de este trabajo ha sido publicada previamente a su entrega, excepto donde se han realizado las aclaraciones correspondientes.



Martín Armellino  
12/09/2019



Maurizio Paolino  
12/09/2019

## **Dedicatoria**

El presente trabajo está dedicado a nuestras familias, por haber dado apoyo a lo largo de nuestras carreras universitarias y a lo largo de la vida. A todas las personas especiales que nos acompañaron en esta etapa, aportando a nuestra formación tanto profesional y como de seres humanos.

## **Agradecimientos**

El presente trabajo de investigación lo dedicamos principalmente a nuestras familias, amigos y seres queridos, quienes nos han brindado su apoyo incondicional a lo largo de esta etapa, acompañando y compartiendo buenos y malos momentos y han sabido comprender nuestras obligaciones y nuestros humores.

A todas las autoridades y personal que forman a la Universidad ORT Uruguay, por confiar en nosotros, abrirnos las puertas y permitir realizar todo el proceso de investigación dentro de su establecimiento educativo.

A nuestra tutora Helena Garbarino por su dedicación y por todas las atenciones e información brindada a lo largo de este proyecto.

A todos ellos, muchas gracias.

## **Abstract**

Hoy en día las organizaciones disponen de enormes cantidades de información, las cuales en muchos casos, representan un aspecto esencial del negocio y donde su correcta utilización es un punto clave para el éxito y continuidad de las mismas. Por ende, es imperativo comprender y poseer la habilidad de poner en practica los conceptos como “gobernanza de datos”, “modelo de madurez” y “Big Data”.

El objetivo general del trabajo, es la construcción de un modelo de madurez para la gobernanza de Big Data.

Para lograr dicho objetivo nos encuadramos en el marco de gobernanza de Big Data propuesto por Sunil Soares y tomaremos como insumo el trabajo titulado “Modelo para la evaluación de buenas prácticas de gobernanza de Big Data” desarrollado por Darío Carlos y Mauro Nascimento, en el que se plantea una extensión y criterio de evaluación de la implementación de dicho marco.

Al evaluar distintas fuentes de literatura respecto a el tema en cuestión, investigamos diferentes modelos de madurez y estándares de gobernanza de datos y de Big Data. Esta investigación puso foco en que los modelos propuestos por la industria hasta el día de hoy no consideran adecuadamente, o no suficientemente, los aspectos relativos a las habilidades blandas y a los aspectos tecnológicos involucrados en los proyectos de Big Data.

A partir de estas bases se implementa dicho modelo de madurez que permita trazar un camino estratégico a las organizaciones en la adopción, evaluación y mejora de las prácticas de gobernanza de Big Data.

## **Palabras clave**

Big data, Cloud computing, Cloud data governance, Data, Data governance, Asset Management, Barriers, Data management, Data Quality, Decision Making, IT Governance, Big Data governance, Hard skills, Soft skills, Machine learning, Maturity Model, Risk Management, Systematic literature review.

# Índice

1. Introducción .....	17
1.1 Descripción del proyecto .....	17
1.2 Objetivos del proyecto .....	18
2. Estado del arte.....	20
2.1. Introducción.....	20
2.1.1. Gobierno corporativo .....	20
2.1.2. Gobierno de TI.....	20
2.1.3. Gobierno de datos.....	21
2.1.4. Gobierno de Big Data.....	26
2.2. Gobernanza de Big Data.....	31
2.2.1. Tipos de Big Data.....	32
2.2.2. Disciplinas.....	33
2.3. Estándar ISO 38500 .....	36
2.4. Desarrollo de instrumento para la evaluación de modelos de madurez para la gobernanza de Big Data que incorpore habilidades duras y blandas.....	43
2.5. CMMI.....	48
3. Investigación .....	61
3.1. Preguntas de investigación .....	61
3.2. Metodología de trabajo .....	62
3.3. Planificación de la revisión sistemática.....	63
3.4. Ejecución de la revisión sistemática.....	65
3.5. Análisis y resultados de la revisión sistemática .....	66
3.6. Respuestas a las preguntas de investigación planteadas.....	70
4. Desarrollo del modelo de madurez.....	76
4.1. Definición del alcance del modelo de madurez.....	76
4.2. Descripción del modelo de evaluación insumo .....	76

4.3. Análisis crítico del modelo de evaluación insumo.....	79
4.4. Modelo de madurez para la gobernanza de Big Data.....	80
4.4.1. Construcción del modelo de madurez.....	80
4.4.2. Descripción del modelo de madurez.....	85
4.4.3. Niveles de madurez.....	86
4.4.4. Disciplinas.....	87
4.4.5. Evaluación de Habilidades Hard.....	88
4.4.6. Evaluación de Habilidades Soft.....	99
4.4.7. Evaluación de resultados.....	100
4.4.8. Evolución dentro del modelo de madurez.....	102
4.4.9. Metodología de validación del modelo de madurez.....	104
5. Gestión de proyecto.....	110
5.1 Metodología de trabajo del equipo.....	110
5.2 Gestión de riesgos .....	110
5.3 Gestión del avance.....	113
6. Conclusiones y trabajos futuros.....	114
7. Referencias Bibliográficas .....	116
Anexo 1 – Modelo de evaluación insumo.....	120
Anexo 2 - Herramienta de gestión de tareas.....	130
Anexo 3- Evaluación.....	132
Anexo 4 – Carta de la empresa evaluada.....	136

## Índice de tablas

Tabla 1. Componentes de la gobernanza de big data.....	31
Tabla 2. Principios del estándar ISO/IEC38500.....	43
Tabla 3. Modelo MIG.....	45
Tabla 4. Comparación representaciones.....	52
Tabla 5. Áreas proceso con sus respectivos niveles de madurez.....	58
Tabla 6. Filtros de búsqueda.....	64
Tabla 7. Bases de datos para búsqueda.....	64
Tabla 8. Etapas de selección de artículos.....	65
Tabla 9. Resultados de la búsqueda con selecciones.....	67
Tabla 10. Autores con mayor actividad.....	67
Tabla 11. Artículos de la selección final.....	69
Tabla 12. Artículos de la selección final del proyecto insumo.....	70
Tabla 13. Ejemplo de registro de niveles Hard.....	102
Tabla 14. Propositiones del estudio del caso .....	105
Tabla 15. Roles participantes del estudio del caso.....	107
Tabla 16. Fuentes de datos del estudio de caso.....	107
Tabla 17. Resultados evaluación.....	109
Tabla 18. Escala de valores de impacto y probabilidad.....	111
Tabla 19. Identificación, calificación, plan de respuesta y plan de contingencia de riesgos..	113

## Índice de ilustraciones

Ilustración 1. DAMA Gobierno de datos.....	22
Ilustración 2. Decisiones del gobierno de datos.....	25
Ilustración 3. Componentes del gobierno de Big Data.....	29
Ilustración 4. Cubo de Sunil Soares.....	32
Ilustración 5. Tareas principales del Gobierno de TIC.....	39
Ilustración 6. Estructura por etapas.....	50
Ilustración 7. Representación continua.....	51
Ilustración 8. Diagramas de ambas representaciones.....	53
Ilustración 9. Niveles de madurez.....	54
Ilustración 10. Niveles de capacidad.....	55
Ilustración 11. Proceso de búsqueda sistemática.....	62
Ilustración 12. Diagrama de actividades para la construcción de un modelo.....	81
Ilustración 13. Niveles de madurez.....	83
Ilustración 14. Framework para gobernanza de Big Data.....	84
Ilustración 15. Gráfica de objetivos.....	103

## Glosario

**Adaptación de la evaluación de CMMI:** selección de opciones dentro del método de evaluación para usar en una instancia específica. La intención de la adaptación de la evaluación es ayudar a una organización a alinear la aplicación del método con sus objetivos comerciales.

**Análisis de riesgos:** Proceso de identificar los riesgos de seguridad, determinar su magnitud e identificar las áreas que necesitan salvaguardar.

**Aseguramiento:** Grado de confianza de que las necesidades de conformidad de la información con leyes, normas, controles y objetivos tanto internos como externos se satisfacen.

**Big data:** Término evolutivo que describe cualquier cantidad voluminosa de datos estructurados, semiestructurados y no estructurados que tienen el potencial de ser extraídos para obtener información.

**Buenas prácticas:** Describe un conjunto coherente de acciones que han rendido de buena manera en un determinado contexto y que se espera que, en contextos similares, rindan con similares resultados.

**Cliente:** Individuo u organización que es responsable de la aceptación del producto y de la autorización del pago a la organización del servicio / desarrollo

**Cloud computing:** Tecnología que permite acceso remoto a softwares, almacenamiento de archivos y procesamiento de datos por medio de Internet, siendo así, una alternativa a la ejecución en una computadora personal o servidor local.

**Componente del modelo CMMI:** cualquiera de los principales elementos arquitectónicos que componen un modelo CMMI. Algunos de los elementos principales de un modelo CMMI incluyen prácticas específicas, prácticas genéricas, objetivos específicos, objetivos genéricos, áreas de proceso, niveles de capacidad y niveles de madurez.

**Confidencialidad:** Propiedad de que la información no está disponible o revelada a individuos, entidades o procesos no autorizados

**Consistencia:** Grado de uniformidad, estandarización y libertad a partir de la contradicción entre los documentos o partes de un sistema o componente.

**Corporación:** Es una entidad legal artificial, con personería jurídica bajo la ley, que tiene una entidad legal separada a pesar de que se compone de una variedad de otras entidades jurídicas y personas reales.

**Datos maestros:** Son los datos básicos de los procesos de negocio, y la relación entre estos.

**Data management:** Se compone de todas las disciplinas relacionadas con gestionar los datos como un recurso valioso.

**Data quality:** Es la ciencia de mantener la información de las organizaciones completa, precisa, consistente, actualizada, única y, lo más importante, válida para el fin que se creó.

**Datos estructurados:** Tienen definido la longitud, el formato y el tamaño de sus datos. Se almacenan en formato tabla, hojas de cálculo o en bases de datos relacionales.

**Datos no estructurados:** Se caracterizan por no tener un formato específico. Se almacenan en múltiples formatos como documentos PDF o Word, correos electrónicos, ficheros multimedia de imagen, audio o video.

**Datos semiestructurados:** Los datos semiestructurados no presentan una estructura perfectamente definida como los estructurados, pero sí presentan una organización definida en sus metadatos donde describen los objetos y sus relaciones, y que en algunos casos están aceptados por convención.

**Director:** Miembro del más alto cuerpo de gobierno de una organización. Incluye propietarios, miembros del consejo, consejeros, socios, ejecutivos senior o similares y oficiales autorizados por el gobierno.

**Disponibilidad:** Propiedad de estar accesible y utilizable bajo demanda por una entidad autorizada.

**Eficacia:** Propiedad de un sistema o producto que representa cómo de bien proporciona seguridad en el contexto de su utilización operativa propuesta o actual.

**Estrategia:** Plan global de desarrollo de una organización que describe el uso eficaz de recursos en apoyo de la organización en sus actividades futuras. Comprende el establecimiento de los objetivos y la propuesta de iniciativas de actuación.

**Evaluación:** Evaluación por un equipo de profesionales entrenado para determinar el estado del proceso actual de las organizaciones, para determinar los problemas de alta prioridad relativos al proceso a los que se enfrenta una organización, y para obtener el soporte organizativo para la mejora del proceso.

**Evidencia:** Datos sobre los que se basa un juicio o conclusión sobre una demanda de aseguramiento. La evidencia puede consistir en la observación, los resultados de las pruebas, los resultados del análisis y las evaluaciones que proporcionan soporte para las demandas asociadas.

**Framework:** por su traducción al español se entiende que es un “Marco de trabajo” o un “Entorno de trabajo” pero es término que abarca un poco más, es un conjunto estandarizado de conceptos, prácticas y criterios para enfocar un tipo de problemática particular que sirve como referencia, para enfrentar y resolver nuevos problemas de índole similar.

**Gestión:** El sistema de procesos y controles necesarios para alcanzar los objetivos estratégicos establecidos por los órganos de gobierno de la organización. La gestión está sujeta a la dirección marcada por la política y seguimiento establecidos a través de la gobernanza corporativa

**Gestión de riesgos:** Proceso de evaluar y cuantificar el riesgo y establecer un nivel aceptable de riesgo para la organización.

**Gestión de TI:** Es el término que se refiere a la “gestión adecuada de recursos tecnológicos para alcanzar los objetivos corporativos al mínimo costo y con máxima rentabilidad”.

**Gobernanza corporativa:** El sistema por el cual se dirigen y controlan las organizaciones. (Adaptado de Cadbury 1992, y OCDE 1999)

**Gobernanza corporativa de TI:** El sistema por el cual se dirige y controla el uso, actual y futuro, de la TI. La Gobernanza corporativa de la TI implica evaluar y dirigir la utilización de la TI para dar soporte a la organización y la monitorización de ese uso para lograr la

consecución de los planes. Incluye la estrategia y políticas para la utilización de la TI en la organización.

**Gobierno de datos:** Consiste en la gestión de datos en pro de los objetivos empresariales, y representa una convergencia de la calidad de datos, gestión, políticas, gestión de procesos de negocio y gestión de riesgos, que comprende el tratamiento de los datos de una organización

**Grupo:** Colección de departamentos, directores e individuos que tienen la responsabilidad de un conjunto de tareas o actividades. El tamaño puede variar desde un individuo único a tiempo parcial hasta varios individuos asignados a tiempo parcial de diferentes departamentos, hasta varios individuos dedicados a tiempo completo.

**Habilidad blanda:** Serie de competencias personales que tienen como eje las relaciones interpersonales y el trabajo en conjunto y equipo. Son una sinergia de habilidades sociales, de habilidades de comunicación, de ser, de interacción con los demás, entre otras.

**Habilidad dura:** Serie de destrezas técnicas requeridas o adquiridas para desempeñar determinadas tareas o funciones.

**Integridad de datos:** Propiedad de que los datos no han sido alterados o destruidos de forma no autorizada.

**Interesados:** Quienes pueden afectar o son afectados por las actividades de una empresa.

**Machine learning:** Es un tipo de inteligencia artificial que proporciona a las computadoras la capacidad de aprender, sin ser programadas explícitamente.

**Medición de procesos:** El conjunto de definiciones, métodos y actividades utilizados para tomar medidas de un proceso y sus productos resultantes con el fin de caracterizar y comprender el proceso.

**Mejora de procesos:** Un programa de actividades diseñado para mejorar el rendimiento y la madurez de los procesos de la organización y los resultados de dicho programa.

**Modelo de CMMI:** Dado que el Marco CMMI puede generar diferentes modelos en función de las necesidades de la organización que lo utiliza, existen múltiples modelos CMMI. En consecuencia, la frase "MODELO CMMI" podría ser cualquiera de las muchas colecciones de

información. La frase "modelos CMMI" se refiere a uno, algunos o la colección completa de modelos posibles que se pueden generar a partir del Marco CMMI.

**Modelo de madurez:** Es un modelo de evaluación de los procesos de una organización.

**Nivel de capacidad:** logro de la mejora del proceso dentro de un área de proceso individual. Un nivel de capacidad se define por las prácticas específicas y genéricas apropiadas para un área de proceso.

**Nivel de madurez:** Grado de mejora del proceso en un conjunto predefinido de áreas de proceso en el que se alcanzan todos los objetivos dentro del conjunto.

**Organización:** Cualquier entidad corporativa, incluyendo asociaciones, clubs, socios, agencias gubernamentales, empresas públicas y privadas.

**Política de seguridad:** Conjunto de procesos que gestionan, protegen y distribuyen reglas, directivas y prácticas que gobiernan los activos, incluyendo información sensible, dentro de una organización y sus sistemas.

**Procedimiento:** Descripción escrita del curso de una acción a llevar a cabo para realizar una tarea dada.

**Proceso:** Secuencia de pasos realizados para un propósito dado.

**Riesgo:** Potencial de que una amenaza dada aprovechará las vulnerabilidades de un activo o grupo de activos para causar pérdida o daño a los activos.

**Risk management:** Es un enfoque estructurado para manejar la incertidumbre relativa a una amenaza, a través de una secuencia de actividades humanas que incluyen la identificación, el análisis y la evaluación de riesgo, para luego establecer las estrategias para su tratamiento, utilizando recursos gerenciales.

**Sistema:** Colección de componentes organizados para cumplir una función específica o conjunto de funciones. Un sistema puede incluir muchos productos. Un producto puede ser el sistema.

**Stakeholder:** Una parte interesada es un grupo o individuo que se ve afectado por el resultado de un proyecto o puede afectar las actividades o la producción del proyecto.

**TI:** Es el uso de cualquier computadora, almacenamiento, redes y otros dispositivos físicos, infraestructura y procesos para crear, procesar, almacenar, proteger e intercambiar todas las formas de datos electrónicos.

**Tolerancia a fallos:** Es un sistema que está capacitado para continuar su funcionamiento cuando algún componente del sistema falla, posiblemente a un nivel más reducido, lo que es mejor a que el sistema falle completamente.

**Validación:** Proceso de evaluar un sistema para determinar si satisface los requisitos especificados.

**Verificación:** Proceso de evaluar un sistema para determinar si los productos de trabajo de una fase dada de desarrollo satisfacen las condiciones impuestas al principio de esa fase.

# 1. Introducción

## 1.1 Descripción del proyecto

En la actualidad las organizaciones han encontrado un gran valor estratégico en la información que disponen. Este valor se convirtió en una necesidad e importancia nunca antes vista en el sector de gobierno de datos, sin importar el tamaño de la organización o en la industria en la que se encuentra. El interés por el buen manejo de datos ha causado en la gran mayoría de los casos la misma problemática: la falta de herramientas que faciliten el manejo y la gestión de datos.

La DAMA[1], nos dice que el gobierno de datos “es el ejercicio de la autoridad y el control (planificación, el seguimiento y la aplicación) a través de la gestión de los activos de datos. La función de Gobierno de Datos guía como se llevan a cabo todas las demás funciones de gestión de datos. Gobierno de Datos es de alto nivel, es la administración ejecutiva de los datos”

De esta definición desprendemos que son necesarias diversas actividades relacionadas a estructuras organizacionales, reglas, normas, derechos de decisión, responsabilidades y métodos de aplicación para manejar información.

Debido a la importancia de la información, la complejidad de los procesos de gobierno y las necesidades de las industrias, diferentes organizaciones han definido marcos de trabajo, modelos y conjuntos de buenas prácticas.

El presente proyecto se encuadra en el marco de gobernanza de Big Data propuesto por Sunil Soares [2] y toma como insumo el trabajo que extiende lo propuesto por Soares llamado “Modelo para la evaluación de buenas prácticas de gobernanza de Big Data” desarrollado por Darío Carlos y Mauro Nascimento [3], además de analizar otros modelos como lo son la ISO /IEC 38500[4] y la Integración del Modelo de Madurez de Capacidades, para poder encontrar patrones en común y evaluar tendencias que permitan determinar cuales son las prácticas aplicadas y como afectan el aumento de valor y de riesgo de las organizaciones.

A partir de estas bases, se busca la construcción de un modelo de madurez que permita trazar un camino estratégico a las organizaciones en la adopción, evaluación y mejora de las prácticas de gobernanza de Big Data.

A continuación, se detallan los objetivos que el trabajo se propone alcanzar. Luego de plantear los mismos se le facilita al lector un capítulo dedicado únicamente al estado del arte, el cual tiene la finalidad de introducir a los interesados con los conceptos y avances más importantes respecto al tema en desarrollo. Una vez que se manejan los conceptos, se plantea la investigación realizada, los objetivos y como se trabajó para lograrlos.

Posteriormente se presenta lo más sustancial del trabajo, que es el modelo de madurez de gobernanza de Big Data. Se detalla su desarrollo, su alcance, análisis y por último la validación del mismo. Finalizando, se podrá apreciar la gestión del proyecto y la conclusión que el equipo llegó luego de haber terminado el trabajo.

## **1.2 Objetivos del proyecto**

A continuación se enumeran los objetivos propuestos para este proyecto, estos objetivos sirven como plan de proyecto y guía para evaluar el avance en el mismo.

### 1. Investigación sobre el tema

1.1. Análisis de material base que tomamos como insumo para nuestro proyecto.

1.2. Búsqueda y análisis de información

1.2.1 Planificación de las búsquedas

1.2.2 Búsqueda y análisis de información en bases de datos de Timbó

1.2.3 Búsqueda y análisis de información en Google

### 2. Construcción del modelo de madurez para Big Data

2.1 Definición del alcance del modelo de madurez

2.2 Desarrollo del modelo de madurez

### 3. Validación de modelo de madurez

3.1 Definición de la metodología para validar

3.2 Validar el modelo de madurez

3.3 Analizar los resultados de la validación

3.4 Correcciones

## **2. Estado del arte**

### **2.1. Introducción**

#### **2.1.1. Gobierno corporativo**

A modo de inicio, se comienza con la definición del concepto más básico del cual proviene la investigación, este es el de gobernanza corporativa.

Analizando el termino por separado, se encuentra que la RAE define gobernanza como el “arte o manera de gobernar que se propone como objetivo el logro de un desarrollo económico, social e institucional duradero, promoviendo un sano equilibrio entre el Estado, la sociedad civil y el mercado de la economía. ”

De modo global la gobernanza es lo que resulta de organizar a las personas con la finalidad de alcanzar los objetivos de la comunidad, de los cuales destaca la protección del territorio, la seguridad de la población y su desarrollo integral.

Por otro lado, el término corporación según la RAE significa, “organización compuesta por personas que, como miembros de ella, la gobiernan” y que tiene un conjunto de derechos y obligaciones legales.

Se entiende, entonces, por el concepto de gobernanza corporativa una manera de gobierno de una organización, de personas que la gobiernan, que tiene un objetivo determinado, o tomando la definición que se encuentra en la norma ISO/IEC 38500 [4], la gobernanza o gobierno corporativo es "el sistema por el cual se dirigen y controlan las organizaciones".

#### **2.1.2. Gobierno de TI**

El gobierno de TI es uno de los tipos de disciplinas que se pueden hallar bajo gobierno corporativo. Esta disciplina tiene principal foco en tecnologías de información (TI), su desempeño y su gestión de riesgos.

Al investigar literatura que trata sobre gobierno de IT, se pueden encontrar distintas definiciones y análisis de este mismo concepto y sobre el impacto que tiene sobre las organizaciones tanto en lo operativo como en lo estratégico.

Al centrarse solo en definiciones, se toma el estándar ISO/IEC 38500[4], en el cual se determina como "el sistema por el cual el uso actual y futuro de TI es dirigido y controlado. La gobernanza de TI implica evaluar y dirigir el uso de las TI para dar soporte a la organización y monitorizar el uso para lograr los planes. Incluye la estrategia y políticas para la utilización de las TI en la organización."

La ITGI (IT Governance Institute) la define como "el liderazgo, las estructuras organizacionales, y los procesos que aseguran que las TI utilizadas por la organización soporten y extiendan estrategias y objetivos de la organización"

Según Mark Weill y Jean Ross [5], la Gobernanza de TI "es el marco que permite definir responsabilidades y tomar decisiones correctas para impulsar los comportamientos deseables en el uso de la TI en las organizaciones".

Ambas definiciones hacen énfasis en la importancia que posee la integración de la TI en la estrategia de la empresa, la cual fortalece su desempeño, aumentando valor, minimizando y controlando los riesgos implícitos en las inversiones de TI.

### **2.1.3. Gobierno de datos**

La Data Governance Institute (DGI), fundada en el 2003, al día de hoy posee extensa cantidad de guías y mejores prácticas sobre el tema en la industria. Esta define gobierno de datos como "un sistema de derechos y responsabilidades de decisión para procesos relacionados con la información, ejecutado de acuerdo con modelos acordados que describen quién puede tomar qué acciones, con qué información y cuándo, bajo qué circunstancias, usar qué métodos".

La DAMA[1] dice que "es el ejercicio de la autoridad y el control (planificación, el seguimiento y la aplicación) a través de la gestión de los activos de datos. La función de Gobierno de Datos guía cómo se llevan a cabo todas las demás funciones de gestión de datos. Gobierno de Datos es de alto nivel, la administración ejecutiva de los datos"

En ambas definiciones se ve que el gobierno de datos tiene un enfoque en la gestión de los activos de datos, diferenciándose notoriamente del concepto de gobierno de TI, el cual se centra en la toma de decisiones de TI, las inversiones, sus aplicaciones y proyectos.

Para ampliar sobre el concepto, se deja de lado la DGI ya que esta no resulta clara, y se toma como referencia el libro "The DAMA Guide to The Data Management Body of Knowledge",

[1] el cual plantea los conceptos con mayor claridad y proporciona mejor visualización al incluir ilustraciones en conjunto a definiciones.

En la siguiente Ilustración 1, se observan los distintos elementos que logran definir al gobierno de datos, respectivamente con sus actividades, tanto de planificación como de control. Muchas de estas actividades forman parte de los principios de la gobernanza corporativa de TI como del gobierno de Big Data, el cual se detallará más adelante en este mismo documento.

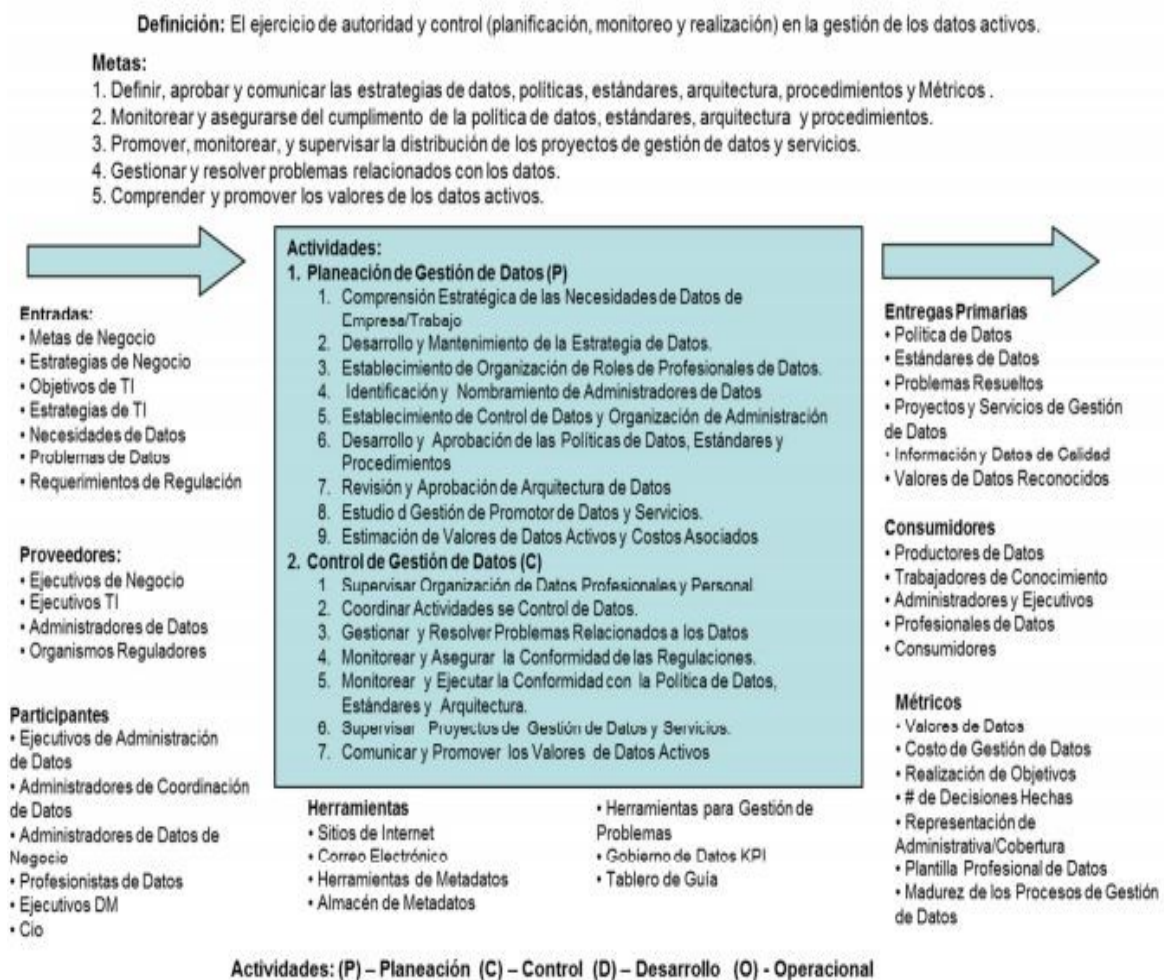


Ilustración 1. DAMA Gobierno de datos [1]

Las actividades de gobierno de datos y de gobierno de Big Data están directamente relacionadas, ambas tratan de establecer ciertos lineamientos y controles sobre los datos que obtiene o genera la organización.

Las principales actividades del gobierno de datos mencionadas en [1] se detallan a continuación:

- Estrategia de datos: programa de gestión de datos para mantener y mejorar la calidad de los datos, la integridad, la seguridad y el acceso. Esta, también puede incluir planes de negocio a implementar para la propia ventaja competitiva y el apoyo de las metas de la empresa. La estrategia de datos debe provenir de una comprensión de las necesidades de los datos inherentes a las estrategias de negocio. Cabe mencionar, que la arquitectura de datos no es lo mismo que la estrategia de estos, aunque la estrategia puede influir en la arquitectura que, a su vez, apoya la estrategia guiando otras decisiones.
- Política de datos: “son breves declaraciones de la administración y las normas fundamentales que rigen la creación, adquisición, la integridad, la seguridad, la calidad y el uso de datos e información” [1]. Las políticas de datos varían entre organizaciones. Estas dictan qué hacer y qué no hacer, mientras que las normas y procedimientos permiten describir el “cómo” hacer algo. Una vez que son establecidas, deben ser comunicadas a toda la organización, y luego de comenzar a aplicarlas se deben monitorear para seguir su aplicación y para permitir la reevaluación. Pueden abarcar temas como el desarrollo y uso de la arquitectura de datos, expectativas de calidad, seguridad de los datos, recuperación y retención, acceso, entre otros.
- Arquitectura de datos: esta actividad se ocupa de que el modelo de datos de la empresa sea desarrollado, mantenido y monitoreado conjuntamente por los arquitectos de datos y administradores de datos.
- Datos estandarizados y procedimientos: aquí es donde se definen estándares de nomenclatura, normas de especificación de requisitos, normas de diseño de base de datos, estándares de arquitectura, entre otros. Los procedimientos son “los métodos documentados, técnicas y los pasos seguidos para llevar a cabo una actividad o tarea específica.” Esta estandarización y uso de procedimientos resulta ser de gran ayuda cuando se considera que en big data hay un alto volumen de datos y de gran variedad de fuentes y tipos.

- Cumplimiento normativo: actividad relacionada con el cumplimiento de las regulaciones de la organización tanto gubernamentales como del sector industrial al que pertenezca. En muchos casos estas normas van a determinar cómo se realizará la gestión de datos. Esto aplica también a la gobernanza de big data ya que va a depender de las normativas existentes.
- Gestión de problemas: actividad a cargo de identificar, gestionar y resolver diferentes tipos de asuntos relacionados con datos. Dentro de los asuntos a tratar, están los problemas de nomenclatura, calidad, seguridad, privacidad, confidencialidad y otras variedades de conflictos (políticos, estándares, procedimientos, etc).
- Proyectos de gestión de datos: aquí se busca aplicar o improvisar la gestión de datos a través de la implementación de proyectos diseñados o mantenidos en esta actividad. Algunos de estos proyectos pueden incluir temas relacionados con la gestión de arquitectura de datos, almacenamiento, datos maestros, metadatos, calidad de datos, entre otros.
- Servicios de gestión de datos: es importante para la organización que exista una formalización y definición de los servicios de gestión de datos que esta va a brindar, manteniendo siempre presente las necesidades del negocio. Esta formalización otorga la oportunidad de poder dimensionar el gasto o inversión que pueda tener cada uno de estos servicios. Por ejemplo, se podría usar para justificar un determinado gasto en una implementación.
- Valoración de activos: los datos y la información son realmente activos que tienen valor para el negocio, ya sea tangible o intangible, y por consiguiente las organizaciones necesitan ponderar de su información disponible. Las organizaciones aplican diferentes enfoques para determinar el valor de sus datos o el beneficio que estos traen. Uno de los usos más en práctica hoy en día es el del análisis de comportamiento e intereses en compras de clientes.
- Comunicación y promoción: esta última actividad debe realizarse constantemente con el fin de educar y promover la importancia y valor de los datos, que son activos de la empresa y son fundamentales para el negocio. El objetivo principal de esta actividad es que los usuarios comprendan las políticas de datos y el compromiso que la

organización tiene con estos. Políticas de calidad, seguridad, protección, acceso, entre otras deben ser comunicadas al personal con el fin de que sean conscientes del programa de gestión de datos y a su vez entiendan los objetivos y expectativas que esta tiene.

- Para que un gobierno de datos sea eficiente, se tiene que trabajar a través de ciertos límites organizativos y de sistema. La toma de decisiones compartida es un elemento característico del gobierno de datos, como se muestra en la Ilustración 2.



Ilustración 2. Decisiones del gobierno de datos [1]

Un programa de gobierno de datos es único en si mismo, siempre tomando en cuenta asuntos de la organización y de su cultura, como también los retos y oportunidades de gestión de datos de la misma. Para lograr un gobierno de calidad, se debe implementar con un programa en curso y un proceso de mejora continua.

Gobierno de datos no es lo mismo que el gobierno de IT. El gobierno de IT toma las decisiones sobre las inversiones en IT, la cartera de aplicaciones de IT y la cartera de proyectos de IT. Este se alinea con las estrategias de TI y las inversiones y estrategias empresariales. Por otro lado, gobierno de datos se encuentra en el corazón de la gestión de los activos de los datos, parte de los activos intangibles que forman al activo TI.

Al día de hoy existen múltiples marcos de trabajo que buscan guiar y establecer una estandarización de los procesos en las organizaciones. La mayoría de estos marcos hacen foco, principalmente, en métodos, procesos y estructuras definidas como se detalla más adelante tomando como ejemplo el modelo CMMI. Una excepción es el estándar ISO/IEC 38500 [4] que incluye como un nuevo factor a considerar el comportamiento humano, o sea habilidades blandas, y el contexto en el que se encuentra el organismo. Este estándar logra diferenciarse del resto al tener presente que una organización se encuentra conformada de personas, y todo proceso y estructuras deben planearse siempre teniendo presente los individuos y crearse de tal manera que estos los puedan aplicar con comodidad, pero principalmente naturalidad.

#### **2.1.4. Gobierno de Big Data**

##### **2.1.4.1. Big Data**

Para explicar el concepto de gobierno de Big Data se busca apoyo en el trabajo de uno de los autores más influyentes de este proyecto, “Big Data Governance” por Sunil Soares [2]. Antes de detallar el concepto completo, se comienza con Big Data. Soares lo define como “los datos que contienen una mayor variedad que llegan en volúmenes crecientes y con una velocidad cada vez mayor”. Esto se conoce como las tres “Vs”. Además, Soares [2] agrega una cuarta “V” para valor.

Las “Vs” en cuestión, son:

- **Volumen:** Son los datos en reposo. Las soluciones de Big Data son un aglomerado de datos y las organizaciones acumulan grandes cantidades de los mismos.
- **Velocidad:** Son los datos en movimiento. Los mismos se deben analizar rápidamente para maximizar el rendimiento que pueda tener para la organización.
- **Variedad:** Son los distintos tipos de datos. Big data debe incluir datos estructurados, así como también no estructurados como lo son texto, audio, vídeo, archivos de logs, etc.
- **Valor:** Este concepto habla directamente del rendimiento. Ese rendimiento que pretenden obtener las empresas que implementan y aplican sistemas de Big Data.

#### **2.1.4.2. Tipos de Big Data**

Según Soares, la gobernanza de Big Data necesita un mayor enfoque en los datos en sí. Debido a esto es que se clasifica Big Data en las siguientes categorías:

- Web y de social media
- Machine-to-machine
- Big transaction data
- Biométricos
- Generado por el humano

A continuación, en el capítulo 2.2, se expande sobre los tipos de Big Data que propone Soares con mayor detalle.

#### **2.1.4.3. Industrias de Big Data**

La implementación de la solución de Big Data va a estar determinada por una variedad distinta de factores. Tanto el tipo de datos como la recolección de los mismos no solo va a variar dependiendo de la industria en la que se encuentre ubicada la empresa sino también del tipo de organización, cantidad de empleados, número de clientes, distribución geográfica y legislación aplicable, para mencionar algunos. Todos estos factores influyen en las decisiones de la gobernanza de datos. Las principales industrias a destacar:

- Gestión de Riesgos
- Seguros
- Salud
- Marketing
- Seguridad de información
- Recursos humanos
- Tecnologías de información

- Atención al cliente

- Ventas

A continuación, en el capítulo 2.2, se expande sobre las industrias de Big Data que propone Soares con mayor detalle.

#### **2.1.4.4. Fuentes de Big Data**

Big Data consiste en concentrar el mayor volumen de información y seleccionar la más relevante. Sin embargo, existen muchos retos y desafíos a la hora de recolectar, almacenar y gestionar los datos, así como mantenerlos actualizados y poder compartirlos. Las organizaciones u organizaciones deben tener presente el impacto negativo de estos desafíos, y es por esto que deben poder asegurar que la calidad de los datos sea de calidad.

Algunas de las fuentes de Big Data más comunes:

- Redes sociales

- Señales GPS

- Etiquetas RFID

- Vídeos en línea

- Datos financieros del mercado

- Sensores de flujo de tráfico

- Imágenes satelitales

- Transmisiones de audio

- Transacciones bancarias

- Imágenes digitales

- Páginas web y su contenido

A continuación, en el capítulo 2.2, se expande sobre las fuentes Big Data que propone Soares con mayor detalle.

#### 2.1.4.5. Componentes del gobierno de Big Data

Mientras que el volumen y la complejidad de los datos continua creciendo, y las organizaciones continúan adquiriendo big data, es imprescindible para estas que posean un buen gobierno de big data. Uno de los grandes problemas hoy presentes es la dificultad que existe en las organizaciones para ser capaces de distinguir los datos de calidad con los datos de insuficiente calidad.

Para priorizar la calidad, el gobierno de big data debe formar parte del programa de gobierno de datos, pero se deben realizar ciertas acciones para adaptarlo a estas soluciones. Se deben modificar políticas puntuales relacionadas con la optimización, privacidad y monetización de la Big Data [2], de modo de alinearse con los objetivos de las distintas partes. El resultado de una implementación de Big Data muchas veces es consecuencia de si se aplicó de manera exitosa o no el plan de gobierno de datos.

Los componentes principales a tener en cuenta en el gobierno de Big Data son los que se muestran en la siguiente Ilustración 3.

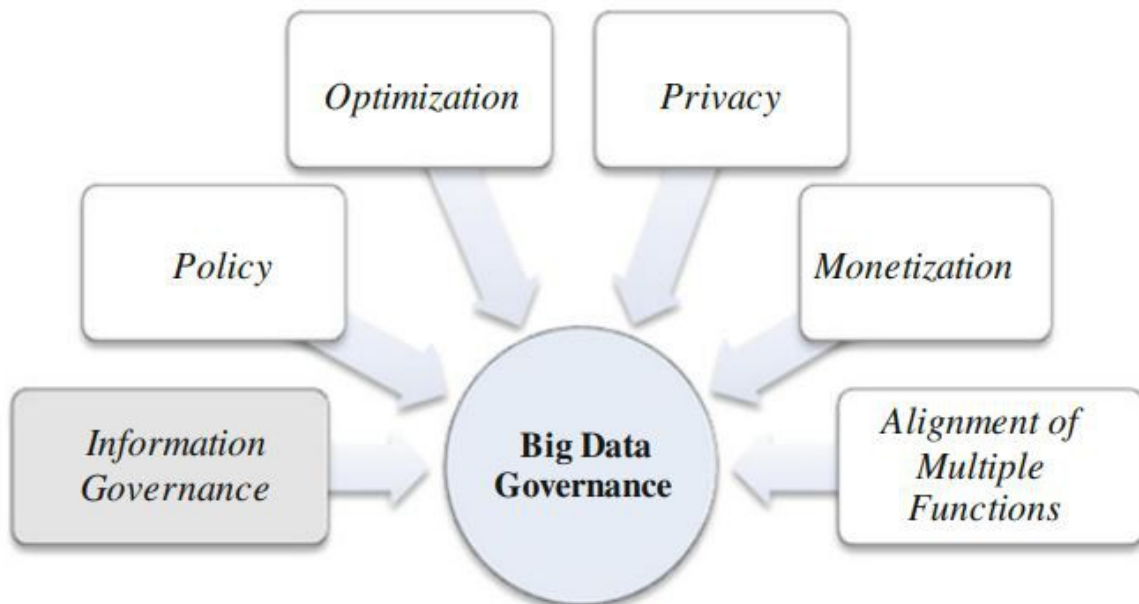


Ilustración 3. Componentes del gobierno de Big Data [2].

A continuación, en Tabla 1, se detallan los componentes clave de la gobernanza de big data.

Componentes	Descripción
Gobernanza de la información	Las organizaciones deben ampliar el alcance actual de la gobernanza de la información para incluir Big Data, usar usuarios especializados en Big Data, así como apoyar la implementación de Big Data en otras disciplinas asociadas, como metadatos, privacidad y datos maestros.
Políticas	Una política de gobernabilidad de datos debe obedecer los requisitos legales y reglamentarios de la organización y de su entorno.
Optimización	<p>Las organizaciones necesitan optimizar y mejorar la calidad de su Big Data en las siguientes áreas:</p> <ul style="list-style-type: none"> <li>• Metadatos: para generar información sobre los inventarios de Big Data</li> <li>• Gestión de la calidad de los datos: para que la Big Data tenga datos “limpios”, al igual que las compañías realizan un mantenimiento preventivo en los conjuntos de datos físicos</li> <li>• Gestión del ciclo de vida de la información: para archivar y retirar grandes volúmenes de datos cuando ya no tenga sentido retenerlos</li> </ul>
Confidencialidad y privacidad	Las organizaciones también deben crear y seguir las políticas y procedimientos apropiados para evitar el uso indebido de Big

Componentes	Descripción
	Data, teniendo en cuenta los riesgos legales y regulatorios al manejar distintas formas de información de identificación personal.
Monetización	La monetización es el proceso de convertir un activo, como los datos, en dinero, vendiéndolo a terceros o usándolo para desarrollar nuevos servicios.
Alineación de múltiples funciones	La gobernanza de Big Data necesita hacer un equilibrio entre los objetivos de las partes en competencia. El gobierno de Big Data las debe reunir para determinar si los ingresos potenciales de los nuevos servicios sopesan los riesgos de reputación y reglamentarios asociados.

Tabla 1: Componentes de la gobernanza de big data

## 2.2. Gobernanza de Big Data

Desde los comienzos de la investigación para la realización del artículo en cuestión, se utiliza como referente el libro “Big Data Governance: An Emerging Imperative” del autor Sunil Soares [2].

En este mismo, el autor toma la gobernanza de big data y la divide en tres dimensiones. En base a estas, presenta un cubo con sus respectivas divisiones obtenidas de la intersección de los diferentes tipos de industria y de big data y también detalla un conjunto de disciplinas de la gobernanza de información o de datos, las cuales señalan diferentes características que se deben tener en cuenta en una organización para la implementación de un buen gobierno de Big Data.

## Big Data Governance Framework

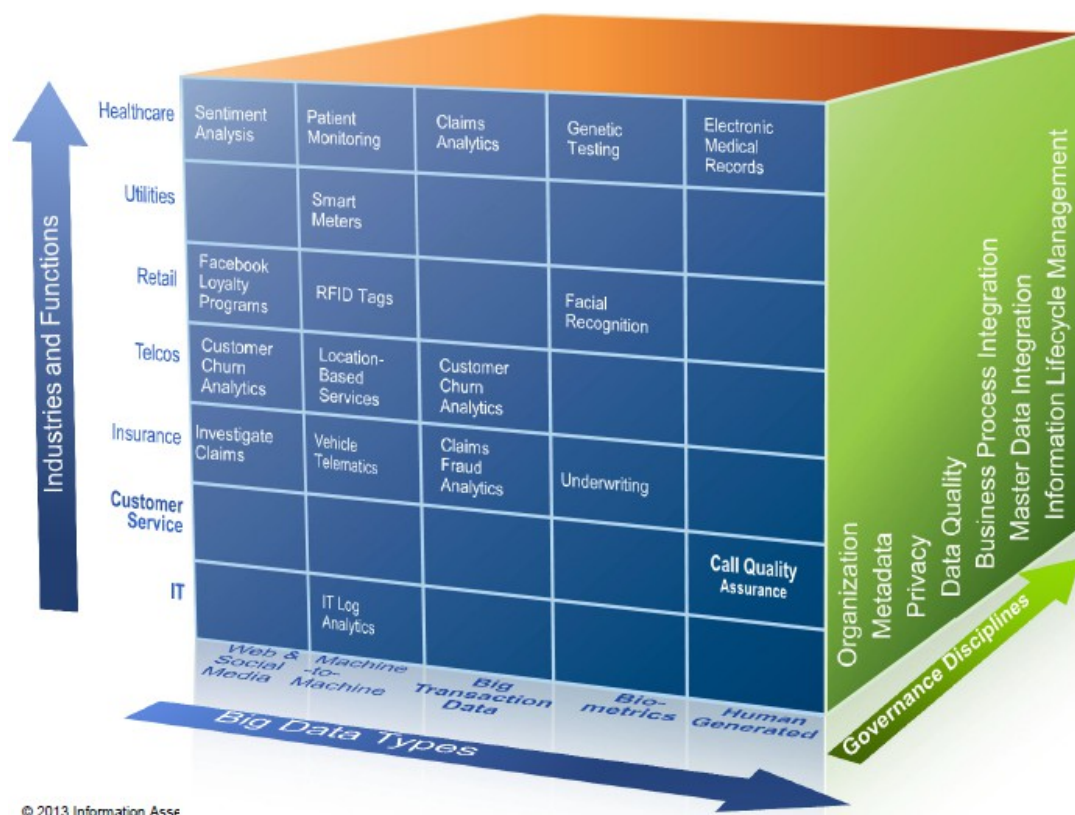


Ilustración 4. Cubo de Sunil Soares [2]

### 2.2.1. Tipos de Big Data

Los tipos de big data que destaca Soares son:

- Web y redes sociales: incluye el estudio de los clicks en redes como Facebook, twitter, LinkedIn y blogs. En lo que se distingue la importancia que tiene incluir estos datos con los datos maestros, esto podría provocar la creación de procesos de negocio con un mejor funcionamiento, así como de programas de lealtad para el cliente. El programa de gobernanza de Big Data debe establecer políticas sobre el uso aceptable de los datos generados por las redes sociales, especialmente porque las regulaciones y precedentes están continuamente en estado de evolución.
- Machine-to-machine data (M2M): hace referencia a las tecnologías que permiten sistemas alámbricos e inalámbricos se comuniquen con otros dispositivos. Como por ejemplo, la recolección de datos que generan sensores.

- Transacciones de big data: este tipo de Big Data Incluye reclamos en el cuidado de la salud, telecomunicaciones, CDR (call detail records) y utilidades. El autor destaca que este tipo de Big Data incrementa en datos semi-estructurados y estructurados, y que de aquí mismo pueden ser obtenidos datos muy valiosos para lo que es metadata, calidad de datos, privacidad y el ciclo de vida de la información.
- Biométrico: engloba la identificación automática de individuos basado en sus características o rasgos anatómicos o de comportamiento. Esta data es creada directamente de las personas como son sus huellas, iris, ADN, retina, rostro, líneas de las manos, voz e inclusive olor corporal entre otros. Hoy en día es muy común que los dispositivos móviles puedan ser desbloqueados con huella digital o reconocimiento facial, y que esta misma data biométrica sea utilizada por apps.
- Datos generados por los humanos: el ser humano es un increíble generador de datos de por sí, notas de voz, emails, documentos, registros electrónicos médicos, etc. Toda esta es información puede llegar a ser muy valiosa y sensible.

### **2.2.2. Disciplinas**

Continuando con el cubo planteado por Soares, en la siguiente cara a detallar se encuentran las disciplinas.

#### **2.2.2.1. Organización**

La organización en su conjunto debe estar integrada con los procesos y el gobierno de la información, la misma debe considerar adoptar todo lo concerniente con la Big Data a su marco general, incluidos los estatutos, la estructura de la organización, roles y responsabilidades. Debe tener personas enfocadas en el gobierno de datos y debería incluir nuevos miembros y roles que puedan proporcionar una perspectiva única sobre la Big Data [2]. Con la adición de la Big Data en la cultura de la organización se pueden llegar a agregar responsabilidades adicionales a los puestos de trabajo existentes. Si se desea obtener un rédito de la implementación de la misma es necesario el apoyo de todas las partes, todos los sectores, pero fundamentalmente de los directivos y altos mandos.

#### **2.2.2.2. Metadata**

La metadata, que se define como los “datos acerca de los datos”. Debe ser integrada a la solución de Big Data de la organización. Esto implica que se incluyan los términos de Big Data en el glosario y en las definiciones de negocios como, por ejemplo, importar metadatos técnicos al repositorio de metadatos, gestionando su linaje y el análisis del impacto dentro del entorno de Big Data [2].

#### **2.2.2.3. Privacidad**

El concepto de privacidad data del 1890[6], y se puede resumir en “el derecho a quedarse solo”. Esta idea de privacidad, fue aplicada a través del mundo por medio de regulaciones y legislaciones que a lo largo de los años lograron formalizarse y expandirse. La gobernanza de big data necesita identificar data sensible y establecer políticas de acuerdo a su uso responsable. Las políticas a su vez, deben tener en cuenta las distintas regulaciones, que pueden variar según el tipo de Big Data, la industria y el país. Dadas las distintas aristas que puede tener este tema, el programa de gobierno de Big Data necesita establecer pautas con respecto al uso aceptable de los medios sociales y a los datos de geolocalización, en el caso de que corresponda [2].

#### **2.2.2.4. Calidad de datos**

La gestión de la calidad de los datos es una disciplina que incluye los métodos para medir, mejorar y certificar la calidad e integridad de los datos de una organización. Debido a su gran volumen, velocidad y variedad, la calidad de la Big Data debe manejarse de forma diferente que los tipos de datos tradicionales. Por ejemplo, la calidad de Big Data podría necesitar ser manejada en tiempo real y resolver problemas relacionados con datos semi-estructurados y no estructurados. La Big data deber ser "lo suficientemente buena" para que la mala calidad de los datos no interfiera en el análisis que se pueda llegar a realizar de modo que esa información pueda ser de ayuda para el negocio [2].

#### **2.2.2.5. Integración de proceso de negocio**

Para que se realice una buena integración con los procesos de negocio se debe identificar aquellos procesos claves que utilizan grandes volúmenes de datos. Se deben definir políticas que respaldan al gobierno o a los administradores de Big Data, tales como políticas de retención, tipos de datos a guardar, prioridad, etc. Generalmente es muy costoso almacenar

grandes cantidades de datos y es sumamente necesario saber cuáles son de importancia para el negocio. Estos datos puede que se guarden por un tiempo determinado o no tengan una fecha de expiración determinada [2].

#### **2.2.2.6. Integración de datos maestros**

Se deben establecer políticas con respecto a la integración de Big Data en el entorno de administración de datos maestros. Se deben definir e implementar políticas para que su uso sea aceptable y herramientas de administración adecuadas para determinar la consistencia de los mismos [2].

#### **2.2.2.7. Gestión del ciclo de vida de la información**

Debido al aumento y a los grandes volúmenes de datos que se manejan, las organizaciones tienen el desafío de comprender los requisitos regulatorios y comerciales que determinan qué datos retener, qué datos archivar y qué datos eliminar. Esto se debe respetar tanto para servidores de archivo como para cualquier otro repositorio de datos que se utilice. Las soluciones de Big Data no están exentas de esta realidad y también deben regularse bajo las mismas definiciones y necesidades comerciales. Se debe aprovechar, en los casos que pueda aplicar, políticas de compresión y archivado, así como también herramientas y mejores prácticas para reducir los costos y mejorar el rendimiento de las aplicaciones [2].

#### **2.2.3. Industrias**

Como fue discutido, la gobernanza varía entre los distintos sectores de la industria, y no es posible que dos organizaciones que se encuentran en distintos sectores tengan éxito aplicando la misma estrategia de gobernanza. Soares siendo consciente de esta problemática, identifica las siguientes industrias:

- **Salud:** donde se destaca el análisis de los sentimientos, monitoreo de pacientes, reclamos analíticos, pruebas genéticas y registros electrónicos médicos
- **Industria de las utilidades:** donde lo más destacado son los contadores inteligentes
- **Retail** (ventas al por menor): aquí destacan los programas de lealtad de Facebook, tags de RFID y el reconocimiento facial

- **Telecomunicaciones:** donde destaca los servicios basados en la localización de los clientes, seguros donde se encuentran investigación de reclamos, investigación de fraudes o aseguramiento
- **Atención al cliente:** donde lo único destacable es la garantía de la calidad de los llamados
- **TI:** donde lo importante aquí es el análisis de los logs.

### 2.3. Estándar ISO 38500

También conocido por su nombre completo ISO/IEC 38500:2015 [4], se trata de uno de los estándares el que no solo analizamos sino que también lo usamos como referencia para la creación del modelo, ya que se trata de una guía aceptada internacionalmente para el gobierno de TI.

El objetivo de este estándar es proveer principios, definiciones y un modelo para que las organizaciones puedan usar a la hora de evaluar, dirigir, y monitorear el uso de TI. De esta manera, proporciona orientación a quienes asesoran, informan o ayudan a los órganos de gobierno. Tales como:

- Gerentes ejecutivos
- Miembros de grupos que monitorean los recursos dentro de la organización
- Proveedores de servicios internos y externos (incluidos consultores)
- Auditores
- Especialistas comerciales o técnicos externos, como especialistas legales o contables, asociaciones minoristas o industriales, u organismos profesionales

En la industria se lo conoce como un estándar de alto nivel; una norma de asesoramiento basada en principios. Además de proporcionar orientación a la dirección de las organizaciones, alienta a estas a utilizar estándares apropiados para apuntalar la gobernanza de TI.

En esencia, todo lo que esta norma propone podría resumirse en tres propósitos fundamentales [7]:

- Asegurar que, si la norma es seguida de manera adecuada, las partes implicadas (directivos, consultores, ingenieros, proveedores de hardware, auditores, etc.), puedan confiar en el gobierno corporativo de TIC.
- Informar y orientar a los directores que controlan el uso de las TIC en su organización.
- Proporcionar una base para la evaluación objetiva por parte de la alta dirección en el gobierno de las TIC.

### **2.3.1. Alcance, Aplicación y Objetivos**

Toda organización que hoy en día utiliza tecnologías de la información, podría aplicar la norma en su gobierno de los procesos de gestión de las TICs para facilitar bases para la evaluación objetiva del gobierno de TIC [7].

Dentro de los beneficios de un buen gobierno de TIC estaría la conformidad de la organización con:

- Los estándares de seguridad
- Legislación de privacidad
- Legislación sobre el spam
- Legislación sobre prácticas comerciales
- Derechos de propiedad intelectual, incluyendo acuerdos de licencia de software
- Regulación medioambiental
- Normativa de seguridad y salud laboral
- Legislación sobre accesibilidad
- Estándares de responsabilidad social

También la búsqueda de un buen rendimiento de las TIC mediante:

- Apropiaada implementación y operación de los activos de TIC
- Clarificación de las responsabilidades y rendición de cuentas en lograr los objetivos de la organización
- Continuidad y sostenibilidad del negocio
- Alineamiento de las TICs con las necesidades del negocio
- Asignación eficiente de los recursos
- Innovación en servicios, mercados y negocios
- Buenas prácticas en las relaciones con los interesados (stakeholders)
- Reducción de costes
- Materialización efectiva de los beneficios esperados de cada inversión en TIC

### **2.3.2. Principios**

La norma define seis principios de un buen gobierno corporativo de TIC:

Responsabilidad: Individuos y grupos dentro de la organización deben comprender y aceptar sus responsabilidades en la oferta y demanda de TI. Aquellos con la responsabilidad sobre una acción también tienen autoridad para realizar la misma.

Estrategia: La estrategia de negocio de la organización debe tener en cuenta las capacidades actuales y futuras de las TI. Los planes estratégicos de TI satisfacen las necesidades actuales y previstas derivadas de la estrategia de negocio.

Adquisición: Las adquisiciones de TI se realizan por motivos válidos, basándose en un análisis apropiado y continuo, con decisiones claras y transparentes. Existe un balance entre beneficios, oportunidades, costos y riesgos, tanto a corto como a largo plazo.

Rendimiento: La TI es apta para dar soporte a la organización, proporcionando los servicios de calidad necesarios para cumplir con las necesidades actuales y futuras.

Conformidad: La función de TI cumple todas las legislaciones y normas aplicables. Las políticas y prácticas al respecto están claramente definidas, implementadas y exigidas.

Factor humano: Las políticas de TI, prácticas y decisiones demuestran respeto por la conducta humana, incluyendo las necesidades actuales y emergentes de toda la gente involucrada.

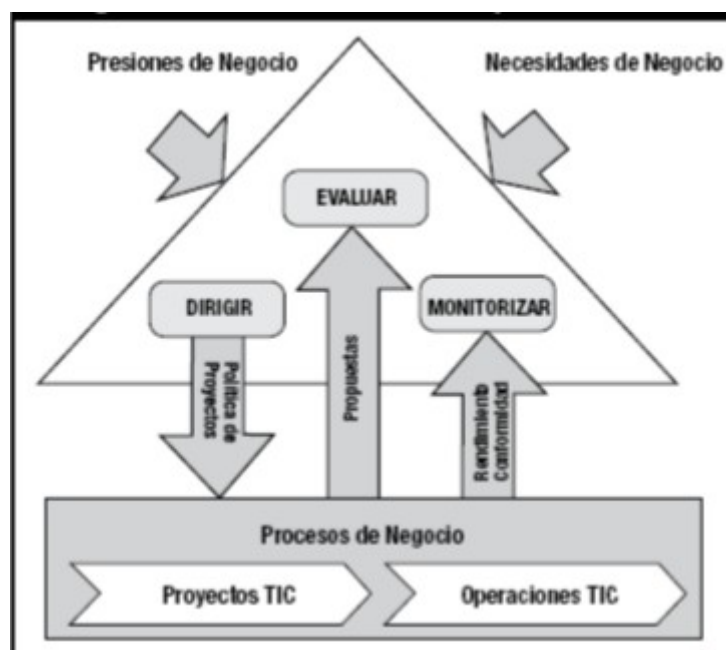
### 2.3.3. Modelo

La dirección ha de gobernar las TIC mediante tres tareas principales:

Evaluar: Examinar y juzgar el uso actual y futuro de las TIC, incluyendo estrategias, propuestas y acuerdos de aprovisionamiento.

Dirigir: Dirigir la preparación y ejecución de los planes y políticas, asignando las responsabilidades al efecto. Asegurar la correcta transición de los proyectos a la producción, considerando los impactos en la operación, el negocio y la infraestructura. Impulsar una cultura de buen gobierno de TIC en la organización.

Monitorizar: Mediante sistemas de medición, vigilar el rendimiento de la TIC, asegurando que se ajusta a lo planificado.



## Ilustración 5. Tareas principales del Gobierno de TIC

### **2.3.4. Orientaciones prácticas**

Para cada uno de los principios, la norma proporciona una breve guía u orientación sobre como evaluar, dirigir y monitorizar la función de TIC .Tabla 2.

Son orientaciones muy generales que no incluyen mecanismos, técnicas o herramientas concretas a utilizar [7].

<b>Principios</b>	<b>Dirigir</b>	<b>Monitorizar</b>	<b>Evaluar</b>
Responsabilidad	Planes con responsabilidad asignada	Mecanismos establecidos	Asignación de responsabilidades
	Recibir información y rendir cuentas	Asignación de responsabilidades	Competencias de responsabilidades
		Desempeño responsable	
Estrategia	Creación y uso de planes de política	Progreso propuestas aprobadas	Desarrollo de TIC y procesos de negocio
	Asegurarse beneficios TI en el negocio	Alcanzar objetivos en plazo establecidos	Evaluar actividades de TIC y alineamiento
	Alentar propuestas innovadoras	Utilizar recursos asignados	Mejores prácticas
		Uso de TIC alcanzando beneficios esperados	Satisfacción interesados
			Valoración y evaluación de riesgos
Adquisición	Activos de TI adquieren manera apropiada	Inversiones y capacidades requeridas	Alternativas propuestas
	Documentos capacidad requerida	Entendimiento interno/externo necesidades del negocio	Propuestas aprobadas
	Acuerdos de provisión respalden necesidades del negocio		Análisis de riesgo/valor
			Inversiones

<b>Principios</b>	<b>Dirigir</b>	<b>Monitorizar</b>	<b>Evaluar</b>
Rendimiento	Asignación de recursos suficientes	Grado TIC sustenta negocio	TIC sustenta procesos de negocio dimensionado y capacidad
	Asignar prioridades y restricciones	Recursos e inversiones priorizando nec. neg.	Riesgos: continuidad de operaciones
	Satisfacer nec. neg.	Políticas de precisión de datos	Decisiones de uso de TIC para apoyo al negocio
	Datos correctos, actualizados, protegidos	Políticas de uso eficiente de TIC	Decisiones de uso de TIC para apoyo al negocio
			Eficacia y desempeño de gobierno de TIC
Cumplimiento	TI cumple obligaciones, normas y directrices	Cumplimiento y conformidad (auditorías/informes)	TIC cumple normas, obligaciones y directrices
	Establecer y aplicar políticas (usó TI interno)	Oportunos, completos, adecuados (necesidades del negocio)	Conformidad de gobierno de TIC
	Personal de TIC cumple directrices, desarrollo y conducta	Actividades de TIC	
	Ética rige acciones desarrolladas por TIC		
Factor Humano	Actividades de TI compatibles con factor humano	Actividades TIC, identificar y prestar atención	Actividades de TIC: Identificar

Principios	Dirigir	Monitorizar	Evaluar
	Informar cualquier individuo, riesgos y problemas	Prácticas de trabajo consistente al uso adecuado de TIC	Actividades de TIC, considera debidamente
	Administración de riesgos según políticas y procedimientos		
	Escalado a los decisores		

Tabla 2. Principios del estándar ISO/IEC 38500 [7]

### 2.3.5. Conclusiones

Como presenta [7], la importancia de tener un gobierno de TI basado en el estándar ISO/IEC 38500 viene dado por las características propias del mundo actual y sus requerimientos.

Es muy importante tener presente que el gobierno de TIC, implica la administración de recursos de los más preciados de una organización y como sistema, su gestión no es responsabilidad exclusiva de una unidad organizativa especializada, sino que implica a los administradores a todos los niveles.

## 2.4. Desarrollo de instrumento para la evaluación de modelos de madurez para la gobernanza de Big Data que incorpore habilidades duras y blandas

Luego de analizar marcos de trabajo y estándares, se halla la misma problemática que los autores del proyecto base “Modelo para la evaluación de buenas prácticas de gobernanza de Big Data” [3], el foco de los marcos de trabajo y estándares se sitúa particularmente en las habilidades duras sin tomar en cuenta las habilidades blandas. Al igual que [3], se encontró una excepción en el caso de la norma ISO/IEC 38500[4] la cual considera el factor humano lo suficiente como para presentarlo en uno de los principios que propone. Se observó que no le da la importancia que merece al no contemplar los diferentes enfoques que requiere.

Cuando se habla de estos tipos de habilidades, por un lado se refiere a todos los requisitos de conocimiento que se deben poseer (duras) y por otro lado a todos aquellos atributos o características que posee una persona y que le permiten interactuar con otras de manera efectiva (blandas). Se tratan de habilidades transversales que incluyen desde habilidades sociales hasta hábitos de trabajo.

Como comenta [3], tomando como referencia a [8], el concepto de gobierno de datos es algo relativamente nuevo y este puede ser distinguido en seis diferentes divisiones, cuatro de ellas son según su alcance:

- Auditoría
- Toma de decisiones
- Conformidad con la perspectiva del gobierno corporativo
- Desempeño del gobierno corporativo
- Las otras dos en cambio se basan en la dirección que toma el gobierno de TI.

El objetivo de [8] es lograr determinar cómo mejorar la efectividad y madurez del gobierno de TI y desarrollar un instrumento de evaluación basado en el modelo MIG (modelo diseñado en una investigación previa de los autores).

El estudio considera los modelos de madurez como “capacidades organizativas” basadas en la visión de recursos utilizada en la literatura de gestión estratégica. Siempre enfatizando el impacto positivo para la organización.

Presenta un modelo de madurez el cual combina habilidades duras y blandas, y agrega el “contexto” como tercer factor clave logrando de esta manera un modelo que abarque todo lo que debe analizarse en una organización. Los autores especialmente mencionan y resaltan la falta de existencia de un marco de trabajo, norma o estándar que incluya habilidades blandas, sin mencionar uno que combine ambas habilidades. El modelo planteado por Daniel Smits y Jon Van Hillegersberg está basado en MIG y se detalla en la Tabla 3 a continuación.

Gobierno	Dominio	Área de foco	Modelo de madurez utilizado
Blando	Comportamiento	Mejora continua	Bessant et al. [9]
		Liderazgo	Collins [10]
	Colaboración	Participación	Magdaleno et al. [11]
		Entendimiento y confianza	Reich et al. [12]
Duro	Estructura	Roles y funciones	CMM(Capability Maturity Model)
	Proceso	Toma de decisiones de TI Planificación Monitoreo	
Contexto	Interno	Cultura Organización informal	Quinn
	Externo	Sector	

Tabla 3. Modelo MIG, adaptado de [8]

Como se puede apreciar, se divide al gobierno de TI en tres grandes gobiernos: blando, duro y su contexto.

Dominios de las habilidades blandas:

Comportamiento: se investiga el impacto que los individuos, grupos y estructuras tienen sobre la conducta dentro de las organizaciones.

Colaboración: aquí se ubican todos los procesos donde se realizan a partir del trabajo de varias personas en conjunto para lograr un resultado difícil de realizar o con un objetivo en común.

A su vez, los dominios tienen dos áreas de foco. Estas áreas para el dominio del comportamiento son:

Mejora continua: se trata de un enfoque para la mejora de procesos operativos que se basa en la necesidad de revisar continuamente las operaciones de los problemas, la reducción de costos oportunidad, la racionalización, y otros factores que en conjunto permiten la optimización.

Liderazgo: definido como un conjunto de habilidades gerenciales o directivas que un individuo tiene para influir en la forma de ser o actuar de las personas o en un grupo de trabajo determinado, haciendo que este equipo trabaje con entusiasmo hacia el logro de sus metas y objetivos.

Y para el dominio de colaboración tenemos las áreas de:

Participación: refiriéndose a la acción de involucrarse en cualquier tipo de actividad de forma intuitiva o cognitiva.

Entendimiento y confianza: definido como la seguridad que posee un individuo hacia otro, y como estos se conocen y comprenden de manera tal de funcionar como un equipo.

Todas las áreas de foco del gobierno blando están basadas en su propio modelo de madurez, los cuales los autores de [8] toman para su publicación. Para validar dichos modelos usan [9, 10, 11, 12].

Continuando los detalles de la Tabla 3, tomamos el gobierno duro, el cual los autores lo definen como dos dominios que son de estructura y de los procesos. Estos dominios se dividen en diferentes áreas de foco.

En el caso de la estructura:

Roles y funciones: se refiere a la realización de una cierta actividad y el responsable de su resultado.

Redes formales: las cuales se basan en canales de comunicación centralizados y descentralizados.

En el caso de los procesos:

Toma de decisiones de TI: el cual es un proceso por medio del cual se obtiene como resultado una o más decisiones con el propósito de dar solución a una situación.

Planificación: como un proceso, parte de la toma de decisiones para alcanzar un futuro deseado.

Monitoreo: definido como el observar una situación de cambios que se pueda producir con el tiempo.

Por último tenemos el gobierno del contexto. En este, se referencia a todo con lo que la organización interactúa, sean estos factores, condiciones o situaciones, externas e internas, que puedan afectar a la misma. Los autores plantean que hay investigaciones que han demostrado que el gobierno de TI es situacional, y es justamente por esto que el contexto se torna en un factor relevante el cual tiene que ser incluido en el modelo.

Los autores utilizan el modelo MIG el cual es un modelo de madurez de área de enfoque. Este mismo utiliza el marco de valores de Quinn, sin tomar la perspectiva de algún otro modelo de madurez existente.

El modelo planteado, indica el modelo de madurez en el cual esta basada cada una de las áreas de foco. Se puede apreciar que las habilidades duras utilizan el modelo CMM y es de este mismo que sus 5 áreas de foco parten de allí. En el capítulo a continuación, se detalla el modelo CMMI, el cual es una integración de varios modelos CMM que se utilizan en distintas áreas de interés.

Después de plantear el modelo, los autores buscan definir un instrumento de evaluación basándose en MIG. Una vez realizada la investigación para poder confeccionar el instrumento deseado logran como resultado, el instrumento MIG, que consta de dos partes.

La primera parte está dirigida al participante y contiene las declaraciones y otras preguntas.

La segunda parte es para el investigador y se utiliza para crear la hoja de resultados.

Ambas partes son separadas para evitar que los resultados no influyan al participante al completar con la evaluación. La primera parte del instrumento evaluativo de MIG, consta de un formulario general con información demográfica y dos cuestionarios. En total son 104 preguntas.

Se puede generar un gráfico de resultados para poder observar el nivel de madurez de la organización al importar los resultados de la primera parte al instrumento de la segunda.

## **2.5. CMMI**

Como se menciona previamente, hasta el momento no existe un modelo de madurez para la gobernanza de Big Data que sea globalmente aceptado por la industria. El objetivo del artículo presente es la realización de dicho modelo y es por este motivo que se optó por analizar otro tipo de modelo de madurez que sí tenga reconocimiento global. Luego de manejar distintas opciones, se tomó la versión 1.3 del Capability Maturity Model Integration (CMMI) o en español como Integración del Modelo de Madurez de Capacidades.

### **2.5.1. Que es**

Tal como dice [13], CMMI es un modelo que contiene las mejores prácticas y proporciona a las organizaciones los elementos críticos que hacen que sus procesos comerciales sean efectivos. Fue creado inicialmente para procesos relacionados con el Desarrollo e Implementación de Software por la Universidad Carnegie-Mellon. Esto se usó por primera vez en 1987, con el nombre de Capability Maturity Model (CMM). Este nombre y sus cinco niveles de representación por etapas se inspiraron en el modelo de madurez de fabricación de Crosby.

La primera vez que se utilizó CMM fue en programas de defensa, pero la realidad es que ha sido ampliamente aceptado, hasta el punto que ha recibido varias correcciones e iteraciones. Debido a este éxito, se han desarrollado varios tipos de CMM para una amplia gama de campos de trabajo, además del software.

Sin embargo, debido a la gran difusión de los modelos de desarrollo de software, comenzaron a aparecer varias confusiones. Por este motivo el gobierno estadounidense terminó financiando un proyecto de dos años en el que participaron más de 200 expertos del mundo industrial y académico, con el objetivo de crear un marco único expandible para la ingeniería de sistemas y software, y el desarrollo de productos. El resultado de este proyecto es CMMI: el modelo más adoptado el día de hoy.

### **2.5.2. Modelos**

Las mejores prácticas de CMMI se publican en documentos llamados modelos, cada uno de los cuales aborda un área de interés diferente. La versión 1.3 proporciona modelos para tres áreas de interés:

CMMI para Desarrollo (CMMI-DEV): Aborda los procesos de desarrollo de productos y servicios.

CMMI para Adquisición (CMMI-ACQ): Aborda los procesos de gestión de la cadena de suministro, adquisición y subcontratación en el gobierno y la industria.

CMMI para Servicios (CMMI-SVC): Se aborda la guía para la prestación de servicios dentro de una organización y a clientes externos.

Con estos tres modelos de las áreas de interés, CMMI asigna uno de cinco niveles de madurez, describiendo qué tan efectivos están siendo dentro de la empresa.

### **2.5.3. Tipos de representaciones**

CMMI está estructurado de la siguiente manera:

- Niveles de madurez (representación por etapas) o niveles de capacidad (representación continua)
- Áreas de proceso
- Objetivos: genéricos y específicos
- Características comunes
- Prácticas: genéricas y específicas

Este capítulo discutirá sobre dos representaciones de CMMI y el resto de los temas se tratarán en capítulos posteriores.

#### **2.5.3.1. Representación por etapas**

La representación por etapas es el método usado en el Software CMM. Es un enfoque que utiliza conjuntos predefinidos de áreas de proceso para definir un camino de mejoramiento en una organización. Esta ruta de mejoramiento es descrita por los niveles de madurez dentro del modelo, los cuales permiten lograr mejoras en los procesos organizativos [13].

La representación por etapas:

- Proporciona una secuencia de mejoras, cada una de ellas sirve como base para la siguiente.
- Permite hacer comparaciones entre las organizaciones por el uso de los niveles de madurez.
- Proporciona una única clasificación que resume resultados de la evaluación y permite establecer comparaciones entre las organizaciones.

Por lo tanto, la representación por etapas proporciona una hoja de ruta predefinida para la mejora organizacional basada en la agrupación probada y el ordenamiento de procesos y las relaciones organizacionales asociadas. No se puede desviar de la secuencia de pasos.

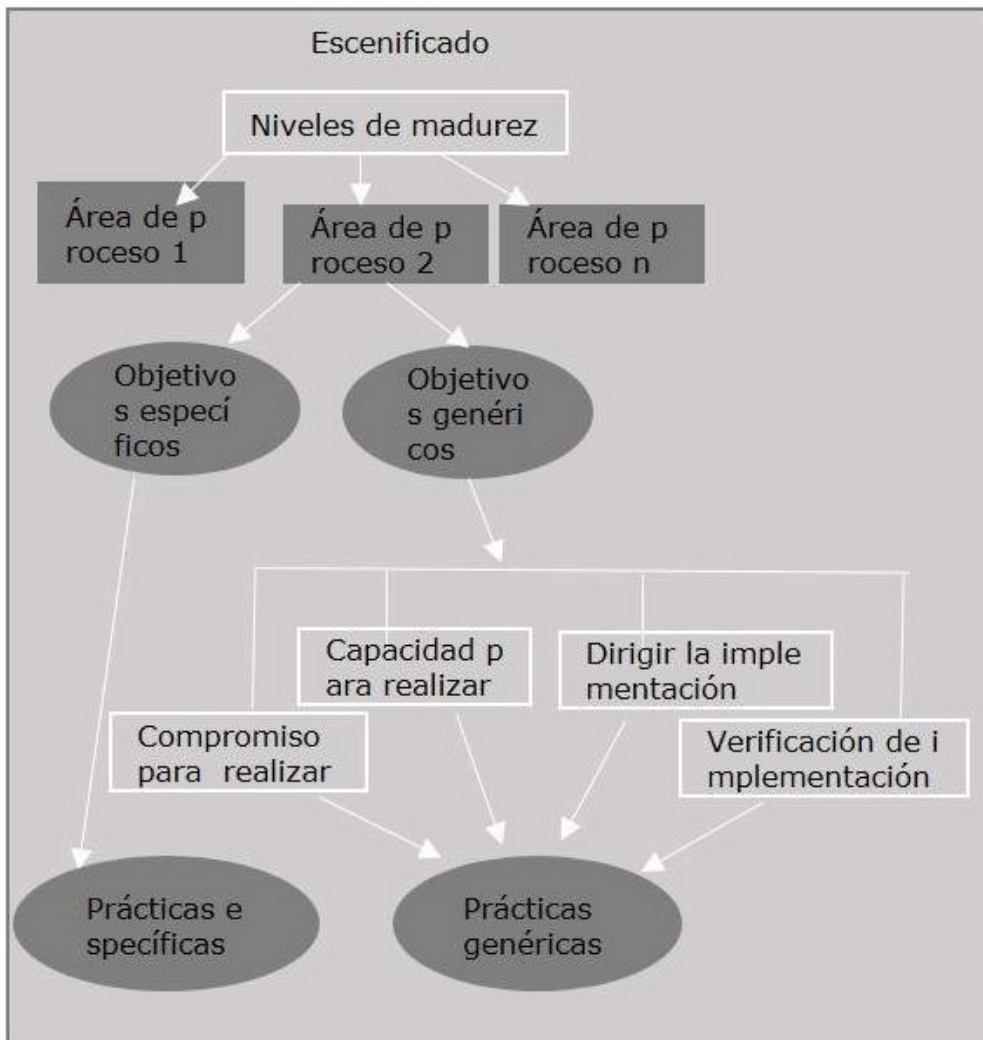


Ilustración 6. Estructura por etapas

### 2.5.3.2. Representación continua

Este enfoque permite a una organización seleccionar un área de proceso y hacer mejoras basadas en el mismo. La representación continua utiliza niveles de capacidad para caracterizar la mejora en relación con un área de proceso individual [13].

La representación continua:

- Permite seleccionar el orden de mejora que mejor se adapte a los objetivos comerciales de la organización y mitiga las zonas de riesgo de la organización.
- Permite comparaciones entre organizaciones dentro de un proceso por área de proceso.

Esta representación, por lo tanto, proporciona flexibilidad a las organizaciones para elegir los procesos de mejora, así como la cantidad de mejoras requeridas.

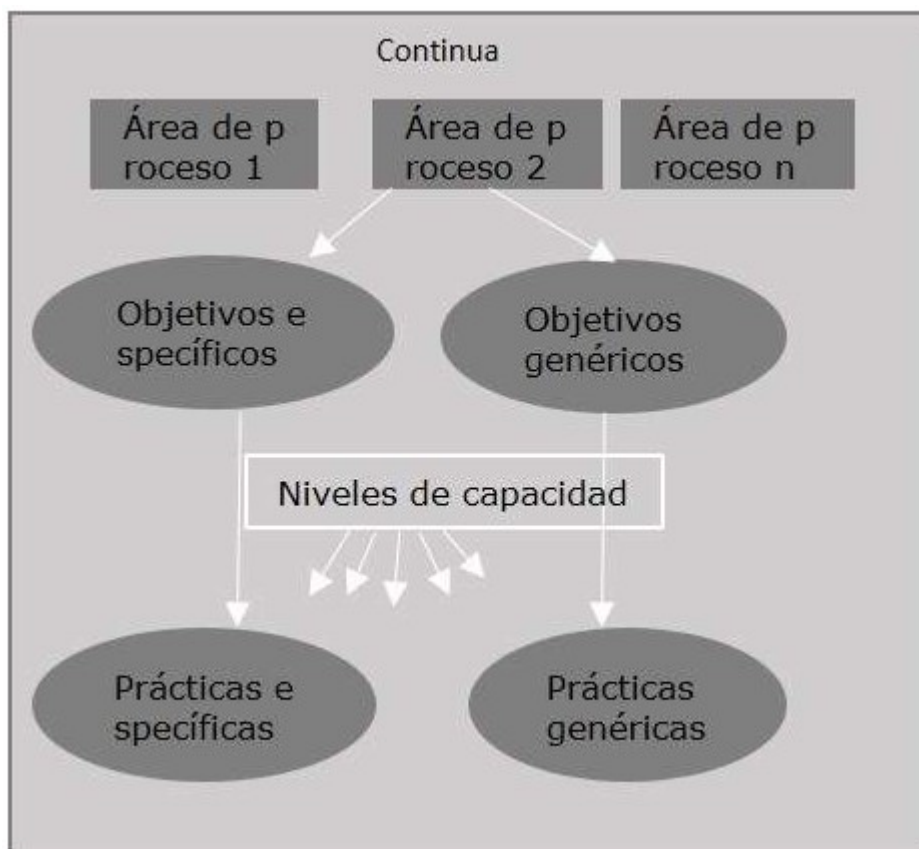


Ilustración 7. Representación continua

### 2.5.3.3. Comparando representaciones

Continua	Etapas
Áreas de proceso están organizadas por categorías de área de procesos.	Áreas de proceso están organizados por niveles de madurez.
Mejora se mide mediante niveles de capacidades. Niveles de Capacidades miden la madurez de un proceso particular de una organización, que va de 0 a 5.	Mejora se mide utilizando los niveles de madurez. Los niveles de madurez miden la madurez de un conjunto de procesos de una organización: oscila entre 1 a 5.
Hay dos tipos de prácticas específicas: base y avanzado. Todas las prácticas específicas aparecen en la representación permanente.	Sólo hay un tipo de práctica específica. Los conceptos básicos y avanzados de las prácticas no son utilizados. Todas las prácticas específicas aparecen en la representación por etapas excepto cuando una base avanzada de par de prácticas aparece en la representación continua, en cuyo caso sólo la práctica avanzada aparece en la representación por etapas.
Niveles de capacidades se usan para organizar las prácticas genéricas.	Características comunes se utilizan para organizar las prácticas genéricas.
Todas las prácticas genéricas están incluidas en cada área de proceso.	Sólo en el nivel 2 y nivel 3 se incluyen las prácticas genéricas.
Clasificación equivalente permite determinar un nivel de madurez de una organización.	No es necesario un mecanismo de equivalencia para respaldar la representación continua porque cada organización puede elegir qué mejorar y cuánto mejorar utilizando la representación por etapas.

Tabla 4. Comparación representaciones [13]

#### 2.5.3.4. Cual usar

Cada representación tiene sus ventajas sobre la otra, algunas organizaciones usan ambas representaciones para abordar requisitos particulares en varios momentos en sus programas de mejora.

La madurez organizacional es el foco de la representación por etapas, mientras que la capacidad del área de proceso es el foco de la representación continua.

La madurez organizacional y la capacidad del área de proceso son conceptos similares. La diferencia entre ellos es que la madurez organizacional pertenece a un conjunto de áreas de proceso en una organización, mientras que la capacidad del área de proceso se ocupa de un conjunto de procesos relacionados con un área de proceso única o práctica específica.

El siguiente diagrama muestra las dos presentaciones. En este diagrama, ML indica el nivel de madurez y PA indica el área de proceso.

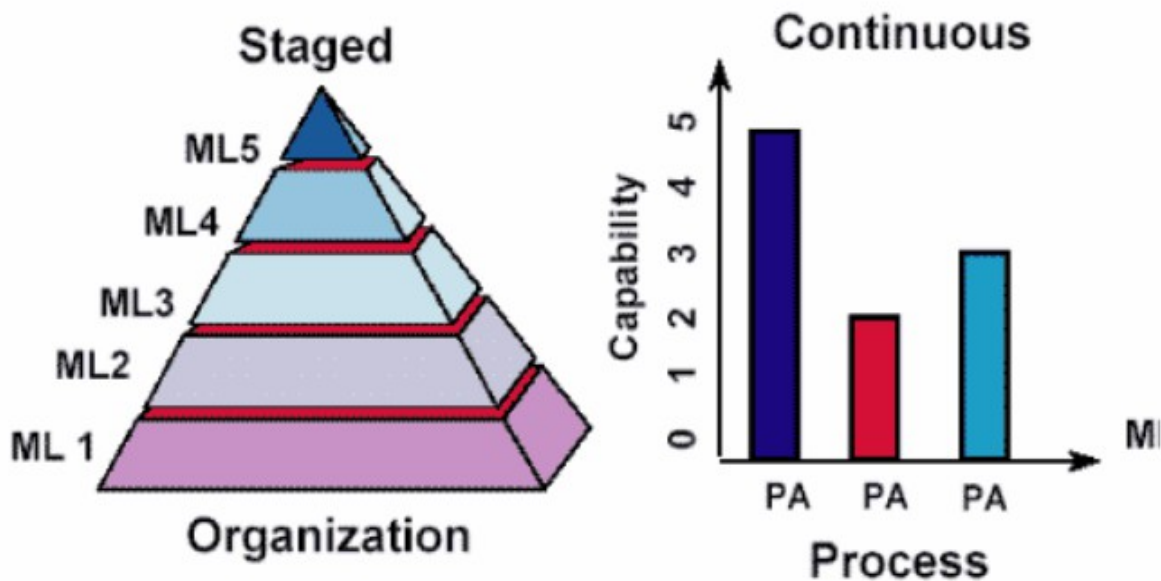


Ilustración 8. Diagramas de ambas representaciones

## 2.5.4. Niveles de madurez

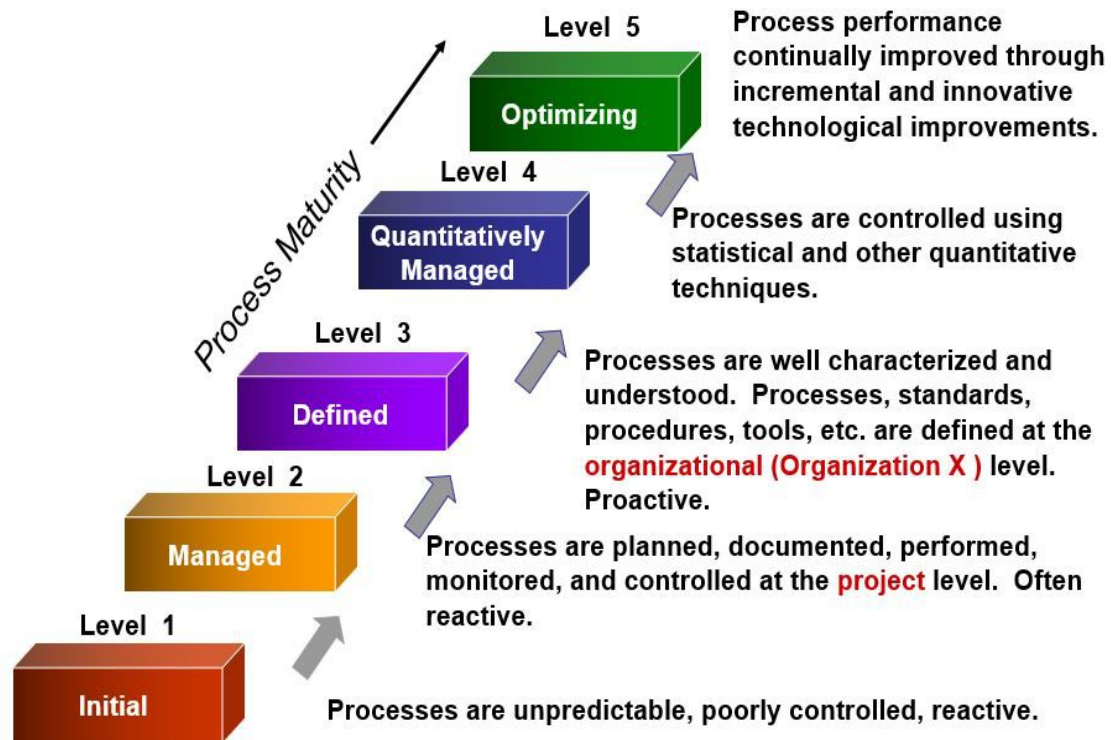


Ilustración 9. Niveles de madurez

Los niveles de madurez [13], también apreciables en la ilustración, en detalle son:

Inicial: Las organizaciones que se encuentran en este nivel no disponen de un ambiente estable para el desarrollo y mantenimiento de software. Aunque se utilicen técnicas correctas de ingeniería, los esfuerzos se ven minados por falta de planificación. El éxito de los proyectos se basa la mayoría de las veces en el esfuerzo personal, aunque a menudo se producen fracasos y casi siempre retrasos y sobrecostos. El resultado de los proyectos es impredecible.

Repetible: En este nivel las organizaciones disponen de una serie de prácticas institucionalizadas de gestión de proyectos, existen unas métricas básicas y un razonable seguimiento de la calidad. La relación con subcontratistas y clientes está gestionada sistemáticamente.

Definido: Además de una buena gestión de proyectos, a este nivel las organizaciones disponen de correctos procedimientos de coordinación entre grupos, formación del personal, técnicas de ingeniería más detallada y un nivel más avanzado de métricas en los procesos.

Gestionado: Se caracteriza porque las organizaciones disponen de un conjunto de métricas significativas de calidad y productividad, que se usan de modo sistemático para la toma de decisiones y la gestión de riesgos. El software resultante es de alta calidad.

Optimizado: La organización completa está volcada en la mejora continua de los procesos. Se hace uso intensivo de las métricas y se gestiona el proceso de innovación.

### 2.5.5. Niveles de capacidad

Un nivel de capacidad es una meseta evolutiva bien definida que describe la capacidad de la organización en relación con un área de proceso. Un nivel de capacidad consiste en prácticas genéricas y específicas relacionadas para un área de proceso que pueden mejorar los procesos de la organización asociados con esa área. Cada nivel es una capa en la base para la mejora continua del proceso.

Por lo tanto, los niveles de capacidad son acumulativos, es decir, un nivel de capacidad más alto incluye los atributos de los niveles más bajos.

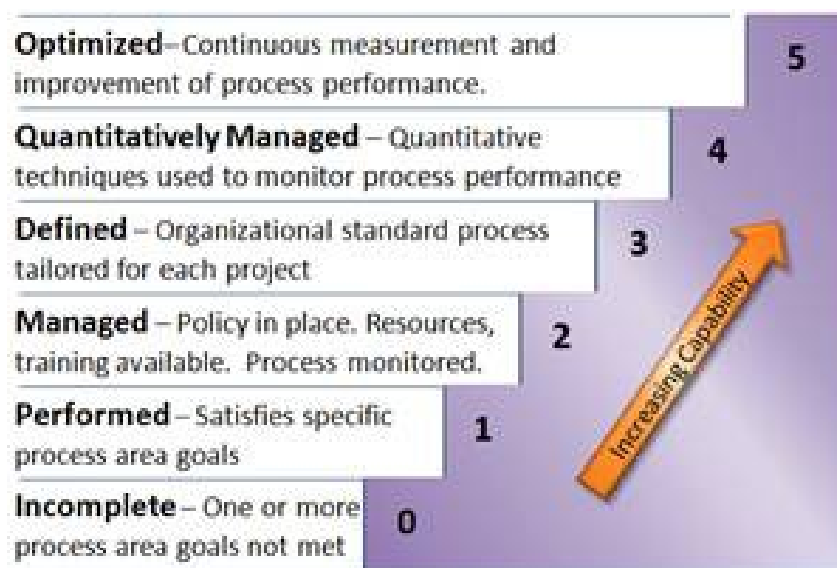


Ilustración 10. Niveles de capacidad.

En los modelos CMMI con una representación continua, hay seis niveles de capacidad.

Incompleto: Un "proceso incompleto" es un proceso que no se realiza o se realiza parcialmente. Uno o más de los objetivos específicos del área de proceso no se cumplen y no existen objetivos genéricos para este nivel, ya que no hay razón para institucionalizar un proceso parcialmente realizado.

Ejecutado: Existe la posibilidad de que el rendimiento no sea estable y no cumpla con objetivos específicos como la calidad, el costo y el cronograma, pero se puede realizar un trabajo útil. Significa que está haciendo algo, pero no puede probar que realmente está funcionando.

Gestionado: Un proceso gestionado se planifica, realiza, supervisa y controla para proyectos individuales, grupos o procesos independientes para lograr un propósito determinado. La gestión del proceso logra tanto los objetivos del modelo para el proceso como otros objetivos, como el costo, el cronograma y la calidad.

Definido: Un proceso definido es un proceso gestionado que se adapta a la política de procesos estándar de la organización de acuerdo a las pautas de la dirección, y aporta productos de trabajo, medidas y otra información de mejora de procesos a los activos del proceso organizacional.

Cuantitativamente gestionado: proceso definido que se controla mediante técnicas cuantitativas y de estadística. Los objetivos cuantitativos para la calidad y el rendimiento del proceso se establecen y utilizan como criterios en la gestión del proceso. La calidad y el rendimiento del proceso se entienden en términos estadísticos y se gestionan durante toda la vida del proceso.

Optimizante: Además de ser un proceso cuantitativamente gestionado, de forma sistemática se revisa y modifica o cambia para adaptarlo a los objetivos del negocio. Se centra en mejorar continuamente el rendimiento del proceso a través de mejoras incrementales e innovadoras.

### **2.5.6. Área de procesos**

La Integración del Modelo de Madurez de Capacidades define un Área de Proceso como "*Un conjunto de prácticas relacionadas en un área que, cuando se implementa colectivamente, satisface un conjunto de objetivos considerados importantes para mejorar en esa área*".

Tanto CMMI para Desarrollo v1.3 como CMMI para Adquisición v1.3 identifican 22 áreas de proceso, mientras que CMMI para Servicios v1.3 identifica 24 áreas de proceso.

Modelo	Nivel de Madurez	Áreas de proceso
CMMI-DEV	2. Repetible	CM - Gestión de configuración
		MA - Medición y Análisis
		PMC - Monitoreo y control de proyectos
		PP - Planificación de proyectos
		PPQA - Aseguramiento de calidad de procesos y productos
		REQM - Gestión de requisitos
		SAM - Gestión de acuerdos de proveedores
	3. Definido	DAR - Análisis y resolución de decisiones
		IPM - Gestión integrada de proyectos
		OPD - Definición del proceso organizacional
		OPF - Enfoque del proceso organizacional
		OT - Entrenamiento Organizacional
		PI - Integración de productos
		RD - Desarrollo de requisitos
		RSKM - Gestión de riesgos
		TS - Solución técnica
		VAL - Validación
		VER - Verificación

	4. Gestionado	OPP - Rendimiento del proceso organizacional
		QPM - Gestión cuantitativa de proyectos
	5. Optimizado	CAR - Análisis casual y resolución
		OPM - Gestión del desempeño organizacional

Tabla 5. Áreas proceso con sus respectivos niveles de madurez

### 2.5.7. Evaluación

Una organización no puede ser certificada en CMMI; en cambio, se evalúa a la organización. Dependiendo del tipo de evaluación, la organización puede recibir una calificación de nivel de madurez (1-5) o un perfil de logro de nivel de capacidad.

Como explica [13], la evaluación CMMI es un examen de uno o más procesos por un equipo capacitado de profesionales que utiliza un modelo de referencia de evaluación como base para determinar las fortalezas y debilidades de una organización.

Las evaluaciones requieren planificación. Al planificar la evaluación de la organización, hay que determinar previamente el alcance de la unidad organizativa, qué disciplinas incluir, si el equipo de evaluación consistirá de miembros internos o externos a la organización, qué proyectos se incluirán, los individuos a ser entrevistados y el tipo o clase de tasación necesaria.

Las evaluaciones consideran tres categorías de componentes del modelo tal como se definen en el CMMI:

Requerido: solo objetivos específicos y genéricos.

Esperado: solo prácticas específicas y genéricas.

Informativo: incluye sub-prácticas y productos de trabajo típicos.

El SEI (Software Engineering Institute) ha publicado dos documentos guía para las evaluaciones CMMI:

Requisitos de evaluación para CMMI (ARC): contiene los requisitos para tres clases de métodos de evaluación Clase A, Clase B y Clase C. Estos requisitos son las reglas para definir cada clase de método de evaluación.

Método estándar de evaluación CMMI para la mejora de procesos (SCAMPI): El documento de descripción del método (Method Description Document - MDD) es actualmente el único método de evaluación Clase A aprobado.

Es decir, SCAMPI satisface todos los requisitos de un método de evaluación ARC Clase A y ha sido aprobado por el SEI.

En total, hay tres clases de métodos de evaluación CMMI: clase A, clase B y clase C.

### **2.5.8. Roles y responsabilidades**

Aquí se hará mención a los principales actores involucrados en el esfuerzo de mejora de procesos. Tener en cuenta que dependiendo de la organización se requerirán más o menos grupos y que una persona puede cumplir muchos de estos roles simultáneamente o en serie, dependiendo del tamaño de la organización y la complejidad del esfuerzo de mejora de procesos.

Patrocinador de PI (Process Improvement): la persona de la organización responsable de supervisar todo el esfuerzo de PI. Es quien generalmente tiene el poder de asignar fondos y personal, y suele estar en el nivel de dirección o superior.

Campeón de PI: esta es la persona de relaciones públicas para el esfuerzo de PI, que puede o no servir como líder de EPG. Es quien comercializa la idea, el enfoque y los resultados de PI.

Líder del grupo de procesos de ingeniería (EPG – Engineering Process Group): esta persona dirige el grupo que revisa los procesos. Asigna tareas a los miembros de la EPG, supervisa sus esfuerzos y planifica las tareas diarias de la EPG.

Miembros de la EPG: estas personas sirven en la EPG como miembros del comité. Son responsables de garantizar que la documentación de mejora del proceso se escriba y se siga.

También son responsables de generar métricas para rastrear el proceso de mejora del proceso. Ellos lideran las PAT.

Equipos de acción de proceso (PAT – Process Action Teams): estos equipos generan la documentación, políticas, procesos, procedimientos, cartas y planes de acción de mejora de procesos.

Socio de transición: por lo general, una o dos personas que son consultores externos contratados para ayudar a establecer, planificar, liderar y monitorear el progreso en la mejora del proceso organizacional. Estas personas aportan experiencia en la mejora de procesos de varias otras organizaciones e industrias.

### **2.5.9. Conclusión**

CMMI ha demostrado ser de gran utilidad a las organizaciones al permitirles tener un conjunto completo de prácticas que se pueden reempaquetar en el flujo de trabajo que deseen. Además, les ha de proporcionar una hoja de ruta para que las prácticas se puedan adoptar de forma gradual para madurar la organización con el tiempo y por último la posibilidad de que las organizaciones puedan obtener una evaluación que brinde una calificación pública reconocida, para demostrar la capacidad de la misma.

## 3. Investigación

### 3.1. Preguntas de investigación

Teniendo ya definido el objetivo general del trabajo que es la construcción del modelo, y habiendo también definido los objetivos o etapas del proyecto, es necesario plantear algunas preguntas de investigación.

Dichas preguntas servirán de guía en la investigación, marcando el foco de la búsqueda del material literario y conduciendo hacia el objetivo general del proyecto.

Se plantean tres preguntas principales.

Pregunta 1.

¿ Existe un modelo de madurez para la gobernanza de big data en la industria ?

Esta primera pregunta busca dilucidar si en el entorno actual de la industria existe un modelo de madurez, que sea referente y aceptado, para la gobernanza de Big Data. Para esto se hará una revisión literaria, analizaremos y buscaremos indicios de que esto sea así.

De ser una respuesta negativa, no clara o que genere dudas se continuará con una siguiente pregunta.

Pregunta 2.

¿ Existe necesidad en la industria de un modelo de madurez para la gobernanza de Big Data ?

Con esta segunda pregunta se busca descifrar si ante la falta de dicho modelo, la industria lo requiere. Es decir, qué impacto positivo tendría dicho modelo en la gestión y éxito de la Big Data.

Pregunta 3.

¿ Es posible construir un modelo de madurez de este tipo ?

Por último, se busca descubrir si el fruto de la investigación brinda herramientas e información necesaria para poder desarrollar un modelo útil para la adopción con éxito de la gobernanza de Big Data.

### 3.2. Metodología de trabajo

Para comenzar debemos encontrar metodologías que permitan definir un proceso confiable en busca de la información deseada. Por esto, para lograr una investigación exitosa y dar un marco científico a la misma, se utilizará como metodología de trabajo la revisión sistemática de la literatura propuesta por Barbara Kitchenham en su artículo[14].

Kitchenham define “Las revisiones sistemáticas tienen como objetivo presentar una evaluación justa de un tema de investigación mediante el uso de una metodología confiable, rigurosa y auditable”. La misma sugiere para las revisiones sistemáticas un proceso en tres etapas o fases principales, planificación de la revisión, realización de la revisión e informe de revisión, como lo muestra la Ilustración 11 propuesta por el proyecto insumo [3].

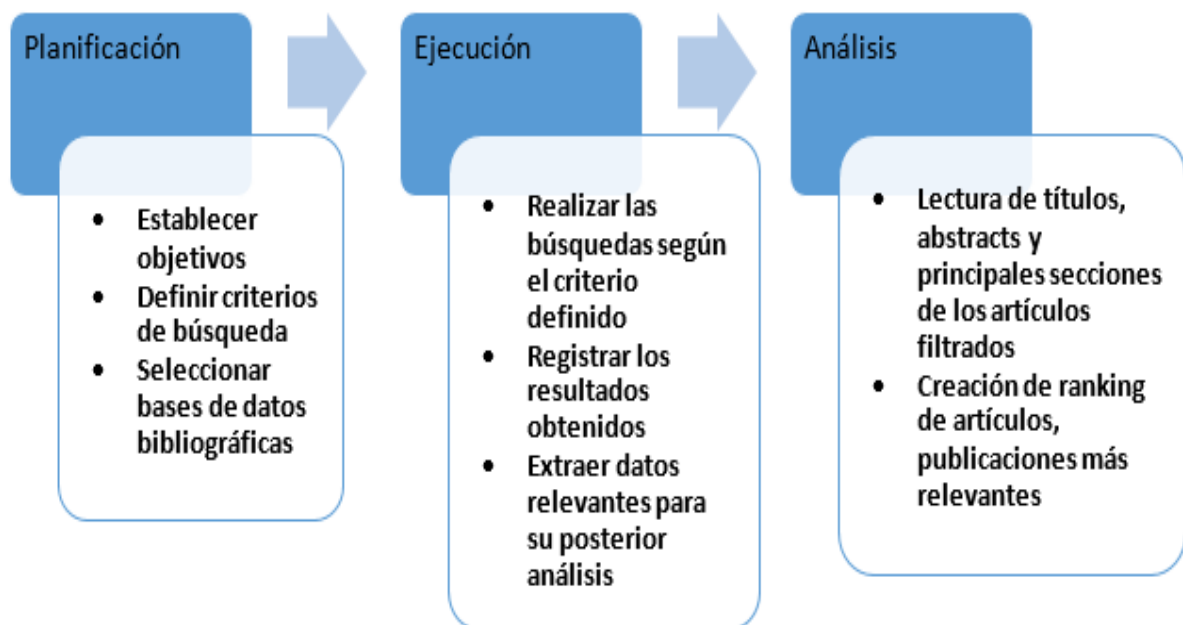


Ilustración 11. Proceso de búsqueda sistemática

En los siguientes capítulos se irán ejecutando estas etapas para desarrollar la investigación.

### 3.3. Planificación de la revisión sistemática

Como primer punto en la planificación Kitchenham [14] plantea la necesidad de definir un protocolo de revisión. Dicho protocolo incluyen entre otros los siguientes elementos:

- Las preguntas de investigación que la revisión pretende responder, osea los objetivos.
- La estrategia que se utilizará para buscar estudios primarios, incluidos los términos de búsqueda y recursos para buscar, los recursos incluyen bases de datos.
- Estudiar criterios y procedimientos de selección.

Para el desarrollo de los elementos planteados se comenzará por definir los criterios de búsqueda y las bases de datos sobre la cuales se trabajará, debido que las preguntas de investigación han sido definidas en capítulos anteriores.

Los criterios de búsqueda permitirán obtener resultados homogéneos a nivel de contenido en cada una de las bases de datos. Se debe tener en cuenta para la definición de estos criterios que se parte de la investigación del proyecto insumo [3], en el cual se realizaron dos búsquedas:

- Búsqueda 1: “Data Governance”
- Búsqueda 2: “Data Governance” + “Maturity Model”

Teniendo ya allanado el camino en cuanto a la gobernanza de datos se enfocamos la búsqueda a los modelos de madurez y la Big Data. Con este foco se define la Búsqueda 1, “maturity models” y la Búsqueda 2, “maturity models” + “big data”. Por otro lado también se definieron los filtros para años, tipos de acceso y sector al que pertenece la información. En la Tabla 6 se muestra el resultado final de la selección de criterios de búsqueda.

<b>Filtro</b>	<b>Valor</b>
<b>Términos de búsqueda</b>	<b>Búsqueda 1: “maturity models” Búsqueda 2: “maturity models” + “big data”</b>
<b>Año de publicación</b>	<b>2010 en adelante</b>
<b>Sector</b>	<b>Computer science, Bussiness</b>
<b>Tipo de acceso</b>	<b>Open Access</b>

Tabla 6. Filtros de búsqueda

Como siguiente elemento se definirán las fuentes de datos. Para esto se enfoca en dos áreas, por un lado la académica y por otro la del sector industrial. Dentro del área académica se seleccionan bases de datos del portal Timbo, con esto se asegurará la confiabilidad de éstas. Para el área del sector industrial se realizó la búsqueda en Google por ser el más utilizado y confiable. En la Tabla 7 se muestra el resultado final de la selección de bases de datos.

<b>Base de datos</b>	<b>URL</b>
<b>Scopus</b>	<b><a href="https://www.scopus.com">https://www.scopus.com</a></b>
<b>SpringerLink</b>	<b><a href="https://link.springer.com">https://link.springer.com</a></b>
<b>IEEE</b>	<b><a href="https://ieeexplore.ieee.org">https://ieeexplore.ieee.org</a></b>
<b>Science Direct</b>	<b><a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a></b>
<b>Google</b>	<b><a href="https://www.google.com">https://www.google.com</a></b>

Tabla 7. Bases de datos para búsqueda

Como último elemento se definen los criterios y procedimientos de selección. Para esto se plantearán diferentes etapas de filtrado o selección hasta llegar a artículos que realmente sean relevantes.

La primer selección estará dada por los propios términos de búsqueda 1 y 2, en esta etapa se registra el número de artículos encontrados bajo estas búsquedas.

La segunda selección se realiza la lectura de títulos y abstract de todos los artículos encontrados en la primera selección. De este modo se filtra y se obtiene los que se cree tienen mayor relevancia.

Por último se realiza la selección final, donde se leen completamente todos los artículos obtenidos en la selección anterior y quedan solo los que se piensa que serán la base para llegar a los objetivos de la investigación. En la Tabla 8, se muestra el resumen de los criterios de selección.

<b>Etapa de selección</b>	<b>Característica</b>
<b>Primer selección</b>	<b>Total de artículos encontrados tras la búsqueda</b>
<b>Segunda selección</b>	<b>Total de artículos que quedan tras leer el abstract</b>
<b>Selección final</b>	<b>Total de artículos que quedan luego de la lectura completa del artículo</b>

Tabla 8. Etapas de selección de artículos

Luego de obtenida la selección final se clasifican los artículos por importancia, es decir de los más a los menos influyentes. Para esto se utiliza una escala tipo Likert cuyos valores van del 1 al 5 donde uno significa no relevante y cinco muy relevante.

### **3.4. Ejecución de la revisión sistemática**

Esta etapa consiste en la ejecución de lo planificado en el artículo anterior. Es decir, se realizan las búsquedas en las diferentes bases de datos, aplicado los criterios de filtros. Para cada una de las búsquedas registra los datos de los artículos discriminados por base de datos, autor.

En el siguiente capítulo se realizará el análisis y la selección de artículos.

### 3.5. Análisis y resultados de la revisión sistemática

En esta etapa se analizarán los artículos encontrados y se irán efectuando las selecciones definidas en la planificación hasta llegar a los artículos finales. Dentro del mismo se mostrarán tablas con diferentes análisis de resultados de las búsquedas.

A continuación en la Tabla 9, se muestran los resultados para las Búsquedas 1 y Búsquedas 2, definidas en el capítulo de planificación. En la misma los resultados se discriminan por base de datos y dentro de cada base se registra la cantidad de artículos que van quedando luego de aplicada la primer selección, segunda selección y la selección final, definidas en el capítulo de planificación.

Base de Datos	Número de publicaciones	Número de publicaciones
	Búsqueda 1	Búsqueda 2
IEEE	Primer selección: 38	Primer selección: 42
	Segunda selección: 6	Segunda selección: 7
	<b>Selección final: 2</b>	<b>Selección final: 2</b>
Springer	Primer selección: 103	Primer selección: 59
	Segunda selección: 8	Segunda selección: 0
	<b>Selección final: 2</b>	<b>Selección final: 0</b>
Scopus	Primer selección: 314	Primer selección: 6
	Segunda selección: 36	Segunda selección: 4
	<b>Selección final: 4</b>	<b>Selección final: 0</b>
Science Direct	Primer selección: 101	Primer selección: 62
	Segunda selección: 4	Segunda selección: 4
	<b>Selección final: 1</b>	<b>Selección final: 1</b>
Google	Primer selección: 300	Primer selección: 300
	Segunda selección: 7	Segunda selección: 10

Base de Datos	Número de publicaciones	Número de publicaciones
	Búsqueda 1	Búsqueda 2
IEEE	Primer selección: 38	Primer selección: 42
	<b>Selección final: 1</b>	<b>Selección final: 0</b>
	<b>Total artículos: 10</b>	<b>Total artículos: 3</b>

Tabla 9. Resultados de la búsqueda con selecciones.

A continuación en la Tabla 10 se muestra el análisis realizado discriminando por autores con mayor actividad, dentro de los artículos encontrados y separados por búsquedas. Si bien existen algunos con mayor cantidad de publicaciones no se puede concluir que exista un referente en el tema.

Autor	Número de publicaciones	Autor	Número de publicaciones
	Búsqueda 1		Búsqueda 2
Alshayeb, M.	3	Enke, D.	1
Werner-Lewandowska	3	Gregory, A.	1
Niazi, M.	3	Hoang, D.B.	1
Enkw, J.	3	Krumeich, J.	1
Arif Ali Khan	2	A. Popov	1
Mohammad Alshayeb	2	Le, N.T.	1
Ismail Keshta	2	Loos, P.	1
Mahmood Niazi	2	Sabarina, K.	1

Tabla 10. Autores con mayor actividad.

Luego de obtenida la selección final, los artículos se clasifican por importancia. Esta clasificación, definida en el capítulo de planificación, va de uno a cinco, siendo cinco la más importante. A continuación en la Tabla 11, se detallan el autor, título y clasificación de los artículos de la selección final.

Autor	Título	Clasificación
Stanford University	Stanford Data Governance Maturity Model [15]	5
João Aguiar, Ruben Pereira, José Braga Vasconcelos, Isaias Bianchi	An overlapless incident management maturity model for multi-framework assessment [16]	5
Diogo Proença, José Borbinha	Maturity Models for Information Systems - A State of the Art [17]	5
Mohammad Khoshgoftar, Omar Osman	Comparison of Maturity Models [18]	5
Gabriel García, Ángeles Félix	Development of Maturity Models: A Systematic Literature Review [19]	5
Daniël Smits, Jos van Hillegersberg	The development of a hard and soft IT governance assessment instrument [8]	5
Zaher Ali Al-Sai, Rosni Abdullah, Mohd heikal	A Review on Big Data Maturity Models [21]	4
Diana Heredia, Wilson Nieto	Propuesta de una Metodología para la Adopción del Big Data [22]	4
Ali Al-Badia , Ali Tarhinia, Asharul Islam	Exploring Big Data Governance Frameworks [23]	4

<b>Autor</b>	<b>Título</b>	<b>Clasificación</b>
Khana		
Daniël Smits, Jos van Hillegersberg	The continuing mismatch between IT governance maturity theory and practice: a new approach [24]	3
Majid Al-Ruithe, Elhadj Benkhelifa	Analysis and Classification of Barriers and Critical Success Factors for Implementing a Cloud Data Governance Strategy [25]	2
Diogo Proença, Ricardo Vieira, José Borbinha	Information Governance Maturity Model Final Development Iteration [26]	2
Guangming Cheng, Yao Li, Zhiwei Gao and Xiaoyin Liu	Cloud Data Governance Maturity Model [27]	1

Tabla 11. Artículos de la selección final.

Por último se realizó una selección, con las características de las anteriores, sobre los artículos del proyecto insumo [3], en donde se extrajeron los artículos que aportan a la construcción del modelo. A continuación en la Tabla 12. se detallan el autor, título y clasificación de los artículos de la selección final de proyecto insumo.

<b>Autor</b>	<b>Título</b>	<b>Clasificación</b>
Carlos Juiz, Ricardo Colomo- Palacios, Beatriz Gómez	Cascading ISO/IEC 38500 based Balanced Score Cards to improve board accountability [28]	5
Isaias Scalabrin Bianchi, Rui Dinis	IT Governance Structures in Brazilian, Dutch and Portuguese Universities [29]	5

Autor	Título	Clasificación
Sousab, Rúben Pereirac, Edimara Lucianod		
Anandasivakumar Ekambaram, Anette. Sørensenb, Heidi Bull-Berga, Nils O.E. Olssonb	The role of big data and knowledge management in improving projects and project-based organization [30]	5

Tabla 12. Artículos de la selección final del proyecto insumo.

### 3.6. Respuestas a las preguntas de investigación planteadas

Antes de comenzar a la construcción del modelo de madurez se debe saber si este es necesario. El mismo será necesario si, no existe uno consolidado en la industria, y si es que la industria lo requiere realmente.

Para dilucidar esto es imperioso responder a las preguntas de investigación planteadas en capítulos anteriores.

#### Pregunta 1.

#### ¿ Existe un modelo de madurez para la gobernanza de Big Data en la industria ?

Luego de los temas discutidos previamente se puede afirmar que el correcto uso de big data es de gran importancia, ya que permite a las organizaciones la posibilidad de obtener información relevante que le conceda a la misma una ventaja competitiva. En la actualidad, todo quehacer humano genera datos: a través de Internet, el uso de sensores, aplicaciones desde computadores y dispositivos móviles, transacciones comerciales de cualquier tipo [31]. El volumen de datos es tan grande e inmanejable, que como respuesta han surgido nuevas ciencias cuyo objetivo son los datos, sus características y sus formas de manipulación. Estas ciencias han buscado, para afrontar estas tareas con menor dificultad, definir marcos de trabajo, modelos de madurez y estándares. ¿Pero han logrado crear estas ciencias un modelo

de madurez para la gobernanza de Big Data que sea globalmente aceptado a través de la industria?

Para afrontar esta cuestión, analizamos algunos de los modelos más comentados y completos, además de apoyarnos en el artículo académico de evaluaciones preliminares de Drus y Hassan [32].

El modelo de madurez propuesto por IBM [33] pone particular foco en los procesos, pero dentro de los elementos que utiliza para evaluación deja completamente de lado la capacidad tecnológica e infraestructura de la empresa. Y aunque dentro de estos elementos de evaluación sí hace mención a la cultura de la empresa, se hace muy superficialmente. Tampoco se encontró material de apoyo para la aplicación del mismo, y parecería que el modelo es meramente descriptivo con el único uso en evaluar la situación de la organización.

Continuando con el modelo propuesto por HortonWorks [34] se encuentra nuevamente un foco fuerte en los procesos de la organización. En este caso la capacidad tecnológica e infraestructura sí son consideradas. Lo que este modelo ignora, aunque sí dedica uno de sus dominios en la organización y sus habilidades, es el uso de habilidades blandas para lograr que la adopción de las nuevas metodologías planteadas por el modelo sean aceptadas.

Por último se analiza el modelo que plantea TDWI [35]. Una vez más se halla un estándar con principal foco en los procesos organizativos, mencionando superficialmente los aspectos tecnológicos y apenas comentarios sobre el uso e importancia de habilidades blandas. Vale destacar que es el único modelo encontrado que toma en consideración dentro de sus dimensiones de evaluación el aspecto cultural y político en el cual se halla la organización.

A partir de esta investigación se observó que varios de los modelos en existencia solo se enfocan en habilidades duras y dentro de éstas, solo buscan establecer procesos para que se utilicen de forma correcta sin tomar en consideración los factores humanos. Además también se nota que dentro de las habilidades duras, encontramos poca exploración de las capacidades tecnológicas con la que cuenta la organización, lo cual significó en algunos casos que empresas más pequeñas no podrían cumplir con procesos básicos según estos modelos.

Se concluye entonces que no existe un modelo de madurez de big data que cubra todos los dominios necesarios para lograr un marco de trabajo globalmente aceptado.

## **Pregunta 2.**

### **¿ Existe la necesidad en la industria de un modelo de madurez para la gobernanza de Big Data ?**

En el capítulo 2, Estado del Arte, se explica qué es y los conceptos relacionados con la gobernanza de Big Data, así como la importancia de la misma. Está claro que en esta era, la tecnología de la información es la nueva revolución industrial, y dentro de la misma la Big Data es una de sus ramas con más empuje e importancia. Por esto, toma relevancia una buena gestión y manejo de la Big Data.

Una estrategia exitosa de Big Data se basa en un plan de gobierno de datos bien ejecutado [36]. De acuerdo a Quinto [36] y otros autores [2], [37], [38] y [17] respecto a la importancia de la gobernanza de Big Data y a la propuesta de [8, 24] se desprende la necesidad de un modelo de madurez para el gobierno de Big Data.

En conjunto con la conclusión anterior y para contestar la pregunta de investigación, se centra en cómo ayudaría un modelo de madurez en la gobernanza de Big Data. De este modo, al saber la utilidad del modelo de madurez, quedará definido si es necesario para la industria.

Como primer enfoque general, dicho modelo de madurez permitirá trazar un camino estratégico a las organizaciones en la adopción, evaluación y mejora de las prácticas de gobernanza de Big Data. Es decir permitirá evaluar el grado de madurez en que está la organización en la actualidad en cuanto a la aplicación de buenas prácticas, y dar una guía para poder llegar al grado de madurez deseado.

Bajando más de nivel, se verá a continuación en las diferentes áreas en las que el modelo incide de forma positiva tomando como referencia el marco de Soares [2] y las disciplinas que él propone.

#### **1. Organización**

En esta disciplina se analiza la estructura de la organización en cuanto a la Big Data. Promueve entre otras cosas la definición de :

- Definir el tipo de de Big Data que más aplique para la organización.

- Creación de una matriz RACI para identificar a los stakeholders y el papel que ocupa cada interesado en cuanto al rol y responsabilidades.
- La creación de nuevos cargos específicos o extender ya existentes para la gobernanza de Big Data.
- Incluir en la gobernanza de la información los aspectos de Big Data

## 2. Metadata

Esta disciplina promueve la definición de procesos, métodos y herramientas para crear una semántica común para negocio y TI. Para esto el programa de gobernanza de Big Data debe adoptar algunas buenas prácticas tales como:

- Implementar un glosario que represente las definiciones de negocio.
- Incluir en el glosario metadata técnica relevante para la Big Data y links a fuentes de datos importantes.
- Recopilar metadata de documentos no estructurados.
- Aprovechar metadatos operativos para monitorear los movimientos de la Big Data.

## 3. Privacidad

Esta disciplina contempla lo referente a las políticas, practicas y controles usados por una organización para mitigar el riesgo y proteger los activos de datos. Para esto el programa de gobernanza de Big Data debe adoptar algunas buenas practicas tales como:

- Identificar datos sensibles.
- Incluir datos sensibles en el repositorio de metadata.
- Tener en cuenta las leyes de privacidad y restricciones de la región.
- Monitorear el acceso a datos

#### 4. Calidad de los datos

Esta disciplina contempla lo referente a métodos para medir, mejorar y certificar la calidad e integridad de los datos. Para esto el programa de gobernanza de Big Data debe adoptar algunas buenas prácticas tales como:

- Establecer con los diferentes interesados métricas y medidas para determinar la calidad de datos
- Nombrar administradores de datos responsables ante el consejo de gobierno de la información para mejorar las métricas a lo largo del tiempo

#### 5. Integración de proceso de negocio

Esta disciplina considera la importancia de integrar la gestión de proceso de negocio y la gobernanza de Big Data. Para esto el programa debe adoptar algunas buenas prácticas tales como:

- Identificar los procesos clave que impactarán en la gobernanza de Big Data
- Construir un mapa de proceso con las actividades clave
- Mapear las políticas de gobierno de big data a las actividades clave en el proceso.

#### 6. Integración de datos maestros.

Esta disciplina considera la importancia de los datos maestros y la integración con la Big Data. Para esto el programa debe adoptar algunas buenas prácticas tales como:

- Mejorar la calidad de los datos maestros para soportar la analítica de Big Data
- Aprovechar la Big Data para mejorar la calidad de los datos maestros
- Extraer información de los datos no estructurados para enriquecer los datos maestros.

#### 7. Gestión del ciclo de vida de la información.

Esta disciplina promueve un enfoque sistemático basado en políticas para la recopilación, el uso, la retención y la eliminación de información. De este modo se gestiona la información durante todo su ciclo de vida.

Dado este análisis queda claro el impacto positivo de un modelo de madurez y la necesidad de la industria de contar con el mismo para la gobernanza eficiente de la Big Data.

### **Pregunta 3.**

**¿ Es posible construir un modelo de madurez de este tipo ?**

Luego de la investigación y analizando los resultados, se ve claramente que el área de la gobernanza de Big Data está aún un poco en pañales. Si bien se sabe de la importancia de la misma, hasta el momento no hay un modelo de madurez estándar que facilite su implementación. Esto se debe a que los impulsos por lograrlo son individuales. Por ejemplo Soares [2] y Stanford [15 ], que fueron parte de los modelos que tomamos como referencia, se basan en el modelo de IBM [33], que se desarrolló por una necesidad de negocio.

Se cree que por ser un área compleja con mucha dependencia del contexto, tipo de Big Data e industrias, se hace difícil un consenso. De todos modos se entiende que existen las bases de conocimiento y expertise como para poder lograr un estándar.

La respuesta final a esta interrogante se va a evacuar en el siguiente capítulo con el desarrollo del modelo de madurez.

## **4. Desarrollo del modelo de madurez.**

### **4.1. Definición del alcance del modelo de madurez**

Para comenzar con la construcción del modelo se debe definir el alcance del mismo. Con la intención de dimensionar el alcance total, se analizaron los diferentes marcos existentes de gobernanza de datos y de Big Data, tales como los de Soares [2], Stanford [15] e IBM [33].

Cada uno de ellos toma diferentes dimensiones, contextos, áreas y componentes que pueden ser en mayor o menor medida aplicables a la realidad de una organización. Debido a la amplitud y al tiempo limitado de este proyecto nos vamos a enfocar, como alcance, en uno de los marcos en particular.

Se tomará como referencia el marco de gobernanza de Big Data de Soares [2], visto en el capítulo 2, Estado del arte, en el cual se proponen tres dimensiones. Dichas dimensiones son tipo de Big Data, disciplinas de la gobernanza de la información e industrias y funciones. Con estos tres elementos conforma un marco completo aplicable a cualquier contexto y tiene en cuenta las diferencias que surgen de los distintos tipos de Big Data e industria.

No obstante el modelo se podría aplicar de forma genérica ya que el esqueleto central del marco aplica para cualquier contexto.

### **4.2. Descripción del modelo de evaluación insumo**

Como ya se mencionó anteriormente, este trabajo toma como insumo un proyecto anterior titulado “Modelo para la evaluación de buenas prácticas de gobernanza de Big Data” desarrollado por Darío Carlos y Mauro Nascimento [3], el cual define un modelo de evaluación de buenas prácticas en la gobernanza de Big Data, ver **Anexo 1**.

A continuación se dará una breve descripción del modelo de evaluación para dar contexto y base al desarrollo del modelo de madurez de este proyecto.

Para la construcción del modelo de evaluación insumo [3] tomaron como base el marco de gobernanza de Big Data de Soares [2], visto en el capítulo Estado del arte.

Del mencionado marco utilizaron el total de disciplinas propuestas por Soares :

- Organización

- Metadata
- Privacidad
- Calidad de datos
- Integración de procesos de negocio
- Integración de datos maestros
- Ciclo de vida de la información

A estas disciplinas se les agregó “*Capacidad tecnológica*”, que no era tomada en cuenta por Soares, y que se explicará al final de este capítulo citando su definición del proyecto base [3].

Para cada una de las disciplinas mencionadas anteriormente se desarrollaron preguntas agrupándolas en habilidades duras y blandas, estas habilidades están definidas en el capítulo 2. Para el desarrollo del set de preguntas sobre habilidades duras, el proyecto insumo [3], hace referencia a un conjunto de interrogantes propuestas por Soares en su libro [2], las cuales tienen como objetivo ayudar a evaluar el nivel de madurez del gobierno de Big data, y añaden otras.

En el caso de las preguntas sobre habilidades blandas, tomaron cuatro áreas que tienen su foco conceptual en Smits y Hillegersberg [8], dichas áreas son Participación, Liderazgo, Mejora continua y Entendimiento y confianza, definidas también en el capítulo 2.

Luego de definido el modelo de evaluación, plantean un modo de evaluarlo, ver **Anexo 1**. Dicho modelo utiliza una escala para responder, que va de 1 a 5 y tiene el siguiente significado:

1. Total desacuerdo, inadecuación, insatisfacción, nunca, etc.
2. Desacuerdo, inadecuación, insatisfacción, casi nunca, etc.
3. Neutro, indiferente
4. Acuerdo, adecuación, satisfacción, casi siempre, etc.
5. Total acuerdo, adecuación, satisfacción, siempre, etc.

De este modo, asignando una puntuación de conformidad con las preguntas, sacando totales y promedio, obtienen un indicador general de la evaluación.

A continuación se citan de [3] la definición y características de la disciplina Capacidad

### ***“Dimensión Capacidad tecnológica***

*La disciplina capacidad tecnológica se agrega al modelo para la evaluación de buenas prácticas de gobernanza de Big Data, debido a que el software de Big Data, como otros, requiere de tareas básicas como lo son:*

- *Instalación*
- *Configuración inicial*
- *Mantenimiento preventivo*
- *Análisis de mejora*
- *Reparación de errores y defectos*

*A su vez, debe trabajar sobre una infraestructura de hardware, la cual tiene cierta complejidad debido al volumen de información que recibe y procesa. Para suplir los requerimientos de estos sistemas se debe cumplir algunos lineamientos:*

- *Alta disponibilidad de sus componentes*
- *Tolerancia a fallos*
- *Redundancia del dato*
- *Manejo de versiones*
- *Administración de cargas*

*Esta dimensión está constituida por el conjunto de conocimientos y habilidades que dan sustento al proceso de implementación, configuración y administración de las soluciones de Big Data. Desde la tecnología, hardware y software, como los recursos que la administran.*

*Las diferentes soluciones de Big Data, sea cual fuere su propósito de desarrollo, no serían posibles sin estos componentes y sin que estos tuvieran un funcionamiento correcto.*

*En el caso del hardware, alta disponibilidad, continuidad del servicio y la redundancia de sus componentes serían aspectos a buscar. Esto depende directamente del hardware con el que se cuente y/o la tecnología que este utilice. Estos aspectos deben ser considerados a la hora de realizar la proyección de la implementación de un sistema de Big Data debido ya que, de no realizarse un correcto dimensionamiento y un análisis de posibles puntos únicos de fallo, se puede incurrir en una falta que puede traer consecuencias no deseadas a futuro (poco cómputo para procesamiento, espacio insuficiente de almacenamiento, denegación de servicio por falla de algún componente de hardware, etc.).*

*Con respecto al software base, se buscaría que el mismo tenga instaladas las actualizaciones concernientes al sistema operativo y a seguridad y se resuelvan advertencias o errores que puedan ir surgiendo.*

*El software propio de Big Data puede requerir de manejo de versiones, ciertas depuraciones, administración de fuentes, entre otras tareas. Las cuales deben realizar recursos humanos debidamente capacitados, ya que de su desempeño va a depender el correcto funcionamiento y salud de estos componentes.*

*Por las razones mencionadas esta disciplina toma suma importancia y se agrega al modelo confeccionado. ”*

### **4.3. Análisis crítico del modelo de evaluación insumo**

Se comenzará el camino de construcción de nuestro modelo de madurez haciendo una crítica a dicho modelo de evaluación [3]. De este modo intentaremos analizar su contenido y estructura, para poder mejorarlo, adaptarlo y llegar a una evaluación más simple, manteniendo su cometido y adicionando lo que creamos correspondiente.

Luego del análisis del proyecto [3], surge como primer punto a mejorar la extensión de la evaluación. Se cree que la cantidad de preguntas es excesiva. Esta crítica se basa en los resultados de validación en cuanto a la cantidad de respuestas que se obtuvo. De un total de 35 expertos a los que se le envió la evaluación, solo 6 respondieron. Uno de los factores que puede haber incidido es la extensión de la misma.

Como segundo punto a mejorar está la estructura y el propósito de la evaluación. El análisis de este aspecto se apoya en el artículo sobre desarrollo de modelos de madurez García y Félix

[19], que definen los elementos de la arquitectura de un modelo. Entre estos elementos está la definición de niveles y sus características, para mediante estos, poder diagnosticar el estado actual en referencia a este modelo. Es decir si se cumplen con determinadas características que marca el nivel, se cumple con este.

Debido a esto, se tendrá que migrar del modo de evaluación con preguntas a un modo de aseveraciones, para poder definir las características del nivel. Esto significa un cambio importante pero necesario para la construcción del modelo, de todos modos se intentara mantener lo sustancial del contenido.

Como tercer punto se va a modificar una de las áreas dentro de las habilidades soft o blandas. Las definiciones de niveles de madurez que plantean en sus respectivos modelos, CMMI [13] y Stanford [15], manejan un último nivel que es de optimización. Este nivel tiene foco en la optimización y la mejora continua, por esto se eliminará de las habilidades soft el área de la Mejora continua, y se tendrá en cuenta en este último nivel.

Por último, se dividirá el modelo en dos capacidades hard y soft, pero que se evaluarán por separado, a diferencia del modelo de evaluación insumo que los unía dentro de cada disciplina.

#### **4.4. Modelo de madurez para la gobernanza de Big Data.**

##### **4.4.1. Construcción del modelo de madurez**

El inicio de la construcción del modelo de madurez se apoya en el artículo de García y Félix [19], el cual describe como implementarlo, desde la metodología, hasta su arquitectura.

Dentro de la metodología, propone una serie de etapas para la implementación. Estas etapas están conformadas por, estado del problema u objetivo, revisión de las soluciones existentes, si existe el modelo adaptarlo, si no existe el modelo, desarrollarlo. En la Ilustración 12 se muestra el proceso completo.

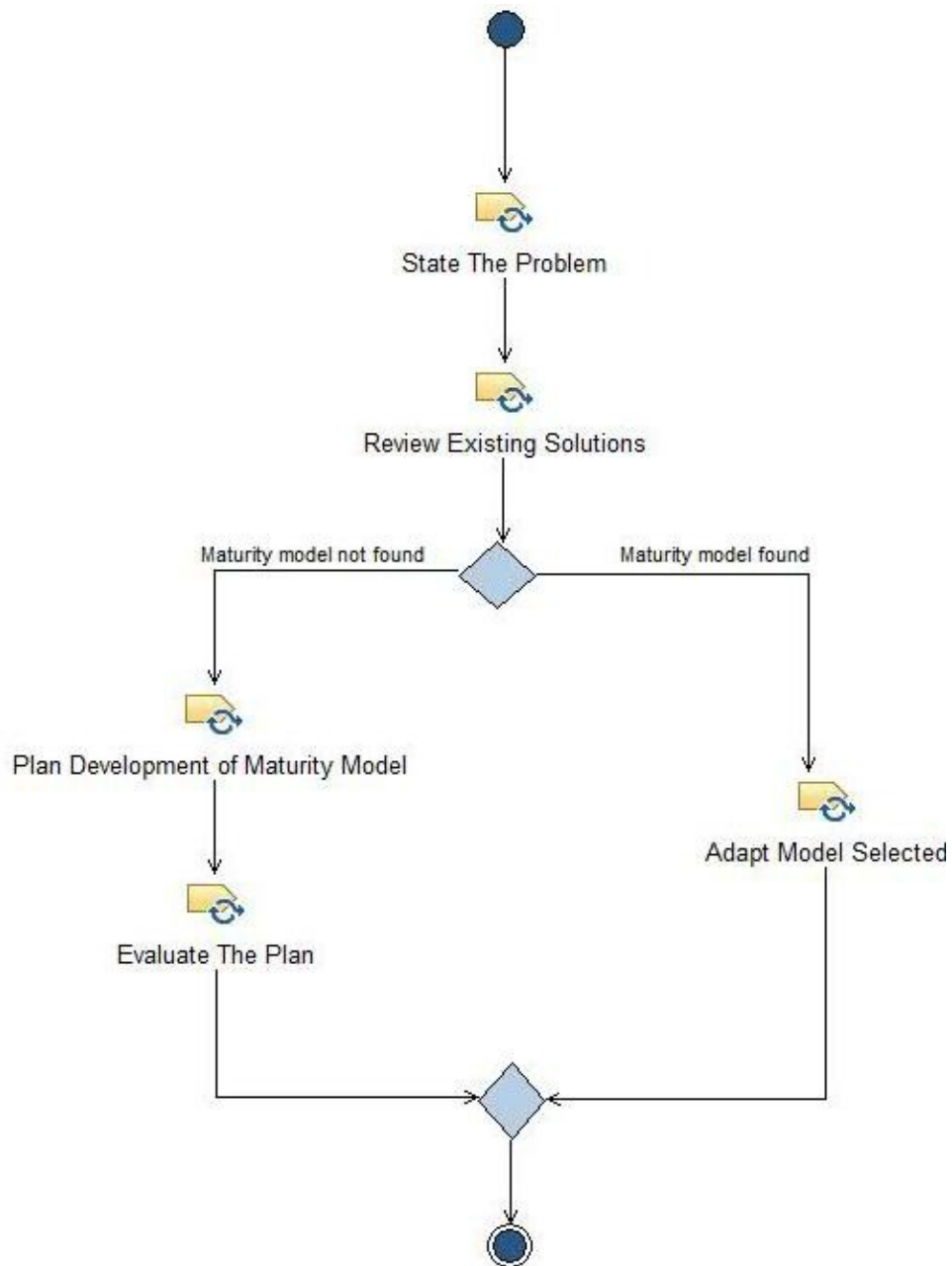


Ilustración 12. Diagrama de actividades para la construcción de un modelo [19]

Por otra parte definen [19] los elementos fundamentales para el desarrollo de modelos de madurez, y por tanto la arquitectura del mismo. Los elementos que promueven son, establecer objetivos para el desarrollo del modelo de madurez, diseñar una arquitectura de modelo, establecer niveles y dimensiones de capacidad.

En cuanto los objetivos, los mismos están claros, y son el desarrollo de un modelo de madurez para la gobernanza de Big Data.

Para establecer la arquitectura se debe comenzar por definir la estructura y sus niveles. Con esta meta se analizan varios libros y artículos para cimentar una base sólida para el modelo del presente proyecto. De toda la literatura se destacarán las que contribuyeron con este modelo.

Se comienza por analizar los artículos de Aguiar, Pereira, Braga y Bianchi [16] y Proença y Borbinha [17], donde comparan los marcos y modelos de madurez más difundidos en la industria, ITIL, COBIT y CMMI [13] entre muchos otros. De este modo se obtuvo una noción clara de las diferentes arquitecturas de modelos más utilizadas en el mercado.

Por otro lado, se adentra en la gobernanza de datos, para esto se investigan varios modelos de madurez, entre los más importantes están los de Soares [2], Stanford [15] e IBM [33].

Finalmente de la conjunción de los modelos de CMMI [13], Soares [2], Stanford [15] e IBM [35], más la impronta propia, se obtuvieron los niveles de madurez y sus características. Dichos niveles se muestran en la Ilustración 13 y en el próximo capítulo se detallan con mayor exactitud.

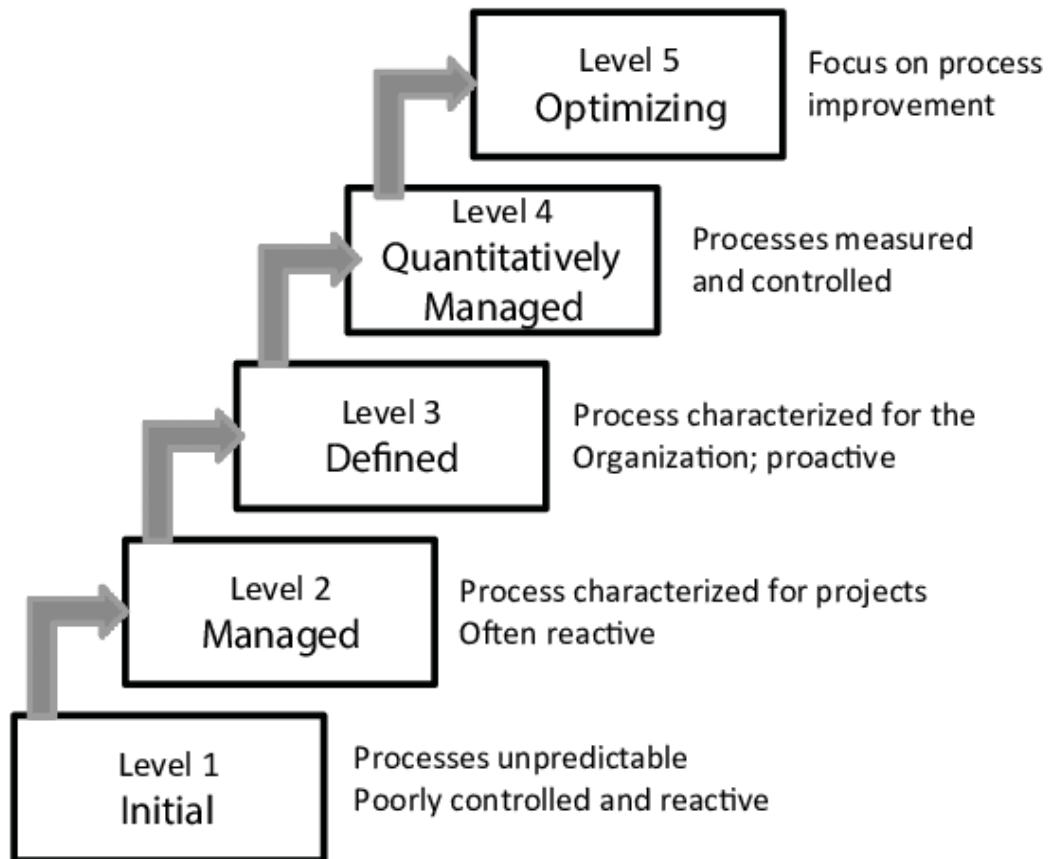


Ilustración 13. Niveles de madurez

Para establecer el siguiente elemento de la arquitectura del modelo propuesto por García y Félix [19], que son las dimensiones de capacidad, se tomará como base las disciplinas (dimensiones) que promueve Soares [2], vistas en capítulo 2, El estado del arte. Soares define siete disciplinas para la gobernanza de Big Data y complementa su marco con los tipos de Big Data e industrias. Serán seleccionadas dichas disciplinas para el modelo del presente proyecto, estas fueron vistas en el capítulo 2, y serán presentadas en el siguiente capítulo. En la siguiente Ilustración 14 se muestra el marco propuesto por Soares con sus componentes.

## Big Data Governance Framework

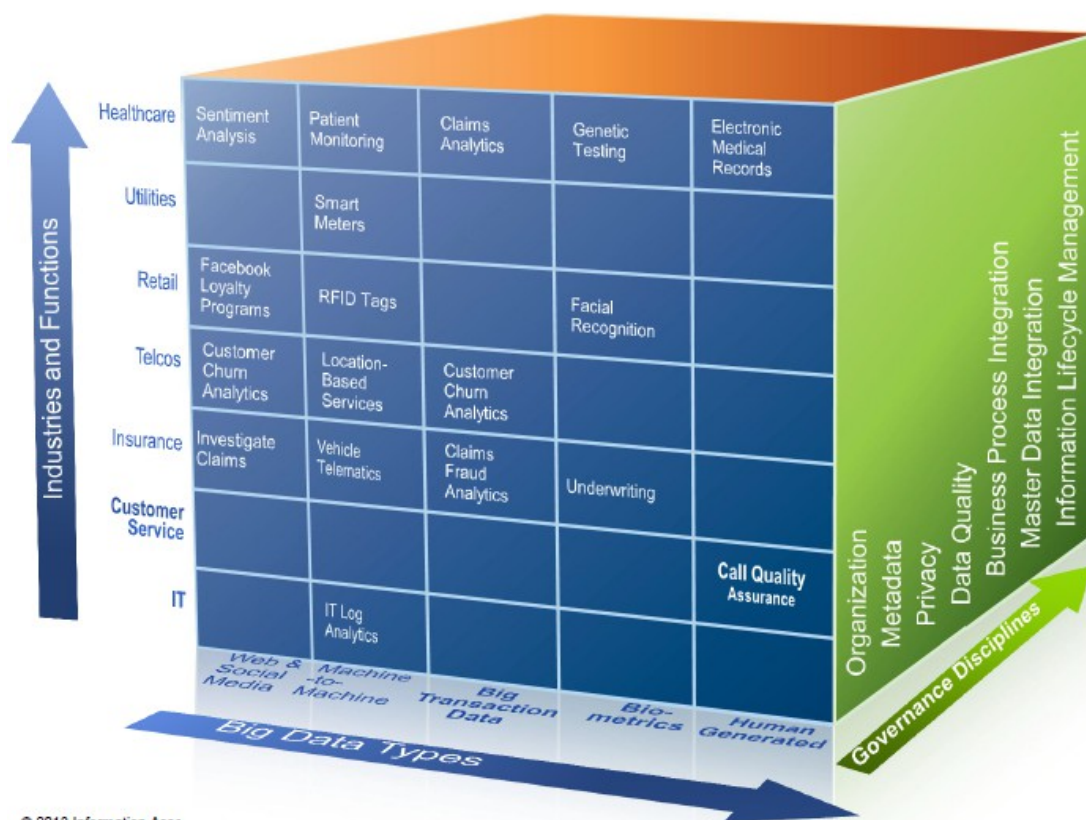


Ilustración 14. Framework para gobernanza de Big Data [2]

A las disciplinas base de Soares se le agregara la “Capacidad tecnológica”, ya definida en el capítulo 4.2, Descripción del modelo de evaluación insumo, que no es tenido en cuenta por el mismo y que se toma del proyecto insumo [3].

Para completar se definirán dos áreas a analizar, las mismas son las habilidades hard y soft, definidas en el capítulo Estado del arte, y extraídas del proyecto insumo [3].

Dentro del área hard se darán un conjunto de características para cada disciplina, por nivel de madurez. Para establecer dichas características hard se toma como base el proyecto insumo [3], el modelo de madurez de gobernanza de datos de Stanford [15], CMMI [13] e impronta propia.

Dentro del área soft, del mismo modo, se darán un conjunto de características por nivel, es decir, el nivel no se definirá por cada disciplina, sino que por el total de ellas. Para establecer dichas características soft se toma como base el proyecto insumo [3], y las planteadas por Smits y Hillegersberg [24].

Por último se adaptan la metodología de Stanford [15] e IBM [33] para determinar en qué nivel de madurez está la organización, y para la planificación de la estrategia a seguir buscando la mejora de dentro de los niveles de madurez del modelo.

#### **4.4.2. Descripción del modelo de madurez**

Este modelo de madurez es una herramienta que permitirá trazar un camino estratégico a las organizaciones en la evaluación, adopción y mejora de las prácticas de gobernanza de Big Data.

Para cumplir con este cometido se utilizará como herramienta un modelo de evaluación. Dicho modelo definirá por un lado, en qué nivel de madurez está actualmente la organización, y por otro lado también hará las veces de guía para lograr la adopción progresiva de buenas prácticas de gobernanza de Big Data. De este modo la organización irá ascendiendo de acuerdo a su planificación hasta llegar al nivel de madurez más alto.

El modelo de evaluación se divide en dos capacidades, la primera de ellas que analiza los aspectos más técnicos será llamada Habilidades Hard, y la segunda capacidad la cual se refiere a capacidades blandas tales como Participación, Liderazgo, Entendimiento y confianza, será llamada Habilidades Soft.

Dentro de las Habilidades Hard se evaluarán ocho disciplinas de la gobernanza de Big Data, Organización, Metadata, Privacidad, Calidad de datos, Integración de procesos de negocio, Integración de datos maestros, Ciclo de vida de la información y Capacidad tecnológica. A cada una de estas se les asignará determinado nivel de madurez de acuerdo a ciertas características cualitativas.

Luego de obtenido el nivel de cada una de las disciplinas se suman sus valores y se dividen por la cantidad total de disciplinas, dicho promedio dará el nivel de madurez hard de la organización.

Las Habilidades Soft se evaluarán del mismo modo, pero a nivel de todas las disciplina en conjunto. Es decir, de acuerdo a las características cualitativas que cumpla la organización se le dará un determinado nivel de madurez soft.

A continuación se describen brevemente los niveles de madurez con sus características, y las diferentes disciplinas de la gobernanza de Big Data.

### **4.4.3. Niveles de madurez**

#### **1. Inicial**

Los procesos suelen ser ad hoc y el entorno no es estable. El éxito refleja la competencia de los individuos dentro de la organización, en lugar del uso de procesos comprobados. Si bien las organizaciones de nivel de madurez uno a menudo producen productos y servicios que funcionan, con frecuencia superan el presupuesto y el cronograma de sus proyectos. Estas organizaciones se caracterizan por una tendencia a abandonar los procesos en el momento de la crisis, y no ser capaz de repetir sus éxitos pasados

#### **2. Administrado**

Los éxitos son repetibles, pero los procesos pueden no repetirse para todos los proyectos en la organización. La administración básica del proyecto ayuda a rastrear los costos y los cronogramas, mientras que la disciplina del proceso ayuda a garantizar que se conserven las prácticas existentes. Cuando se implementan estas prácticas, los proyectos se realizan y administran de acuerdo con sus planes documentados, sin embargo, aún existe el riesgo de exceder las estimaciones de costo y tiempo.

#### **3. Definido**

El conjunto de procesos estándar de la organización se utiliza para establecer la coherencia en toda la organización. Los estándares, las descripciones de procesos y los procedimientos para un proyecto se adaptan a partir del conjunto de procesos estándar de la organización para adaptarse a un proyecto particular o unidad organizativa.

#### **4. Administrado cuantitativamente**

Las organizaciones establecen objetivos de calidad cuantitativos tanto para el proceso como para el mantenimiento. Los subprocesos seleccionados contribuyen significativamente al rendimiento general del proceso y se controlan mediante técnicas estadísticas y otras técnicas cuantitativas.

#### **5. Optimización**

Los objetivos cuantitativos de mejora de procesos para la organización están firmemente establecidos y se revisan continuamente para reflejar los cambios en los objetivos de negocios, estos se utilizan como criterios para gestionar la mejora continua de procesos.

#### **4.4.4. Disciplinas**

##### **1. Organización**

En esta disciplina se analiza la estructura de la organización con respecto a la Big Data buscando tener los procedimientos y roles necesarios para soportar con éxito a la misma.

##### **2. Metadata**

Esta disciplina promueve la definición de procesos, métodos y herramientas para crear una semántica común para negocio y TI.

##### **3. Privacidad**

Esta disciplina contempla lo referente a las políticas, practicas y controles usados por una organización para mitigar el riesgo y proteger los activos de datos.

##### **4. Calidad de los datos**

Esta disciplina contempla lo referente a métodos para medir, mejorar y certificar la calidad e integridad de los datos.

##### **5. Integración de proceso de negocio**

Esta disciplina considera la importancia de integrar la gestión de proceso de negocio y la gobernanza de Big Data.

##### **6. Integración de datos maestros**

Esta disciplina considera la importancia de los datos maestros y la integración con la Big Data.

##### **7. Gestión del ciclo de vida de la información**

Esta disciplina promueve un enfoque sistemático basado en políticas para la recopilación, el uso, la retención y la eliminación de información.

## 8. Capacidad tecnológica

Esta disciplina refiere al entrono de infraestructura y tecnología necesario para llevar adelante con éxito la Big Data.

### 4.4.5. Evaluación de Habilidades Hard

A continuación se detalla el modelo de evaluación de habilidades hard. Por cada disciplina de la gobernanza de Big Data se muestra un cuadro donde se describen las características cualitativas propias de cada nivel de madurez.

<b>ORGANIZACIÓN</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>1</b>	<p>No están definidos o no se es cociente de:</p> <ul style="list-style-type: none"><li>- Las disciplinas de la gobernanza de Big Data</li><li>- Los roles relacionados con la Big Data</li><li>- La documentación de políticas, procesos, procedimientos formales de Big Data.</li></ul> <p>Los requisitos de datos son impulsados de forma individual.</p>
<b>2</b>	<p>Se han definido las disciplinas de la gobernanza de Big Data</p> <p>Se han definido, pero solo para algunos proyectos o disciplinas:</p> <ul style="list-style-type: none"><li>- Los roles relacionados con la Big Data</li><li>- Las políticas, procesos, procedimientos formales de Big Data.</li></ul> <p>La gestión es reactiva.</p>
<b>3</b>	<p>Las políticas, procesos, procedimientos de la gobernanza de Big Data están estandarizados, documentados y accesibles a nivel de la organización.</p> <p>Hay una definición de arquitectura empresarial incipiente.</p> <p>Se ha definido una matriz RACI para cada proceso identificado, aunque aun no se cumplen con</p>

<b>ORGANIZACIÓN</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
	<p>todos los roles. También existe todavía algún rol funcionalmente aislado.</p> <p>La gestión es proactiva.</p>
<b>4</b>	<p>Existe un marco de gobernanza de datos en la organización, la Big Data se ha integrado ha dicho marco, y se tiene claro su rol dentro de la gobernanza de la organización .</p> <p>La arquitectura empresarial esta claramente definida.</p> <p>Los roles están totalmente claros e integrados, cada rol cumple su función y sabe como colabora en los objetivos de Big Data.</p> <p>Se planifica el trabajo a realizar en Big Data.</p> <p>Se evaluá y controla el cumplimiento de todas las políticas, procesos, procedimientos mediante técnicas estadísticas y otras técnicas cuantitativas.</p>
<b>5</b>	<p>Se evalúa periódicamente si las políticas, procesos, procedimientos y roles siguen siendo compatibles y efectivos con las premisas u objetivos.</p> <p>Se toman acciones correctivas sobre las falencias encontradas en la evaluación.</p> <p>La arquitectura empresarial responde a los objetivos estratégicos de la empresa.</p> <p>Se investigan, analizan e incorporan nuevas tendencias en la gobernanza de Big Data.</p>

<b>METADATA</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>1</b>	<p>No existe una comprensión del valor de los metadatos.</p> <p>No existen o son inconsistentes las políticas, procesos, procedimientos o roles relacionados con la metadata</p> <p>No existen un glosario de datos definido o si existe es incompleto y de baja calidad.</p>
<b>2</b>	<p>Se definen algunas políticas, procesos, procedimientos a nivel de proyecto.</p> <p>Se han definido roles responsables de la metadata de la Big Data .</p> <p>Se han identificado las fuentes de Big Data de datos operacionales estructurados.</p> <p>Se ha definido en un glosario el significado en términos técnicos de la Big Data para datos estructurados.</p> <p>El glosario esta consolidado y disponible para algunos interesados.</p>
<b>3</b>	<p>Están instauradas y documentadas las políticas, procesos y procedimientos para la generación de la nueva metadata y mantenimiento de la existente.</p> <p>Hay una definición de arquitectura de datos incipiente.</p> <p>Los participantes comprenden claramente las responsabilidades asociadas con sus roles.</p> <p>La recopilación de metadatos sobre contenido estructurado se automatiza.</p> <p>Las políticas de metadata en torno al gobierno de Big Data están disponibles para todos los interesados.</p>
<b>4</b>	<p>La arquitectura de datos esta claramente definida.</p> <p>Se identifican, almacenan e incorporan al glosario de datos, los datos no estructurados.</p> <p>Las auditorías y validación regular de metadatos se adoptan como políticas oficiales de datos de la organización.</p> <p>Está sincronizada y centraliza la metadata perteneciente a los distintos repositorios existentes de Big Data.</p> <p>Los metadatos se recopilan automáticamente de la mayoría de los sistemas de gestión de bases de datos relacionales y sistemas externos.</p> <p>Se ha definido en un glosario el significado en términos del negocio de la Big Data</p>

<b>METADATA</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>5</b>	<p>La arquitectura de datos responde a los objetivos estratégicos de la empresa.</p> <p>Se crea un grupo de gestión de metadatos dedicado para avanzar estratégicamente en las capacidades de metadatos y aprovechar de manera más efectiva los metadatos existentes.</p> <p>Las políticas, proceso y procedimientos de metadatos cubre datos estructurados y no estructurados</p> <p>Se evalúa periódicamente si las políticas, procesos, procedimientos y roles siguen siendo compatibles y efectivos con las premisas u objetivos.</p> <p>Se toman acciones correctivas sobre las falencias encontradas en la evaluación.</p>

<b>PRIVACIDAD</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>1</b>	<p>No existen políticas, practicas o controles para proteger los activos de datos</p> <p>No están definidos los roles relacionados con la privacidad</p> <p>No se han identificado los datos sensibles.</p> <p>No se tienen en cuenta las leyes de privacidad y restricciones de la región.</p> <p>No se controla el acceso a datos.</p>
<b>2</b>	<p>Se definen algunas políticas, procesos, procedimientos a nivel de proyecto.</p> <p>Se han definido roles responsables de la privacidad de activos de datos.</p> <p>Se ha identificado la Big Data que contiene datos sensibles.</p> <p>Se registra el acceso a la Big Data</p>
<b>3</b>	<p>Están instauradas y documentadas las políticas, proceso y procedimientos entorno a la privacidad de activos de datos, y se distribuyen como mejores prácticas.</p> <p>Los participantes comprenden claramente las responsabilidades asociadas con sus roles.</p> <p>Se ha marcado en el repositorio de metadata de Big Data los datos sensibles identificados.</p>

<b>PRIVACIDAD</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
	<p>Se comienza a tener en cuenta las leyes de privacidad y restricciones de la región</p> <p>Se controla el acceso a datos.</p>
<b>4</b>	<p>Las políticas de privacidad de activos de datos se convierten en políticas oficiales de la organización y se audita el cumplimiento de las mismas.</p> <p>Se han tomado en cuenta las leyes regionales e internacionales que afectan el uso de la Big Data</p> <p>Se monitorea y audita el acceso a datos.</p> <p>Los niveles de privacidad y seguridad son establecidos por el grupo de manejo de Big Data y por el directorio</p> <p>Los cambios realizados en niveles de privacidad y seguridad son formalmente finalizados y comunicados al todos los interesados.</p>
<b>5</b>	<p>Se monitorea y analiza de manera periódica para encontrar nuevas amenazas y su posible impacto en la privacidad de los activos de datos.</p> <p>Se utilizan indicadores para dimensionar y dar forma a la mejora continua.</p> <p>Se evalúa periódicamente si las políticas, procesos, procedimientos y roles siguen siendo compatibles y efectivos con las premisas u objetivos.</p> <p>Se toman acciones correctivas sobre las falencias encontradas en la evaluación.</p>

<b>CALIDAD DE DATOS</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>1</b>	<p>No existen políticas, practicas o procedimientos para la gestión de calidad de datos.</p> <p>La calidad de los datos no esta formalmente definida, es decir no esta claro el concepto de calidad de datos que tiene la organización.</p> <p>No están definidos los roles relacionados con la calidad de datos.</p> <p>Las personas realizan esfuerzos, poco frecuente, de calidad de datos, comúnmente ad hoc, según sea necesario y corrigen manualmente los problemas de datos identificados.</p>
	<p>Las políticas, practicas o procedimientos se han definido para algunas actividades relacionadas con la calidad de los datos y se han seguido de manera inconsistente.</p>

<b>CALIDAD DE DATOS</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>2</b>	<p>Un pequeño grupo de individuos es capacitado y realiza perfiles para evaluar la calidad de los datos , para luego establecer una línea de base a ser utilizada en proyecto.</p> <p>Se han identificado los datos críticos que aseguran el éxito del programa de Big Data en la organización</p> <p>Los perfiles definidos para la calidad de datos están disponibles para su uso en cualquier parte del ciclo de vida de la Big Data.</p>
<b>3</b>	<p>Están instaurado y documentados las políticas, practicas o procedimientos sobre perfiles de calidad de datos y el desarrollo de estándares.</p> <p>Se han definido todos los roles para garantizar la calidad de los datos y sus responsabilidades están claras.</p> <p>Se definen métricas de calidad e informes de calidad.</p> <p>La calidad de datos se convierte un actividad programada en los planes del proyecto.</p>
<b>4</b>	<p>Las políticas de calidad de datos se convierten en políticas oficiales de la organización y se audita el cumplimiento de las mismas.</p> <p>Los expertos en calidad de datos se identifican en toda la organización y participan en todos los proyectos de mejora de la calidad de datos.</p> <p>Para los datos críticos identificados se ha definido la matriz de calidad de datos <i>Elemento de dato - A qué y cómo afecta la falta de calidad - Responsable</i></p> <p>Se controlan los resultados de las métricas de calidad.</p> <p>La solución de problemas de calidad de datos está integrada en la plataforma de informes de calidad.</p>
<b>5</b>	<p>Se evalúa periódicamente si las políticas, procesos, procedimientos y roles siguen siendo compatibles y efectivos con las premisas u objetivos.</p> <p>Se utilizan indicadores para dimensionar y dar forma a la mejora continua.</p> <p>Se realiza un seguimiento y se informa de forma centralizada el cumplimiento de la calidad de los datos oficiales de la organización.</p> <p>Las correcciones de la calidad de los datos se implementa tanto en datos en reposo (en bases de datos) como en datos en vuelo (en ETL y como mensajes entre sistemas).</p>

<b>INTEGRACIÓN DE PROCESO DE NEGOCIO</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>1</b>	<p>No existen políticas, practicas o procedimientos para la integración de los procesos de negocios con la gobernanza de Big Data.</p> <p>No se tiene conciencia de la importancia de alinear los proceso de negocio con la gobernanza de Big Data</p> <p>No se han identificado los procesos de negocio que se verán afectados por el gobierno de Big Data.</p> <p>No están definidos los roles relacionados con la integración de los procesos de negocio en con la gobernanza de Big Data.</p>
<b>2</b>	<p>Se han identificado a los stakeholders del negocio para el programa de gobierno de Big Data.</p> <p>Se han definido roles responsables de la integración de proceso de negocio.</p> <p>Se han identificado algunos procesos de negocio que se verán afectados por el gobierno de Big Data.</p> <p>Se definen algunas políticas, procesos, procedimientos para algún proceso de negocio.</p>
<b>3</b>	<p>Están estandarizadas y documentados las políticas, practicas o procedimientos que contribuyen con la integración delos procesos de negocio.</p> <p>Hay una definición de arquitectura de procesos de negocio incipiente.</p> <p>Se han definido todos los roles para garantizar la integración y sus responsabilidades están claras.</p> <p>Se ha construido un mapa de procesos con las actividades clave para cada uno de los mismos.</p> <p>Se ha establecido la relación entre las políticas del gobierno de Big Data con las actividades en los procesos mapeados.</p>
<b>4</b>	<p>Las políticas de integración de procesos se convierten en políticas oficiales de la organización y se audita el cumplimiento de las mismas.</p> <p>La arquitectura de procesos de negocio esta claramente definida y controla su alineación con las arquitecturas de datos y tecnología.</p> <p>Se han cuantificado los beneficios financieros de la gestión de Big Data.</p> <p>Se monitorea y audita los beneficios financieros como indicador de la integración de proceso de negocio a la gobernanza de Big Data.</p>

<b>INTEGRACIÓN DE PROCESO DE NEGOCIO</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>5</b>	<p>Se evalúa periódicamente si las políticas, procesos, procedimientos y roles siguen siendo compatibles y efectivos con las premisas u objetivos.</p> <p>Se utilizan indicadores para dimensionar y dar forma a la mejora continua.</p> <p>La arquitecturas de proceso de negocios, datos y tecnología se corresponden entre si y están afiliadas a los objetivos estratégicos de la empresa.</p> <p>Están totalmente alineadas IT y la Big Data con los objetivos estratégicos de la organización</p>

<b>INTEGRACIÓN DE DATOS MAESTROS</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>1</b>	<p>No existen políticas, practicas o procedimientos para la integración de datos maestros con la Big Data .</p> <p>Comprensión inconsistente de conceptos y beneficios de gestión de datos maestros.</p> <p>Existe una gestión limitada de los datos maestros.</p> <p>No están definidos los roles relacionados con la integración de datos maestros con la Big Data .</p> <p>No hay definiciones formales de qué datos se consideran datos maestros institucionales.</p>
<b>2</b>	<p>Se identifican roles específicos para datos maestros y desarrollan una definición básica y un modelo de datos maestros.</p> <p>Se definen dominios institucionales de datos maestros y se documentan los sistemas que almacenan datos maestros.</p> <p>Los datos maestros se identifican, gestionan y aprovisionan manualmente a través de extractos, transferencias de archivos o cargas manuales.</p>
<b>3</b>	<p>Están estandarizadas las políticas, practicas o procedimientos sobre integración de datos maestros.</p> <p>Se han definido todos los roles responsables de la integración de datos maestros.</p> <p>Se ha mejorado la calidad de los datos maestros para soportar la analítica de Big Data.</p>

<b>INTEGRACIÓN DE DATOS MAESTROS</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
	<p>Se utiliza Big Data para mejorar la calidad de los datos maestros.</p> <p>Están aprobados los modelos de datos maestros.</p>
<b>4</b>	<p>Se audita el cumplimiento de las políticas y estándares de uso de datos maestros.</p> <p>Se ha establecido un programa de mejora de la calidad y consistencia de las claves (keys) para dar soporte al programa de gobierno de Big Data.</p> <p>Se ha enriquecido los datos maestros a través de procesos de extracción en texto u otros datos no estructurados.</p> <p>Se han considerado las políticas vigentes en las distintas redes sociales y su impacto en la integración con los datos maestros</p>
<b>5</b>	<p>Se evalúa periódicamente si las políticas, procesos, procedimientos y roles siguen siendo compatibles y efectivos con las premisas u objetivos.</p> <p>El hub de datos maestros multidominio maneja todo el aprovisionamiento y la administración de datos maestros.</p> <p>EL consejo de administración de datos maestros asumen la responsabilidad de hacer cumplir las políticas de datos maestros en toda la organización.</p> <p>Se utilizan indicadores para dimensionar y dar forma a la mejora continua.</p>

<b>CICLO DE VIDA DE LA INFORMACIÓN</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>1</b>	<p>No existen políticas, practicas o procedimientos relacionados con el ciclo de vida de la información.</p> <p>No se tiene conciencia de la importancia de gestionar el ciclo de vida de la información en cuanto a costos, reglamentación legal y requerimientos del negocio.</p> <p>No están definidos los roles relacionados gestión del ciclo de vida de la información.</p>
<b>2</b>	<p>Se han definido roles responsables del ciclo de vida de la información.</p> <p>Se definen algunas políticas para la recopilación, el uso, la retención y la eliminación de</p>

<b>CICLO DE VIDA DE LA INFORMACIÓN</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
	<p>información.</p> <p>Se comprende la importancia del manejo del ciclo de vida de la información en cuanto a el costo de almacenamiento y los requerimientos de negocio.</p>
<b>3</b>	<p>Están estandarizadas y documentados las políticas, practicas o procedimientos que contribuyen a la gestión del ciclo de vida de la información .</p> <p>Comprende las regulaciones legales que aplican a la retención de Big Data para cada tipo de industria y región pertinente.</p> <p>Se lleva registro del volumen de almacenamiento para la Big Data.</p> <p>Se comprime y archiva la Big Data para reducir los costos de TI y mejorar el rendimiento de las aplicaciones</p>
<b>4</b>	<p>Se lleva control y auditan el volumen de almacenamiento para la Big Data .</p> <p>Se llevan control y auditan los costos del almacenamiento.</p> <p>Se desecha de forma responsable la Big Data que ya no se requiere según las regulaciones y las necesidades del negocio.</p> <p>Se han cuantificado los beneficios de administrar el ciclo de vida de la información.</p>
<b>5</b>	<p>Se gestiona totalmente el ciclo de vida de los datos, incluso los capturados por streaming.</p> <p>Se gobiernan adecuadamente los datos capturados a través de las redes sociales</p> <p>Se evalúa periódicamente si las políticas, procesos, procedimientos y roles, siguen siendo compatibles y efectivos con las premisas u objetivos.</p> <p>Se utilizan indicadores para dimensionar y dar forma a la mejora continua.</p>

<b>CAPACIDAD TECNOLÓGICA</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>1</b>	<p>No existen políticas, practicas o procedimientos relacionados con la capacidad tecnológica.</p> <p>Poco conocimiento de las tecnologías de Big Data.</p>

<b>CAPACIDAD TECNOLÓGICA</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
	<p>No se han identificado las necesidades de infraestructura para dar soporte a los proyectos de Big Data.</p> <p>No se tiene conciencia de la importancia de las capacidades tecnológicas de la Big Data.</p> <p>No están definidos los roles relacionados a la tecnología.</p>
<b>2</b>	<p>Se han definido roles responsables del tecnología para la Big Data.</p> <p>Se han identificado las necesidades de infraestructura para dar soporte a los proyectos de Big Data, pero sin estar estandarizado.</p> <p>Se definen algunas políticas entorno a las capacidad tecnológica, pero inconsistente.</p> <p>Se comprende la importancia de la tecnología para el éxito en la gestión de la Big Data.</p>
<b>3</b>	<p>Están estandarizadas y documentados las políticas, practicas o procedimientos que contribuyen a la gestión de las capacidad tecnológica.</p> <p>Hay una definición de arquitectura de tecnología incipiente.</p> <p>Se ha evaluado la disponibilidad de la solución de Big Data .</p> <p>Existe un plan de recuperación de desastres.</p> <p>Las soluciones tecnológicas se adoptan como mejores prácticas y se ponen a disposición en toda la institución</p>
<b>4</b>	<p>La arquitectura de tecnología esta claramente definida.</p> <p>Se lleva registro y control de los cambios en infraestructura de la Big Data.</p> <p>Se mide y controla el impacto de la negación de servicio.</p> <p>Se mide y controla la performance de la infraestructura .</p> <p>Se gestiona la infraestructura de archivos de la Big Data .</p>
<b>5</b>	<p>La arquitectura de tecnología responde a los objetivos estratégicos de la empresa.</p> <p>Todas las disciplinas y aspectos de la Big Data y su gobernanza tienen una solución tecnológica disponible.</p> <p>Se gestiona totalmente las capacidades tecnológicas.</p> <p>Se evalúa periódicamente si las políticas, procesos, procedimientos y roles, siguen siendo</p>

<b>CAPACIDAD TECNOLÓGICA</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
	<p>compatibles y efectivos con las premisas u objetivos.</p> <p>Se utilizan indicadores para dimensionar y dar forma a la mejora continua.</p>

#### **4.4.6. Evaluación de Habilidades Soft**

A continuación se detalla el modelo de evaluación de habilidades soft. A diferencia del modelo de evaluación hard, no se divide por disciplina de la gobernanza de Big Data, sino que se evaluó para todas en conjunto. Se muestra un cuadro donde se describen las características cualitativas propias de cada nivel de madurez para las habilidades soft Participación, Liderazgo, Entendimiento y confianza.

<b>Participación, Liderazgo, Entendimiento y confianza</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
<b>1</b>	<p>No existen políticas, prácticas o procedimientos relacionados con las habilidades Soft tales como Participación, Liderazgo, Entendimiento y confianza.</p> <p>Conocimiento muy limitado del propósito o valor de Big Data.</p> <p>Conocimiento muy limitado de las políticas, estándares y mejores prácticas de gobernanza de Big Data.</p> <p>Las personas no tienen conocimiento de su papel dentro del programa de gobernanza de Big Data.</p>
<b>2</b>	<p>Los ejecutivos son conscientes de la existencia del programa de gobernanza de Big Data.</p> <p>Poco conocimiento del programa fuera de la alta dirección.</p> <p>Las acciones en el programa de gobernanza de Big Data están dadas por el cumplimiento de directivas por parte del personal sin tener clara conciencia de su rol.</p> <p>El liderazgo solo está motivado en los altos cargos, a nivel de cargos más bajos solo se da por</p>

<b>Participación, Liderazgo, Entendimiento y confianza</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
	<p>impulsos personales.</p> <p>Se busca fomentar el trabajo en equipo en la organización.</p>
<b>3</b>	<p>Están documentados las políticas, practicas o procedimientos que contribuyen a la gobernanza de Big Data mediante las habilidades Soft.</p> <p>Los ejecutivos entienden cómo la gobernanza de Big Data se beneficia / impacta en la organización,. Los ejecutivos promueven activamente la gobernanza de Big Data dentro de sus grupos.</p> <p>Se fomentan los planes de comunicación para que los interesados conozcan el programa.</p> <p>Esta estandarizado la promoción del liderazgo dentro de los grupos, pero aun falta arraigo en la cultura de la organización.</p>
<b>4</b>	<p>Se fomenta el liderazgo a través de factores cualitativos tales como, poder de decisión, motivación, trasmisión de información y roles claros.</p> <p>Se monitorea o controla cómo ocurren los procesos de colaboración.</p> <p>Los miembros de la organización se comprometen con sus tareas y responsabilidades, teniendo una reconocimiento formal.</p> <p>Los empleados entienden en qué parte del proceso se encuentran y cómo colaboran con el objetivo final.</p>
<b>5</b>	<p>Existen planes de carrera para el personal involucrado en Big Data.</p> <p>El directorio y los stakeholders están involucrados en el plan estratégico del modelo de Big Data .</p> <p>Los lideres apuntalan la gestión de la Big Data y tiene un rol fundamental.</p> <p>Se evalúa periódicamente si las políticas, procesos, procedimientos que contribuyen a la gobernanza de Big Data mediante las habilidades soft.</p> <p>Se utilizan indicadores para dimensionar y dar forma a la mejora continua.</p>

#### **4.4.7. Evaluación de resultados**

A continuación se describe el modo de evaluación de resultados para poder determinar el nivel de madurez en cada habilidad y disciplina.

Una organización está en el nivel uno de madurez, si cumple con todas las características del mismo. Una organización está en el nivel dos de madurez, si cumple con todas las características del nivel uno y con todas las características del nivel dos. De este modo sucesivamente, para estar en un nivel hay que primero cumplir con todos los niveles anteriores.

Para medir el nivel madurez de los aspectos cualitativos del modelo de gobernanza de Big Data de una organización, se utiliza las siguiente Tabla 13 como ejemplo. En ésta, en el caso de la evaluación hard, se debe registrar su puntaje por cada disciplina y luego se calcula el promedio de las mismas. El promedio alcanzado para el total de las disciplinas es el nivel de madurez de la organización en el área hard.

En el caso de la evaluación soft, el puntaje obtenido será el nivel de madurez de la organizacional en el área soft.

La primera evaluación de una organización debe hacerse lo más temprano posible en el programa de gobernanza de Big Data. En el momento de la evaluación inicial, se debe determinar con qué frecuencia se evaluará el programa de gobernanza de Big Data. La frecuencia de las evaluaciones puede depender de muchos factores, incluidos los recursos disponibles para el programa o la madurez de la organización en el momento inicial.

Se recomienda tener un plan de evolución dentro del modelo, es decir fijar un rumbo, metas a alcanzar, y de este modo trazar un camino guía para obtener los objetivos.

<b>Disciplina</b>	<b>Nivel</b>
<b>Organización</b>	2
<b>Metadata</b>	3
<b>Privacidad</b>	3
<b>Calidad de datos</b>	2

<b>Disciplina</b>	<b>Nivel</b>
<b>Integración de procesos de negocio</b>	3
<b>Integración de datos maestros</b>	3
<b>Ciclo de vida de la información</b>	4
<b>Capacidad tecnológica</b>	3
<b>Nivel de madurez hard (promedio)</b>	<b>2,9</b>

Tabla 13. Ejemplo de registro de niveles Hard

#### **4.4.8. Evolución dentro del modelo de madurez**

Un punto fundamental es la planificación de la evolución dentro del modelo de madurez. Para esto se deben establecer objetivos a corto y mediano plazo que permitan a la organización crear una hoja de ruta. Dicha hoja de ruta será la guía para lograr un gobierno de Big Data exitoso.

Dentro del proceso la evaluación inicial, la organización y/o la junta de gobierno de datos, deben establecer objetivos de madurez. Dichos objetivos deben fijarse tanto a corto como mediano plazo y deben estar alineados con la planificación de evaluaciones previstas del nivel de madurez.

La estrategia para la selección de los objetivos depende del enfoque y necesidades de la organización, se puede ir evolucionado de manera integral o hacer foco en disciplinas que tengan mayor valor estratégico. Para la selección de los objetivos se debe tener en cuenta que sean medibles y alcanzables.

Para implementar la hoja de ruta, se crea un gráfico de dispersión similar al ejemplo que se muestra en la Ilustración 15. En dicho gráfico se marca el nivel actual, en naranja, para cada disciplina. Luego de acuerdo a los objetivos planteados, se marca, con verde, el nivel objetivo, es decir la meta a llegar para la próxima evaluación. Junto a la gráfica es bueno

registrar las medidas necesarias para lograr dicha meta y un plan de acción que ponga en ejecución esas medidas.

De acuerdo con el cronograma acordado en la evaluación inicial, se van realizando las diferentes evaluaciones del nivel de madurez y cotejando con la gráfica de objetivos. De este modo se controla si la evolución es la esperada, y de lo contrario se toman medidas correctivas.

Por ultimo se deberían hacer revisiones periódicas de las metas y sus resultados, para de este modo poder tomar acciones en caso de que los mismos ya no estén alineados con la estrategia actual.

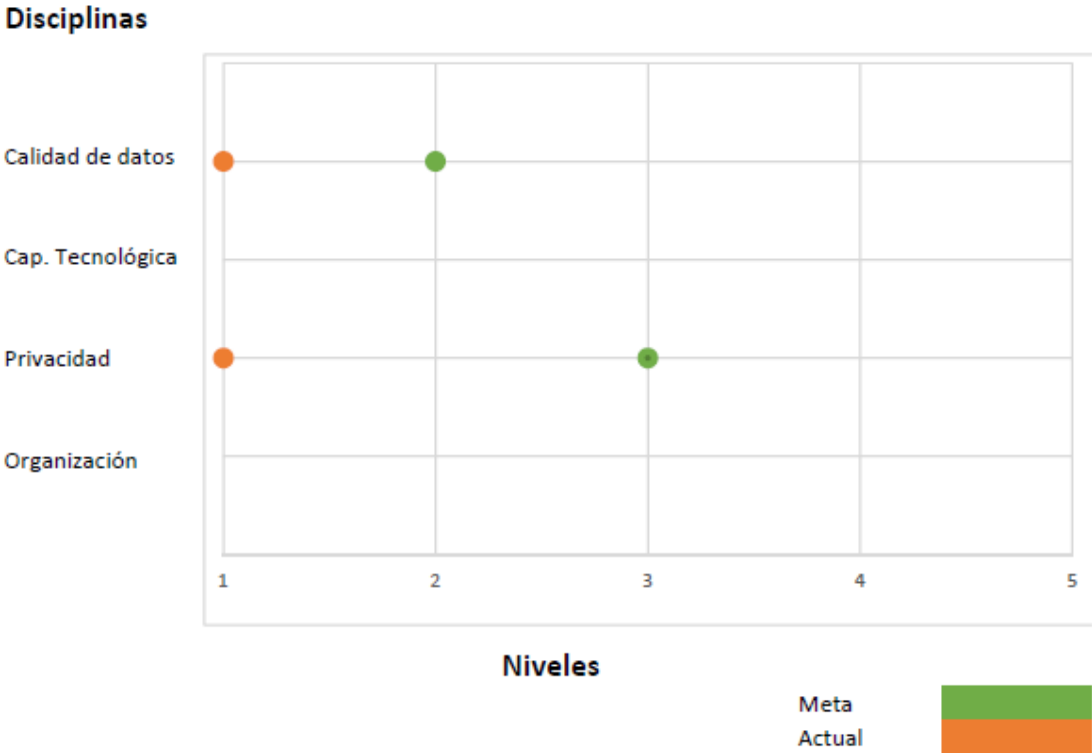


Ilustración 15. Gráfica de objetivos

#### **4.4.9. Metodología de validación del modelo de madurez**

El objetivo principal de la validación es obtener feedback para saber si nuestro modelo podía ser realmente aplicado en una organización, es decir, si toda la teoría planteada condice con la realidad en el área de la Big Data.

Para esto, se contaba con dos metodologías de validación. Por un lado, consultar a expertos en el tema y obtener su opinión, aunque basada en experiencia, también teórica. Por otro lado, se podía aplicar una metodología empírica, llevando a la práctica el modelo planteado en este proyecto a una organización que le interesase implementar el gobierno de Big Data.

Del proyecto insumo [3] se tenía una validación hecha por expertos, que a pesar de ser solo sobre el modelo de evaluación, fue de utilidad. Por este motivo se optó por la validación empírica, que se cree es la que puede dejar mayor aprendizaje, una feedback real y práctico, para poder validar y mejorar el presente modelo.

#### **4.4.10. Validación empírica de la solución propuesta**

##### **Introducción**

En el actual capítulo se desarrolla la introducción y planificación del proceso de validación empírica del modelo de madurez de gobernanza de Big Data propuesto.

La validación de la solución propuesta está diseñada en 4 fases:

**Fase 0:** Evaluación por expertos. Esta fase considera las opiniones vertidas por expertos consultados en el proyecto desarrollado Darío Carlos y Mauro Nascimento [3]. Tanto el modelo de evaluación como las oportunidades de mejora han sido el punto de partida del presente trabajo.

**Fase 1:** Esta fase corresponde al diseño del estudio de caso basado en la metodología propuesta por Robert K. Yin [20]. El estudio se desarrollará en una PyME nacional.

Ejecución del estudio de caso. El estudio se dividió en dos etapas: análisis de la situación actual, roadmap que permita alcanzar mayores niveles de madurez.

**Fase 2A:** La primera etapa es un análisis de la situación actual, para lo cual en varias entrevistas con distintos actores de la organización se medirá el nivel de madurez con que se trabajan los proyectos relacionados a Big Data.

**Fase 2B:** La segunda etapa es, a partir de la implementación de buenas prácticas de gobernanza de big data en el contexto de proyectos específicos, alcanzar al menos un nivel de madurez superior.

**Fase 3:** Análisis de los resultados obtenidos en las fases anteriores.

El alcance del proyecto actual permite cumplir la fase 2A, siendo el punto de partida para la incorporación de buenas prácticas y el proceso de medición asociado (fases 2B y 3).

## FASE 1

Para alcanzar los objetivos, con fecha 19 de julio, se contactó a la empresa “EFA Laboratorios (Antía, Moll y Cia SA)” (de aquí en más, la empresa), donde se implementará el estudio de caso. El acuerdo y alcance del estudio se definió con fecha 20 de agosto Ver **Anexo 4**.

De acuerdo con Schramm, 1971, citado por [20], esta metodología es adecuada ya que este estudio es descriptivo y exploratorio donde lo que se busca es explicar un fenómeno que ocurre en tiempo presente.

	<b>Proposiciones</b>
<b>P1</b>	¿Cómo influye la evaluación de la madurez en los procesos de gobierno de big data en la calidad de los mismos?
<b>P2</b>	¿Cómo afecta el nivel de madurez en relación a la adquisición de habilidades blandas de los actores que intervienen en proyectos de big data en relación al éxito de los mismos?

Tabla 14. Proposiciones del estudio del caso

## La empresa

La empresa donde se llevó a cabo el caso de estudio es un laboratorio que se especializa en la fabricación, comercialización y distribución de especialidades farmacéuticas. Inició sus actividades en mayo de 1951. Desde el año 1972, se instaló en su sede actual, en Montevideo.

Allí se encuentra su sede comercial y productiva ya que el edificio fue adaptado para cumplir con las necesidades requeridas para un laboratorio farmacéutico.

Actualmente el laboratorio produce, comercializa y distribuye más de 160.000 unidades farmacéuticas mensuales, lo que significa el 1,5% de las unidades físicas del mercado farmacéutico uruguayo. La empresa representa en Uruguay, desde 1986 a uno de los primeros centros europeos de investigación terapéutica y el primer laboratorio farmacéutico independiente francés, presente en 140 países del mundo y fuente permanente de innovación farmacéutica al servicio del hombre. Características generales de la empresa:

1. Empresa nacional con 1,5% del mercado.
2. Es una PyME que no supera los 85 empleados.
3. El departamento de TI consta de 5 personas en total
4. La empresa tiene interés en desarrollar proyectos basados en grandes volúmenes de datos, provenientes de sus procesos productivos.
5. La empresa es representativa de las PyMEs del mercado uruguayo, sin características particulares que la distinguan del resto. Esto es importante ya que este caso es único y por tanto la empresa debe representar la media de las PyMEs uruguayas.

A continuación, se detallan los roles participantes:

<b>Rol o Cargo</b>	<b>Descripción</b>
<b>Gerente General</b>	La persona que ocupa este cargo, también forma parte de la Junta Directiva.
<b>CIO</b>	Gerente de Sistemas
<b>Gerente de RRHH</b>	El gerente de recursos humanos
<b>Gerente de Producción</b>	El gerente de producción, persona relativamente nueva en la organización (1 año). Comparte parte de las responsabilidades respecto de la gestión de materias primas con el Gerente de Compras.

<b>Rol o Cargo</b>	<b>Descripción</b>
<b>Personal de TI</b>	El personal de TI tiene poca diferenciación de roles.

Tabla 15. Roles participantes del estudio del caso

Procedimiento de recolección de datos

<b>Fuentes de datos</b>	<b>Proposiciones</b>
<b>Entrevistas personales</b>	
Se llevaron adelante entrevistas informales con los diferentes participantes del caso.	Pr1, Pr2
<b>Revisión de registros documentales</b>	
Contratos de outsourcing	Pr1
Currículum de los empleados	Pr2
Análisis documental de proyectos	Pr1

Tabla 16. Fuentes de datos del estudio de caso y su relación con las proposiciones del mismo

## **FASE 2A**

Esta fase establece la línea base del proyecto, al medir el nivel de madurez actual se establece el punto de partida para el proyecto de mejora definido por la gerencia.

La empresa ha emprendido un proyecto utilizando datos provenientes de parte de sus procesos de producción y control de calidad, en particular los que controlan la homogeneidad de los productos activos seleccionados (ej. Furosemida).

La categorización de la big data es “generada por máquinas”, proviene de un cromatógrafo (Waters - Alliance 2695 con detector UV), la cual es enriquecida con información estadística

histórica y datos provenientes de condiciones medioambientales. El objetivo del proyecto es “predictivo”, detectar micro-condiciones medioambientales que afectan la homogeneidad de las mezclas.

## **Evaluación**

La evaluación se comenzó el 10 de setiembre y se evaluaron las disciplinas metadata, calidad de datos, capacidad tecnológica y habilidades blandas. La misma continuará a un ritmo de 1 entrevista por semana involucrando a los diversos actores del proyecto.

El método de recolección de datos fue la entrevista personal en la que intervinieron los actores descritos en la tabla 15.

Para las habilidades blandas, la evaluación dio como resultado que cumplen completamente con el nivel 1 de madurez y con un 20% del nivel de madurez 2. Lo cual da un nivel total de habilidades blandas de 1,2. Ver en Anexo 3 las respuestas a la evaluación.

A continuación en la tabla 17 se muestra los resultados para habilidades hard de la disciplinas evaluadas hasta el momento y el nivel de madurez actual. El promedio se hizo solo sobre las disciplinas evaluadas y dio un nivel resultado de 2,2 para las habilidades hard. Ver en Anexo 3 las respuestas a la evaluación.

<b>Disciplina</b>	<b>Nivel</b>
<b>Organización</b>	0
<b>Metadata</b>	2
<b>Privacidad</b>	0
<b>Calidad de datos</b>	3
<b>Integración de procesos de negocio</b>	0

<b>Disciplina</b>	<b>Nivel</b>
<b>Integración de datos maestros</b>	<b>0</b>
<b>Ciclo de vida de la información</b>	<b>0</b>
<b>Capacidad tecnológica</b>	<b>1,5</b>
<b>Nivel de madurez hard</b>	<b>2,2</b>

Tabla 17. Resultados evaluación

Como conclusión, la evaluación de la organización hasta el momento, en cuanto a la metodología y el modelo, es muy satisfactoria. Todo esto está apuntalado por la empresa seleccionada, que realmente siente la necesidad de adoptar la gobernanza de Big Data.

## **5. Gestión de proyecto**

### **5.1 Metodología de trabajo del equipo**

El equipo de proyecto se conformó con dos estudiantes, uno trabajando desde Uruguay y otro radicado en Alemania. Este hecho fue el primer gran desafío del trabajo dado la imposibilidad de reunirse físicamente, el uso horario con diferencia de cinco horas y ciertas actividades que requieren presencia física.

Por este motivo, más allá de la tradicional planificación de tareas y la gestión del proyecto, se tuvo que coordinar cada una de las reuniones con el tutor en horarios factibles para todos. Por otra parte el trabajo entre los integrantes del equipo también se tuvo que realizar en horarios poco comunes, tales como las madrugadas. Estos contactos fueron vía WhatsApp y se utilizaron herramientas para compartir documentación tales como Dropbox.

Partiendo de este primer reto que fue el trabajo a distancia, se realizó un análisis de los diferentes riesgos en el proyecto.

### **5.2 Gestión de riesgos**

Esta actividad tiene como objetivo la planificación de la gestión de los riesgos.

Dentro de la misma se realizó la identificación de los riesgos, análisis cualitativo de estos, planificación de respuesta para mitigarlos y por último las acciones de contingencia en caso de que ocurran. Se realizará el control de los mismos durante todo el proyecto.

El objetivo de este análisis es minimizar el impacto negativo en el proyecto de los posibles factores de riesgo. Para el análisis cualitativo definimos valores que nos indiquen el grado de impacto y probabilidad de ocurrencia, de la multiplicación de estos surgirá la magnitud de riesgo para su gestión.

Luego de haber comenzado el proyecto e ir realizando la planificación de tareas y de proyectar las diferentes etapas a proceder, se realizó una breve identificación de los riesgos. A estos, se les aplica un análisis cualitativo por medio del cual se halla la probabilidad de ocurrencia y el impacto de cada uno de estos riesgos identificados. Para impacto se define una escala de valores entre 0 – 5, y para la probabilidad de ocurrencia se define una escala de valores entre

0.0 – 1.0. Seguidamente se propone un plan de respuesta como también de contingencia para cada riesgo identificado.

Según el PMBOK [39], el riesgo de un proyecto es un evento o condición incierta que, de producirse, tiene un efecto positivo o negativo en uno o más de los objetivos del proyecto, tales como el alcance, el cronograma, el costo y la calidad. Un riesgo puede tener una o más causas y, de materializarse, uno o más impactos. Una causa puede ser un requisito especificado o potencial, un supuesto, una restricción o una condición que crea la posibilidad de consecuencias tanto negativas como positivas. Por lo tanto, es de vital importancia realizar una adecuada gestión de riesgos donde se dé inicio en la primera fase del proyecto, para llevar seguimiento y controles adecuados de estos en cada una de las fases del desarrollo del proyecto, hasta la aceptación del producto final.

Impacto	Descripción	Probabilidad	Descripción
0	Ninguno	0.0	No probable
1	Marginal	0.2	Poco probable
2	Poco importante	0.4	Probable
3	Importante (puede retrasar el proyecto)	0.6	Muy probable
4	Crítica (puede detener el proyecto)	0.8	Altamente probable
5	Catastrófica (fracaso del proyecto)	1.0	Es un hecho

Tabla 18. Escala de valores de impacto y probabilidad

Riesgo	Imp.	Prob.	Respuesta	Contingencia
Falta de conocimiento sobre el tema	4	0.2	Previo a comenzar se buscan artículos y material de calidad respecto al tema. Se estudia dicho material	Replantearse si el material utilizado es el apropiado. Buscar nueva información o se selecciona otro material de lo previamente investigado

Falta de conocimiento de metodologías de investigación	3	0.2	Previo a comenzar se investigan metodologías y luego se define metodología de trabajo	Replantearse si el material utilizado es el apropiado. Buscar nueva información
Metodología de investigación no proporciona resultados buscados	3	0.2	Previo a comenzar se investigan varias metodologías de trabajo y se elige la que mejor sirve al equipo	Utilizar otra de las metodologías que se tuvieron en cuenta antes del comienzo o buscar nuevas
No encontrar información o información poco fiable sobre el tema	5	0.2	Se define un periodo de búsqueda de información relevante al proyecto tomando de varias fuentes distintas	Se busca nueva información utilizando nuevas fuentes
Modalidad de trabajo a distancia	2	1	Se tiene en cuenta al seleccionar metodología de trabajo	Modificar metodología de trabajo establecida
No poder dedicarle tiempo al proyecto	5	0.2	Se toma en cuenta cantidad de horas de trabajo que cada individuo del equipo posee por semana y se define alcance	Modificar alcance establecido
Falta de compromiso	3	0.2	El equipo se informa sobre el tema del proyecto, posibles metodologías de trabajo que se utilizaran y horas de trabajo necesarias por semana previo a comprometerse	Se investigan y se aplican técnicas de motivación para aumentar el compromiso del equipo
No poder continuar con el proyecto por problemas personales	5	0.4	Los integrantes del equipo deben de informarse en toda área del proyecto	Si un integrante del equipo está disponible continua con el proyecto o si esto no es posible se pospone
Que el modelo no sea aceptado	5	0.2	Se trabaja con tutor con experiencia previa validando modelos de madurez	Modificar modelo tomando en cuenta comentarios de expertos que no aceptaron modelo presentado

No poder validar el modelo	4	0.4	Se define método para la validación del modelo	Cambiar método de validación del modelo
----------------------------	---	-----	--	---

Tabla 19. Identificación, calificación, plan de respuesta y plan de contingencia de riesgos

### 5.3 Gestión del avance

Para la planificación y la gestión de las tareas del proyecto se utilizó la herramienta web Trello. En la misma se definió un tablero del estilo Kanban con tres columnas. La primera columna dedicada para las tareas a realizar, la segunda columna las tareas que se encuentran actualmente en proceso y por último, en la tercera columna, las tareas que fueron exitosamente finalizadas.

Al comienzo se definieron una serie de tareas, pero vale destacar que a lo largo del proyecto se fueron introduciendo nuevas que no habían sido consideradas al principio del mismo. A medida que se comenzaron las tareas, estas fueron asignadas al integrante responsable de ejecutarla, y de ser necesario una fecha de finalización, para luego colocarlas en la columna de tareas en proceso. Una vez que la tarea se completa, se sitúa la misma en la columna final.

Esta forma de trabajo permitió al equipo un control sobre en qué tarea está trabajando cada uno, cuáles son las tareas restantes, y cuáles son las tareas ya realizadas. Obteniendo de esta manera el control sobre el avance del proyecto en general y una ayuda visual para las reuniones semanales de avance. Para más detalle del tablero comentado ver Anexo 2.

## 6. Conclusiones y trabajos futuros

Estamos viviendo un momento único en la historia dado que comenzamos a transitar por la cuarta revolución industrial. Esta es la revolución de la información, y por ende, de las tecnologías de la información. Dentro de este contexto tan amplio entra a tener una importancia fundamental la gestión de datos y más que eso la gestión de grandes volúmenes de datos, es decir la Big Data. Una estrategia exitosa de Big Data se basa en un plan de gobierno de datos bien ejecutado que incluya la gobernanza de Big Data.

En este proyecto se intentó dilucidar si existía un modelo de madurez que ayudara a la evaluación, implementación y mejora del gobierno de de Big Data. Y de no ser así, intentar desarrollar uno.

Para esto se realizó una revisión literaria, planteándose una metodología de investigación la cual permitiera obtener resultados exitosos. Dentro de esta metodología se definieron búsquedas y filtro para ir refinando la información hasta obtener un conjunto de datos de calidad.

En base a estos datos de calidad se comenzó un análisis, el que permitió contestar las preguntas de investigación y cimentar las bases del modelo planteado en este proyecto. De las respuestas surgió la no existencia de un modelo de madurez aceptado y la amplia necesidad de este en la industria .

Una vez detectada la necesidad, se comenzó el desarrollo de este modelo. Para esto nos apoyamos en diferente literatura y modelos de madurez que existen en la industria pero por impulsos individuales o de organización, no se estandarizó y con carencias en ciertos aspectos como el de habilidades blandas y tecnológicas.

Finalmente este modelo se nutrió del material y experiencia de la academia e industria en consecuencia de la investigación. Como resultado, creemos que se obtuvo un modelo que contempla todos los aspectos de la gobernanza de Big Data, e incluyó nuevas dimensiones como las habilidades blandas y tecnología. Apostamos también a que sea aplicable en la industria, como lo demuestra la validación, y una verdadera guía en la adopción de la gobernanza de Big Data.

A futuro se piensa en seguir con la validación comenzada, vista en el capítulo de validación y ampliarlo a otras organizaciones. De este modo poder lograr un modelo robusto, que contemple todo el contexto del entorno de la Big Data y que puede ser referente en el mercado.

## 7. Referencias Bibliográficas

1. M. Brackett, *The DAMA Guide to The Data Management Body of Knowledge* , 2009.
2. S. Soares, *Big Data Governance> An Emerging Imperative*. USA: MC PRESS, 2012.
3. D. Carlos y M. Nascimento, “Modelo para la evaluación de buenas prácticas de gobernanza de Big Data”, Proyecto, Universidad ORT Uruguay, Montevideo, Uruguay, 2019.
4. ISO/IEC, *ISO/IEC 38500:2015. Corporate Governance for Information Technology*, 2015.
5. P. Weill and J. Ross, *IT Governance: How Top Performers Manage IT Decision Rights for Superior Results*, 2004.
6. S. Warren and L. Brandeis, *The Right to Privacy*. *Harvard Law Review*, Vol. IV No5, December , 1890
7. M. Ballester, “Gobierno de las TIC ISO/IEC 38500”. [OnLine] Disponible : [https:// www.isaca.org/Journal/archives/2010/Volume-1/Pages/Gobierno-de-las-TIC-ISO-IEC-385001.aspx](https://www.isaca.org/Journal/archives/2010/Volume-1/Pages/Gobierno-de-las-TIC-ISO-IEC-385001.aspx)
8. D. Smits and J. van Hillegersberg, "The development of a hard and soft IT governance assessment instrument," *Procedia Computer Science*, vol. 121, pp. 47-54, enero 2017.
9. J. Bessant, S. Caffyn, and M. Gallagher, "An evolutionary model of continuous improvement behaviour," *Technovation*, vol. 21, pp. 67-77, enero 2001.
10. J. Collins, *Level 5 Leadership: The Triumph of Humility and Fierce Resolve* vol. 79, 2001.
12. B. H. Reich and I. Benbasat, "Measuring the Linkage between Business and Information Technology Objectives," *MIS Quarterly*, vol. 20, pp. 55-81, 1996.
13. M. Chrissis, M. Konrad and S. Shrum , *CMMI Guía para la integración de procesos y la mejora de productos*, 2006
14. B. Kitchenham, *Procedures for Performing Systematic Reviews* vol. 33, 2004.

15. Stanford University. "Stanford Data Governance Maturity Model". Retrieved April 19, 2016, [OnLine] Disponible : <http://web.stanford.edu/dept/pres-provost/cgi-bin/dg/wordpress/wp-content/uploads/2011/11/StanfordDataGovernanceMaturityModel.pdf>
16. J. Aguiar, R. Pereira, B. Vasconcelos, and I. Bianchi. "An overlapless incident management maturity model for multi-framework assessment (ITIL, COBIT, CMMI-SVC)". *Interdisciplinary Journal of Information, Knowledge, and Management*, 13, 137-163. , 2018
17. D. Proença and J. Borbinha, "Maturity Models for Information Systems - A State of the Art," *Procedia Computer Science*, vol. 100, pp. 1042-1049, enero 2016.
18. M. Khoshgoftar<sup>1</sup> and Omar Osman, "Comparison of Maturity Models", School of Housing, Building and Planning, Universiti Sains Malaysia, Pulau Pinang, Malaysia
19. G. García and A. Félix "Development of Maturity Models: A Systematic Literature Review", Instituto de Tecnologías y Sistemas de Información, University of Castilla-La Mancha, España
20. R. Yin, *Case study research: design and methods*. 4 ed, Beverly Hills, CA: Sage Publications, 2014
21. Z. Ali Al-Sai, R. Abdullah and M. heikal husin, "A Review on Big Data Maturity Models", in IEEE Jordan international joint conference on electrical engineering and information technology, 2019
22. D. Heredia y W. Nieto "Propuesta de una Metodología para la Adopción del Big Data", Universidad Simón Bolívar, Barranquilla, Colombia
23. A. Al-Badia , A.Tarhinia and A.Islam Khana, " Exploring Big Data Governance Frameworks " in The 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks, 2018
24. D. Smits and J. van Hillegersberg, "The continuing mismatch between IT governance maturity theory and practice: a new approach," *Procedia Computer Science*, vol. 138, pp. 549-560, enero 2018.

25. M. Al-Ruithe and E. Benkhelifa, "Analysis and Classification of Barriers and Critical Success Factors for Implementing a Cloud Data Governance Strategy," *Procedia Computer Science*, vol. 113, pp. 223-232, enero 2017.
26. D. Proença, R. Vieira and J. Borbinha “ Information Governance Maturity Model Final Development Iteration ”, in Conference: International Conference on Theory and Practice of Digital Libraries , 2017
27. G. Cheng, Y. Li, Z. Gao and X. Liu “Cloud Data Governance Maturity Model”, China Electronic Product Reliability and Environmental Testing Research Institute Guangzhou, 510610, P. R. China
28. C. Juiz, R. Colomo-Palacios, and B. Gómez, "Cascading ISO/IEC 38500 based Balanced Score Cards to improve board accountability," *Procedia Computer Science*, vol. 138, pp. 417-424, enero 2018.
29. I. S. Bianchi, R. D. Sousa, R. Pereira, and E. Luciano, "IT Governance Structures in Brazilian, Dutch and Portuguese Universities," *Procedia Computer Science*, vol. 121, pp. 927-933, enero 2017.
30. A. Ekambaram, A. Ø. Sørensen, H. Bull-Berg, and N. O. E. Olsson, "The role of Big Data and knowledge management in improving projects and project-based organizations," *Procedia Computer Science*, vol. 138, pp. 851-858, enero 2018.
31. D. Heredia y W. Nieto, “Propuesta de una Metodología para la Adopción del Big Data”, in 12th Iberian Conference on Information Systems and Technologies (CISTI) 2017
32. S. Mohd & N. Hafizah, “BIG DATA MATURITY MODEL – A PRELIMINARY EVALUATION”.Universiti Tenaga Nasional, Malaysia, 2017
33. IBM Data Governance Council. “*The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance*”. Retrieved April 19, 2016, [OnLine] Disponible : [https://www-935.ibm.com/services/uk/cio/pdf/leverage\\_wp\\_data\\_gov\\_council\\_maturity\\_model.pdf](https://www-935.ibm.com/services/uk/cio/pdf/leverage_wp_data_gov_council_maturity_model.pdf)

34. V. Dhanuka, “ Big Data Maturity Model”, 2016, [OnLine] Disponible : <http://hortonworks.com/wp-content/uploads/2016/04/Hortonworks-Big-Data-Maturity-Assessment.pdf>
35. F. Halper and K. Krishnan “TDWI Big Data Maturity Model Guide”, 2014, [OnLine] Disponible : <https://tdwi.org/whitepapers/2013/10/tdwi-big-data-maturity-model-guide.aspx>
36. B. Quinto, "Big Data Governance and Management," in *Next-Generation Big Data: A Practical Guide to Apache Kudu, Impala, and Spark*, ed Berkeley, CA: Apress, 2018, pp. 495-506.
37. V. Morabito, "Big Data Governance," in *Big Data and Analytics: Strategic and Organizational Impacts*, ed Cham: Springer International Publishing, 2015, pp. 83-104.  
BigDataGov - MM - A 21/02/2019
38. M. Serrato and J. Ramirez, "The Strategic Business Value of Big Data," in *Big Data Management*, F. P. García Márquez and B. Lev, Eds., ed Cham: Springer International Publishing, 2017, pp. 47-70.
39. Project Management Institute. *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*. Newtown Square, Pa: Project Management Institute, 2004.

## Anexo 1 – Modelo de evaluación insumo

### Modelo original para la evaluación de buenas prácticas de gobernanza de Big Data []

#### Organización

<b>Habilidades HARD</b>	
1	¿Se han identificado los procesos de negocio donde se genera Big Data?
2	¿Se han identificado los procesos de negocio que requieren decisiones basadas en Big Data?
3	¿Se han identificado los roles existentes dentro de la organización que serán responsables en alguna medida de la Big Data?
4	¿Se han definido nuevos roles dentro de la organización responsables en alguna medida de la Big Data?
5	¿Los nuevos roles definidos, tienen en cuenta las características de cada uno de los tipos de Big Data?
6	¿Se ha definido una matriz RACI para cada proceso identificado?
7	¿Existe un marco de gobernanza de datos en la organización?
8	¿Se han integrado al marco de gobernanza de datos la Big Data ?
9	¿Se planifica el trabajo a realizar en Big Data?
10	¿Cuántas personas están encargadas de la gestión de la Big Data a nivel organización?
11	¿Cuántas personas están encargadas de la gestión de la Big Data a nivel de sector?
<b>Habilidades SOFT</b>	
Participación	
12	¿Se busca fomentar el trabajo en equipo en el sector?
13	¿Existen planes de comunicación?
14	¿Se monitorea o controla cómo ocurren los procesos de colaboración?
15	¿Los procesos son formalmente finalizados y comunicados al personal?
16	¿Existen planes de carrera para el personal involucrado en Big Data?
17	¿El directorio y los stakeholders están involucrados en el plan estratégico del modelo de Big Data?
Liderazgo	
18	¿Qué tipo de liderazgo se practica en el área de Big Data?
19	¿Qué tipo de liderazgo se busca fortalecer o priorizar para el área de Big Data?
20	¿Son incorporados factores cualitativos (motivación, comunicación adecuada) a la hora de tomar decisiones?
Mejora continua	
21	¿Se evalúa periódicamente si los procesos siguen siendo compatibles con las premisas u objetivos?
22	¿Se registran y analizan las fortalezas y debilidades?
23	¿Las lecciones aprendidas se capturan y transmiten?
24	¿Se tiene el tiempo necesario para que el equipo de manejo de Big Data lea o cree procedimientos y/o manuales?

25	¿El equipo de manejo de Big Data está en contacto con los usuarios para entender sus necesidades?
26	¿El equipo de manejo de Big Data está en contacto con los usuarios para entender cómo utilizan sus recursos?
27	¿Se evalúan las capacidades de los RRHH respecto de los proyectos de Big Data?
<b>Entendimiento y confianza</b>	
28	¿Los miembros de la organización se comprometen con sus tareas y responsabilidades?
29	¿Se reconoce de alguna manera formal (no necesariamente financiera) la contribución de los empleados a los procesos de mejora?
30	¿Los gerentes apoyan el experimento no castigando los errores sino fomentando el aprendizaje de ellos?
31	¿Se utiliza la delegación para brindar poder decisión a los empleados?
32	¿Los empleados entienden en qué parte del proceso se encuentran y cómo colaboran con el objetivo final?
33	¿Se utilizan las reuniones cara a cara para establecer una relación de confianza antes de usar medios de comunicación habituales?
34	¿El directorio toma seriamente cualquier riesgo reportado por el sector de manejo de Big Data aunque haya sido consecuencia de un error del equipo?

## Metadata

<b>Habilidades HARD</b>	
1	¿La organización cuenta con un glosario de datos definido?
2	¿Se ha definido en un glosario el significado en términos del negocio de la Big Data?
3	¿Se han considerado las características tecnológicas en la definición de la metadata para Big Data?
4	¿Se han identificado los datos sensibles dentro de la Big Data?
5	¿Se han marcado en el glosario los datos sensibles encontrados en la Big Data?
6	¿Se sincroniza la metadata perteneciente a los distintos repositorios de Big Data?
7	¿Se han identificado las fuentes de Big Data de datos operacionales?
8	¿Se ha definido la metadata de los datos operacionales fuente de Big Data?
9	¿Se ha incluido en la metadata de la Big Data información que asegure el linaje de la misma?
10	¿Se almacenan datos no estructurados? ¿Se definen índices para hacer más efectivas las búsquedas en dicho tipo de dato?
11	¿Se han definido roles responsables de la metadata de la Big Data ?
12	¿Existe colaboración entre sectores, en distintos procesos de creación-modificación de Big Data?
<b>Habilidades SOFT</b>	
<b>Participación</b>	
13	¿Se han definido roles responsables de la metadata de la Big Data ?
14	¿Existe colaboración entre sectores en distintos procesos de creación y/o modificación de metadata?
15	¿Se han comunicado a la organización o a los diferentes sectores las acciones realizadas sobre la metadata?
16	¿Se ha definido la metadata de acuerdo con la unidad de negocio dueña de la Big Data?

Liderazgo	
17	¿Existe un líder que administre los la metadata de la Big Data?
18	¿Cuál es el tipo de líder que se busca para la gestión de la metadata?
19	¿Cuáles son las habilidades que busca en un líder encargado de la metadata?
Mejora continua	
20	¿Se evalúa periódicamente si los procedimientos siguen siendo compatibles con las premisas u objetivos?
21	¿Las personas (como individuos y/o grupos) inician y llevan a cabo las actividades y participan en el proceso de mejora continua?
22	¿Las personas usan indicadores para dimensionar y dar forma a la mejora continua?
23	¿Se ha definido un período para verificar la calidad de la metadata?
Entendimiento y confianza	
24	¿Los miembros del grupo están al tanto de sus tareas y responsabilidades?
25	¿Se realiza una asignación de tareas adecuadas para que los miembros del equipo estén al tanto de sus responsabilidades y funciones?
26	¿Los individuos evalúan que los cambios a realizar estén en sintonía con los objetivos del sector o de la organización?

## Privacidad

Habilidades HARD	
1	¿Se ha identificado la Big Data que contiene datos sensibles?
2	¿Se ha marcado en el repositorio de metadata de Big Data los datos sensibles identificados?
3	¿Se han tomado en cuenta las leyes regionales que afectan el uso de la Big Data?
4	¿Se han previsto aquellas situaciones de internacionalidad en el uso de Big Data?
5	¿Se registra el acceso a la Big Data?
6	¿Se han definido roles específicos para el acceso a datos sensibles dentro de la Big Data?
7	¿Se controlan dichos accesos?
Habilidades SOFT	
Participación	
8	¿Existe visibilidad entre lo que realizan diferentes sectores de la organización?
9	¿Los cambios realizados en niveles de privacidad y seguridad son formalmente finalizados y comunicados al personal?
10	¿Los niveles de privacidad y seguridad son establecidos por el grupo de manejo de Big Data y por el directorio?
11	¿Los datos que se consideren sensibles son comunicados formalmente al área que los administra?
Liderazgo	
1	¿Existe un líder que administre los datos sensibles de la Big Data?

2	
1	3 ¿Cuál es el tipo de líder que se busca para la seguridad de los datos de la Big Data?
1	4 ¿Cuáles son las habilidades que busca en un líder encargado de la privacidad de Big Data?
Mejora continua	
1	5 ¿Se evalúa periódicamente si los procedimientos siguen siendo compatibles con las premisas u objetivos?
1	6 ¿Las personas (como individuos y/o grupos) inician y llevan a cabo las actividades y participan en el proceso de mejora continua?
1	7 ¿Las personas usan indicadores para dimensionar y dar forma a la mejora continua?
1	8 ¿Existe algún ciclo formal de búsqueda y resolución de problemas de accesos?
1	9 ¿Se registran y analizan las oportunidades y posibles amenazas?
2	0 ¿Se monitorea y analiza de manera periódica para encontrar nuevas amenazas y su posible impacto?
Entendimiento y confianza	
2	1 ¿Los miembros del grupo están al tanto de sus tareas y responsabilidades?
2	2 ¿Se realiza una asignación de tareas adecuadas para que los miembros del equipo estén al tanto de sus responsabilidades y funciones?
2	3 ¿Los individuos evalúan que los cambios a realizar estén en sintonía con los objetivos del sector o de la organización?
2	4 ¿Los individuos y grupos monitorean / miden el impacto que tiene las tareas que realizan en los objetivos estratégicos o sectoriales?

## Calidad de datos

Habilidades HARD	
1	¿Se han identificado los datos críticos que aseguran el éxito del programa de Big Data en la organización?
2	¿Se han establecido los intervalos de confianza para medir la calidad de los datos críticos identificados en la Big Data?
3	¿Para los datos identificados se ha definido la matriz de calidad de datos? <i>Elemento de dato - A qué y cómo afecta la falta de calidad - Responsable</i>
4	¿Se han utilizado los datos semi o no-estructurados para mejorar la calidad de los datos escasamente poblados en la Big Data?

5	¿Se utiliza analítica en tiempo real para determinar en memoria la calidad de los datos en lugar de utilizar espacio en disco?
<b>Habilidades SOFT</b>	
Participación	
6	¿Los individuos y grupos monitorean / miden los resultados de su actividad de mejora y el impacto que tiene en los objetivos estratégicos o sectoriales?
7	¿Existe un mecanismo de incentivo que promueva una cultura de calidad de datos en la Big Data?
8	¿Se han definido roles responsables de la calidad de la Big Data ?
Liderazgo	
9	¿Existe un líder que administre las calidad de los datos de la Big Data?
10	¿Cuál es el tipo de líder que se busca para asegurar la calidad de los datos de la Big Data?
11	¿Cuáles son las habilidades que busca en un líder encargado de la calidad de los datos de la Big Data?
Mejora continua	
12	¿Se evalúa periódicamente si los procedimientos siguen siendo compatibles con las premisas u objetivos?
13	¿Las personas (como individuos y/o grupos) inician y llevan a cabo las actividades y participan en el proceso de mejora continua?
14	¿Las personas usan indicadores para dimensionar y dar forma a la mejora continua?
15	¿Existe algún ciclo formal de búsqueda y resolución de problemas que apunte a la calidad de los datos de la Big Data?
16	¿Se registran y analizan las posibles fortalezas y debilidades?
17	¿Cuándo los procesos son finalizados son formalmente comunicados al personal?
18	¿Se realizan verificaciones periódicas de las fuentes de datos de la Big Data?
Entendimiento y confianza	
19	¿Los miembros del grupo están al tanto de sus tareas y responsabilidades?
20	¿Se realiza una asignación de tareas adecuadas para que los miembros del equipo estén al tanto de sus responsabilidades y funciones?
21	¿Los individuos evalúan que los cambios a realizar estén en sintonía con los objetivos del sector o de la organización?
22	¿Los individuos y grupos monitorean / miden el impacto que tiene las tareas que realizan en los objetivos estratégicos o sectoriales?

## Integración de proceso de negocio

<b>Habilidades HARD</b>	
1	¿Se han identificado los procesos de negocio que se verán afectados por el gobierno de Big Data?
2	¿Ha identificado a los stakeholders del negocio para el programa de gobierno de Big Data?
3	¿Se ha construido un mapa de procesos con las actividades clave para cada uno?
4	¿Se ha establecido la relación entre las políticas del gobierno de Big Data con las actividades en los procesos mapeados?
5	¿Ha cuantificado los beneficios financieros de la gestión de Big Data?
6	¿Que tan alineada esta IT y la Big Data con los objetivos de la organización?
<b>Habilidades SOFT</b>	
Participación	
7	¿Existe colaboración entre sectores en distintos procesos de creación, modificación de datos de la Big Data?
8	¿Existe visibilidad entre lo que realizan diferentes sectores de la organización?
9	¿Existe algún monitoreo o control el cual verifique cómo ocurren los procesos de colaboración?
10	¿Los empleados entienden en qué parte del proceso se encuentran y cómo colaboran con el objetivo final?
11	¿Existen planes de comunicación para poder integrar y fomentar la participación de todos los miembros del equipo?
Liderazgo	
12	¿Existe un líder que administre la integración con los procesos de negocio?
13	¿Cuál es el tipo de líder que se busca para la integración con los procesos de negocio?
14	¿Cuáles son las habilidades que busca en un líder encargado de la integración con los procesos de negocio?
Mejora continua	
15	¿Las personas (como individuos y/o grupos) inician y llevan a cabo las actividades y participan en el proceso de mejora continua?
16	¿Existe algún ciclo formal de búsqueda y resolución de problemas en la integración con los procesos de negocio?
17	¿Cuándo los procesos son finalizados son formalmente comunicados al personal?
18	¿Se verifican periódicamente los procesos con el fin de encontrar cambios o posibles mejoras?
19	¿Existe un mapeo de los procesos que cambiaron que ayude a los empleados a visualizar el nuevo flujo de trabajo?
Entendimiento y confianza	
20	¿Los miembros del grupo están al tanto de sus tareas y responsabilidades?
21	¿Se realiza una asignación de tareas adecuadas para que los miembros del equipo estén al tanto de sus responsabilidades y funciones?
22	¿Los individuos evalúan que los cambios a realizar estén en sintonía con los objetivos del sector o de la organización?
23	¿Los individuos y grupos monitorean / miden el impacto que tiene las tareas que realizan en los objetivos?

estratégicos o sectoriales?
-----------------------------

## Integración de datos maestros

<b>Habilidades HARD</b>	
1	¿Se ha mejorado la calidad de los datos maestros para soportar la analítica de Big Data?
2	¿Se utiliza Big Data para mejorar la calidad de los datos maestros?
3	¿Se ha establecido un programa de mejora de la calidad y consistencia de las claves (keys) para dar soporte al programa de gobierno de Big Data?
4	¿Se ha enriquecido los datos maestros a través de procesos de extracción en texto u otros datos no estructurados?
5	¿Se han considerado las políticas vigentes en las distintas redes sociales y su impacto en la integración con los datos maestros?
<b>Habilidades SOFT</b>	
Participación	
6	¿Existe colaboración entre sectores en distintos procesos de integración con los datos maestros de la Big Data?
7	¿Se han comunicado a la organización o a los diferentes sectores las acciones realizadas sobre los datos maestros de la Big Data?
8	¿Se han definido roles responsables de la integración con los datos maestros de la Big Data?
Liderazgo	
9	¿Existe un líder que administre la integración de los datos maestros de la Big Data?
10	¿Cuál es el tipo de líder que se busca para realizar la integración de los datos maestros de la Big Data?
11	¿Cuáles son las habilidades que busca en un líder encargado de la integración de los datos maestros de la Big Data?
Mejora continua	
12	¿Las personas (como individuos y/o grupos) inician y llevan a cabo las actividades y participan en el proceso de mejora continua?
13	¿Existe algún ciclo formal de búsqueda y resolución de problemas en la integración de los datos maestros de la Big Data?
14	¿Cuándo los procesos son finalizados son formalmente comunicados al personal?
15	¿Poseen algún ciclo formal de búsqueda y resolución de problemas de las fuentes utilizadas?
Entendimiento y confianza	

1 6	¿Los miembros del grupo están al tanto de sus tareas y responsabilidades?
1 7	¿Se realiza una asignación de tareas adecuadas para que los miembros del equipo estén al tanto de sus responsabilidades y funciones?
1 8	¿Los individuos evalúan que los cambios a realizar estén en sintonía con los objetivos del sector o de la organización?
1 9	¿Los individuos y grupos monitorean / miden el impacto que tiene las tareas que realizan en los objetivos estratégicos o sectoriales?

### Ciclo de vida de la información

<b>Habilidades HARD</b>	
1	¿Comprende las regulaciones legales que aplican a la retención de Big Data para cada tipo de industria y región pertinente?
2	¿Se lleva control y registro del volumen de almacenamiento para la Big Data?
3	¿Se lleva control del costo de almacenamiento de la Big Data?
4	¿Se comprime y archiva la Big Data para reducir los costos de TI y mejorar el rendimiento de las aplicaciones?
5	¿Comprende los requerimientos de negocio que manejan la retención de la Big Data?
6	¿Tiene un proceso para documentar retenciones legales en Big Data en base a litigios en curso?
7	¿Se gestiona el ciclo de vida de los datos capturados por streaming?
8	¿Se gobiernan adecuadamente los datos capturados a través de las redes sociales?
9	¿Se desecha de forma responsable la Big Data que ya no se requiere según las regulaciones y las necesidades del negocio?
<b>Habilidades SOFT</b>	
Participación	
1 0	¿Se ha definido un mecanismo de participación en el ciclo de vida de la Big Data?
1 1	¿Los sectores utilizan un ciclo de vida definido a nivel de organización o varía según el sector?
Liderazgo	
1 2	¿Existe un líder que administre el ciclo de vida de la Big Data?
1 3	¿Cuál es el tipo de líder que se busca para la administración del ciclo de vida de la Big Data?
1 4	¿Cuáles son las habilidades que busca en un líder encargado de la administración del ciclo de vida de la Big Data?

1	
5	¿Los líderes delegan responsabilidades referentes a la administración del ciclo de vida de la Big Data?
<b>Mejora continua</b>	
1	¿Las personas (como individuos y/o grupos) inician y llevan a cabo las actividades y participan en el
6	proceso de mejora continua?
1	
7	¿Existe algún ciclo formal de búsqueda y resolución de problemas en la ciclo de vida de la información?
1	
8	¿Cuándo los procesos son finalizados son formalmente comunicados al personal?
1	
9	¿Se registran y analizan cambios en los ciclos de vida definidos?
2	
0	Basados en políticas de retención de datos de la Big Data, ¿los usuarios purgan información innecesaria?
<b>Entendimiento y confianza</b>	
2	
1	¿Los miembros del grupo están al tanto de sus tareas y responsabilidades?
2	¿Se realiza una asignación de tareas adecuadas para que los miembros del equipo estén al tanto de sus
2	responsabilidades y funciones?
2	¿Los individuos evalúan que los cambios a realizar estén en sintonía con los objetivos del sector o de la
3	organización?
2	¿Los individuos y grupos monitorean / miden el impacto que tiene las tareas que realizan en los objetivos
4	estratégicos o sectoriales?

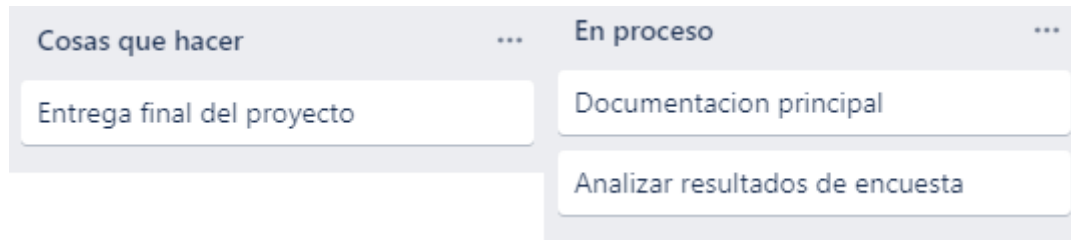
## Capacidad tecnológica

<b>Habilidades HARD</b>	
1	¿Se han identificado las necesidades de infraestructura para dar soporte a los proyectos de Big Data?
2	¿Se ha evaluado la disponibilidad de la solución de Big Data?
3	¿Se gestiona la infraestructura de archivos de la Big Data?
4	¿Se lleva registro y control de los cambios en infraestructura de la Big Data?
5	¿Se ha medido el impacto de la negación de servicio?
6	¿Existe un plan de recuperación de desastres?
7	¿Se mide y controla la performance de la infraestructura?
<b>Habilidades SOFT</b>	
<b>Participación</b>	
8	¿Se ha definido un mecanismo para evaluar la capacidad de tecnología de la Big Data?
9	¿Se ha realizado un análisis en conjunto sobre la capacidad tecnológica necesaria para brindar el servicio de Big Data?

Liderazgo	
1 0	¿Existe un líder que administre la capacidad tecnológica de la Big Data?
1 1	¿Cuál es el tipo de líder que se busca para la administración de la capacidad tecnológica de la Big Data?
1 2	¿Cuáles son las habilidades que busca en un líder encargado de la administración la capacidad tecnológica de la Big Data?
1 3	¿Los líderes delegan responsabilidades referentes a la administración de la capacidad tecnológica de la Big Data?
Mejora continua	
1 4	¿Las personas (como individuos y/o grupos) inician y llevan a cabo las actividades y participan en el proceso de mejora continua?
1 5	¿Existe algún ciclo formal de búsqueda y resolución de problemas de la capacidad tecnológica?
1 6	¿Cuándo los procesos son finalizados son formalmente comunicados al personal?
1 7	¿Se registran y analizan cambios en la capacidad tecnológica?
Entendimiento y confianza	
1 8	¿Los miembros del grupo están al tanto de sus tareas y responsabilidades?
1 9	¿Se realiza una asignación de tareas adecuadas para que los miembros del equipo estén al tanto de sus responsabilidades y funciones?
2 0	¿Los individuos evalúan que los cambios a realizar estén en sintonía con los objetivos del sector o de la organización?
2 1	¿Los individuos y grupos monitorean / miden el impacto que tiene las tareas que realizan en los objetivos estratégicos o sectoriales?

## Anexo 2 - Herramienta de gestión de tareas

Se adjuntan imágenes, a modo ilustrativo, del tablero de tareas utilizado a lo largo del proyecto.



Hecho		...	
Preparación y entrega de Carpeta de Anteproyecto	Apr 5	M	P
Preparar Entrega de avance de proyecto , ppt etc.	Jun 18		
Responder preguntas definidas con la investigacion			
Leer el material inicial			
Planificación inicial, definición de etapas del proyecto. Lectura,Búsqueda, Construcción del modelo,Validación del Modelo		HG	M P
Definir preguntas a responder con la investigacion			
Definir objetivo del proyecto			
Estado del Arte	0/4		P
Refinación de búsqueda en Google			
Definir como vamos a validar el modelo de madurez proyecto			
Gestión del Riegos		M	P
Selección de Herramientas para Gestión del proyecto	1	M	P
Lectura ISO 38500		M	P
Lectura Libro Big Data Governance de S. Soares		M	P
Criticas a la Evaluación de Madurez del Soares		M	P
Analizamos si en un empresa que tiene implementada ISO 38500 se puede implementar el Marco de Soares. Criticas a Soares con Foco en ISO 38500		M	P
Construcción del modelo de madurez			
Definición de Metodología de Trabajo (Primaria). Definición de búsquedas de bibliográfica a realizar		M	P
Lectura de material inicial ""The continuing mismatch between IT governance maturity theory and practice: a new approach,""		M	P
Búsquedas de informacion en Timbo (academica) sobre la temática		M	P
Búsquedas de información Google (industria) y Google académico sobre la temática		M	P
Proceso de análisis y refinamiento de material encontrado		M	P
Primer lectura de doc. 302		M	P
Esqueleto inicial de documento con algunos items ()		M	P
Refinación de búsquedas en Timbo		M	P
Refinamiento de artículos del Proyecto base		M	P
Busquedas en Timbo clasificadas por Autor		M	P
Lectura de artículos del Proyecto que tomamos como base		M	P

## Anexo 3- Evaluación

En las siguientes tablas se muestra las respuestas afirmativas marcadas con amarillo, de la organizacional evaluada

METADATA	
Nivel de Madurez	Características
1	<p>No existe una comprensión del valor de los metadatos.</p> <p>No existen o son inconsistentes las políticas, procesos, procedimientos o roles relacionados con la metadata</p> <p>No existen un glosario de datos definido o si existe es incompleto y de baja calidad.</p>
2	<p>Se definen algunas políticas, procesos, procedimientos a nivel de proyecto.</p> <p>Se han definido roles responsables de la metadata de la Big Data .</p> <p>Se han identificado las fuentes de Big Data de datos operacionales estructurados.</p> <p>Se ha definido en un glosario el significado en términos técnicos de la Big Data para datos estructurados.</p> <p>La glosario esta consolidado y disponible para algunos interesados.</p>
3	<p>Están instauradas y documentadas las políticas, proceso y procedimientos para la generación de la nueva metadata y mantenimiento de la existente.</p> <p>Hay una definición de arquitectura de datos incipiente.</p> <p>Los participantes comprenden claramente las responsabilidades asociadas con sus roles.</p> <p>La recopilación de metadatos sobre contenido estructurado se automatiza.</p> <p>Las políticas de metadata en torno al gobierno de Big Data están disponibles para todos los interesados.</p>
4	<p>La arquitectura de datos esta claramente definida.</p> <p>Se identifican, almacenan e incorporan al glosario de datos, los datos no estructurados.</p> <p>Las auditorías y validación regular de metadatos se adoptan como políticas oficiales de datos de la organización.</p>

	<p>Esta sincronizada y centraliza la metadata perteneciente a los distintos repositorios existentes de Big Data.</p> <p>Los metadatos se recopilan automáticamente de la mayoría de los sistemas de gestión de bases de datos relacionales y sistemas externos.</p> <p>Se ha definido en un glosario el significado en términos del negocio de la Big Data</p>
5	<p>La arquitectura de datos responde a los objetivos estratégicos de la empresa.</p> <p>Se crea un grupo de gestión de metadatos dedicado para avanzar estratégicamente en las capacidades de metadatos y aprovechar de manera más efectiva los metadatos existentes.</p> <p>Las políticas, proceso y procedimientos de metadatos cubre datos estructurados y no estructurados</p> <p>Se evalúa periódicamente si las políticas, procesos, procedimientos y roles siguen siendo compatibles y efectivos con las premisas u objetivos.</p> <p>Se toman acciones correctivas sobre las falencias encontradas en la evaluación.</p>

<b>CALIDAD DE DATOS</b>	
<b>Nivel de Madurez</b>	<b>Características</b>
1	<p>No existen políticas, practicas o procedimientos para la gestión de calidad de datos.</p> <p>La calidad de los datos no esta formalmente definida, es decir no esta claro el concepto de calidad de datos que tiene la organización.</p> <p>No están definidos los roles relacionados con la calidad de datos.</p> <p>Las personas realizan esfuerzos, poco frecuente, de calidad de datos, comúnmente ad hoc, según sea necesario y corrigen manualmente los problemas de datos identificados.</p>
2	<p>Las políticas, practicas o procedimientos se han definido para algunas actividades relacionadas con la calidad de los datos y se han seguido de manera inconsistente.</p> <p>Un pequeño grupo de individuos es capacitado y realiza perfiles para evaluar la calidad de los datos , para luego establecer una línea de base a ser utilizada en proyecto.</p> <p>Se han identificado los datos críticos que aseguran el éxito del programa de Big Data en la organización</p> <p>Los perfiles definidos para la calidad de datos están disponibles para su uso en cualquier parte del</p>

	ciclo de vida de la Big Data.
3	<p>Están instaurado y documentados las políticas, practicas o procedimientos sobre perfiles de calidad de datos y el desarrollo de estándares.</p> <p>Se han definido todos los roles para garantizar la calidad de los datos y sus responsabilidades están claras.</p> <p>Se definen métricas de calidad e informes de calidad.</p> <p>La calidad de datos se convierte un actividad programada en los planes del proyecto.</p>
4	<p>Las políticas de calidad de datos se convierten en políticas oficiales de la organización y se audita el cumplimiento de las mismas.</p> <p>Los expertos en calidad de datos se identifican en toda la organización y participan en todos los proyectos de mejora de la calidad de datos.</p> <p>Para los datos críticos identificados se ha definido la matriz de calidad de datos <i>Elemento de dato - A qué y cómo afecta la falta de calidad - Responsable</i></p> <p>Se controlan los resultados de las métricas de calidad.</p> <p>La solución de problemas de calidad de datos está integrada en la plataforma de informes de calidad.</p>
5	<p>Se evalúa periódicamente si las políticas, procesos, procedimientos y roles siguen siendo compatibles y efectivos con las premisas u objetivos.</p> <p>Se utilizan indicadores para dimensionar y dar forma a la mejora continua.</p> <p>Se realiza un seguimiento y se informa de forma centralizada el cumplimiento de la calidad de los datos oficiales de la organización.</p> <p>Las correcciones de la calidad de los datos se implementa tanto en datos en reposo (en bases de datos) como en datos en vuelo (en ETL y como mensajes entre sistemas).</p>

CAPACIDAD TECNOLÓGICA	
Nivel de Madurez	Características
1	<p>No existen políticas, practicas o procedimientos relacionados con la capacidad tecnológica.</p> <p>Poco conocimiento de las tecnologías de Big Data.</p> <p>No se han identificado las necesidades de infraestructura para dar soporte a los proyectos de Big</p>

	<p>Data.</p> <p>No se tiene conciencia de la importancia de las capacidades tecnológicas de la Big Data.</p> <p>No están definidos los roles relacionados a la tecnología – HAY UN ROL DE TI</p>
2	<p>Se han definido roles responsables del tecnología para la Big Data.</p> <p>Se han identificado las necesidades de infraestructura para dar soporte a los proyectos de Big Data, pero sin estar estandarizado.</p> <p>Se definen algunas políticas entorno a las capacidad tecnológica, pero inconsistente.</p> <p>Se comprende la importancia de la tecnología para el éxito en la gestión de la Big Data.</p>
3	<p>Están estandarizadas y documentados las políticas, practicas o procedimientos que contribuyen a la gestión de las capacidad tecnológica.</p> <p>Hay una definición de arquitectura de tecnología incipiente.</p> <p>Se ha evaluado la disponibilidad de la solución de Big Data .</p> <p>Existe un plan de recuperación de desastres.</p> <p>Las soluciones tecnológicas se adoptan como mejores prácticas y se ponen a disposición en toda la institución</p>
4	<p>La arquitectura de tecnología esta claramente definida.</p> <p>Se lleva registro y control de los cambios en infraestructura de la Big Data.</p> <p>Se mide y controla el impacto de la negación de servicio.</p> <p>Se mide y controla la performance de la infraestructura .</p> <p>Se gestiona la infraestructura de archivos de la Big Data .</p>
5	<p>La arquitectura de tecnología responde a los objetivos estratégicos de la empresa.</p> <p>Todas las disciplinas y aspectos de la Big Data y su gobernanza tienen una solución tecnológica disponible.</p> <p>Se gestiona totalmente las capacidades tecnológicas.</p> <p>Se evalúa periódicamente si las políticas, procesos, procedimientos y roles, siguen siendo compatibles y efectivos con las premisas u objetivos.</p> <p>Se utilizan indicadores para dimensionar y dar forma a la mejora continua.</p>

## Anexo 4 – Carta de la empresa evaluada

**EFA LABORATORIOS**  
ANTIA, MOIL & CIA. S.A.

Montevideo, 10 de setiembre de 2019

A quien corresponda  
Universidad ORT Uruguay

Como parte de los objetivos de calidad que han distinguido a *EFA Laboratorios S.A.* se encuentra el proyecto de migración tecnológica que desde 2018 el laboratorio viene desarrollando.

En este contexto nos comprometemos a participar del proyecto "*Modelo de madurez para la gobernanza de Big Data*" desarrollado en la Universidad ORT por estudiantes de la carrera Licenciatura en sistemas, dirigidos por la Dra. Helena Garbarino. Se brindará información relevante – previamente acordada – sobre el desarrollo del proyecto "*Estabilidad de mezclas*" con el objetivo de obtener recomendaciones de mejora.

Saluda atentamente,

  
Q. F. Joaquin Torterolo  
Químico Analista