**ETAPA 1**

**ACTA DE CONSTITUCIÓN DEL PROYECTO INFORMÁTICO**

Escuela de Informática y Telecomunicaciones

Julio 2025

1. Identificación del Proyecto

| **Nombre de Proyecto** |
| --- |
| DataGov Live |

1. Integrantes del Equipo de Trabajo

| **Nombre estudiante** | Ramirez, Sebastian. |
| --- | --- |
| **Rut** | 20.237.961-3 |
| **Carrera** | Ingeniería Informática. |
| **Sede** | Duoc UC sede Padre Alonso Ovalle. |

| **Nombre estudiante** | Vargas, Rodrigo. |
| --- | --- |
| **Rut** | 20.744.660-2 |
| **Carrera** | Ingeniería Informática. |
| **Sede** | Duoc UC sede Padre Alonso Ovalle. |

1. Registro de Control de Cambios

| **Revisión** | **Fecha** | **Páginas** | **Descripción del Cambio** | **Autor** |
| --- | --- | --- | --- | --- |
| 1 | 26/08/2025 | Todas | Primera Revisión del Documento | RV |
| 2 | 29/08/2025 | Todas | Segunda Revisión del Documento viendo el Feedback | RV |

# 

# ÍNDICE DE CONTENIDO

[ÍNDICE DE CONTENIDO 3](#_heading=h.xvbr31hz63ln)

[Glosario 4](#_heading=h.3znysh7)

[Abstract 6](#_heading=h.gbogjisnrs35)

[Declaración del Proyecto Tecnológico 7](#_heading=h.tyjcwt)

[Definir la problemática a solucionar o necesidad a satisfacer 8](#_heading=h.3dy6vkm)

[Relación del proyecto APT con las competencias del perfil de egreso. 9](#_heading=h.2ye19nx1ko6y)

[Relación del proyecto APT con tus intereses profesionales. 10](#_heading=h.72uagzdim64t)

[Argumento sobre la factibilidad del proyecto dentro de la asignatura. 10](#_heading=h.3wbuguhv40qr)

[Toma de roles 11](#_heading=h.t5f3mfwnnctq)

[Definición del tiempo asociado al desarrollo del Proyecto Tecnológico. 12](#_heading=h.4d34og8)

[Gestionar los Interesados del Proyecto (gestionar la comunicación entre el cliente y el equipo de desarrollo del Proyecto Tecnológico - Canales de comunicación). 13](#_heading=h.jz193kez297u)

[Metodología implementada 14](#_heading=h.tr3vaomtyfw)

[BIBLIOGRAFÍA 15](#_heading=h.44sinio)

# Glosario

**PALABRA:** Significado de la palabra o acrónimo.

**QA:** Quality Assurance. En español, Analista de Calidad. Tiene como objetivo principal realizar pruebas a soluciones tecnológicas en desarrollo por medio planes de prueba y diversas herramientas de testeo, que permiten asegurar un buen funcionamiento libre de errores.

**ARQUITECTURA TECNOLÓGICA:** Es un conjunto de procesos que se unen para formar una estructura tecnológica, de la cual depende el proyecto.

**SPA**: Sociedad por Acciones.

**E-COMMERCE**: Conocido como comercio electrónico; consiste en la distribución, compra, venta, marketing y suministro de productos o servicios a través de internet.

**INFRAESTRUCTURA CLOUD**: Conocida como la infraestructura informática en la nube; es la recopilación de elementos de hardware y software necesarios para hacer posible la informática de nube

**TI** : Tecnologías de la Información.

**EDT**: Representación gráfica, descriptiva, de las tareas que permitirán concretar el proyecto.

**FRONT END**: El **frontend** es la parte del desarrollo web que se dedica a la parte frontal de un sitio web, en pocas palabras del diseño de un sitio web

**INTERFAZ**: Dispositivo capaz de transformar las señales generadas por un aparato en señales comprensibles por otro

**PLATAFORMA CHATBOT:** Programa de software que se ocupa de mantener conversaciones con usuarios a través de mensajes de texto, mejorando el servicio de atención al cliente.

**BACKEND:** Se define como sistema corporativo que se utiliza para dirigir una web o empresa, tales como sistemas de gestión de pedidos, inventario y procesamiento de suministro, este recoge información de los usuarios u otros sistemas de tratamiento de datos en la compañía.

**FRONTEND:** Este sería la parte de un dispositivo o programa que permite a un usuario acceder directamente. Son todas las tecnologías de diseño y

desarrollo web que corren en el navegador y que se encargan de la interactividad con los usuarios.

**SCRUM**: Es un marco de trabajo para desarrollo ágil de software, este se considera un proceso en el que se aplican de manera regular un conjunto de buenas prácticas para trabajar colaborativamente, en equipo y obtener el mejor resultado posible de proyecto.

**SPRINTS:** Nombre que va a recibir cada uno de los ciclos o iteraciones que vamos a tener dentro de un proyecto según la metodología de Scrum.

**SCRUM MASTER:** Se define como el encargado de llevar la jefatura del proyecto, dándole el nombre de líder de proyecto.

**SCRUM TEAM:** Es un pequeño equipo de personas que incluye un Scrum Master, un Product Owner y los Developers, los cuales estarán compuestos por profesionales multifuncionales con las habilidades necesarias para cumplir el objetivo del producto final.

**PRODUCT OWNER:** Éste será la persona encargada de dirigir el equipo hacia el mejor resultado posible, porque conoce muy bien el producto final.

**PHP:** lenguaje de programación interpretado​ del lado del servidor y de uso general que se adapta especialmente al desarrollo web

**OPEN METADATA:** Plataforma open-source para gestión de metadatos y gobiernos de datos.

**METADATA:** “Datos sobre los datos” - información que describe otros datos.

**DATA GOVERNANCE:** Conjunto de procesos para garantizar calidad, seguridad y disponibilidad de datos.

**DATA CATALOG:** Catálogo que organiza y documenta los activos de datos de una organización.

**MVP (MINIMUM VIABLE PRODUCT):** Producto mínimo viable con features esenciales.

# Abstract

An improvement was requested for the data governance project, which was implemented based on Open MetaData. The scope of the current project is explained below, detailing the creation of a system that allows companies to access their data systems to determine where the data is located, who owns it, is it up-to-date, and where it belongs. The system improves metadata governance in IT services companies or SMEs, improving management and automating processes.

The project is structured in modules:

Metadata Catalog Module Interface: Provides semantic search and advanced filters, unified registration of all data assets, automatic structure analysis, and statistics.

Data Quality Module Interface: Configurable rules engine management, scheduled execution of quality tests, and detection of deviations and outliers.

Data Lineage Module Interface: Automatic data provenance tracing, visualization of dependencies, and impacts.

Governance Module Interface: Creation and management of governance policies, automatic verification of regulations, complete logging of activities and changes.

This system is essential for improving the efficiency and management of enterprise metadata. The project objectives are to provide a high-quality and effective product that benefits processes.

Other highlights include the progress of project management, highlighting the following stages already executed: Planning and preparation stage; Requirements specification; Open MetaData core architecture; Diagramming.

There are three different methodologies for its execution; however, for this project, it was decided to use the agile methodology for technical development (installation, configuration, integration with data sources). This was followed by the Data Governance Knowledge Framework (DAMA-DMBOK), based on security, quality, and privacy. The implementation methodology was phased (roadmap), which details the phases, objectives, and key deliverables. Detailed follow-up was provided for each phase.

After analyzing these methodologies, another positive aspect achieved was that, for our project, clients required a short follow-up period.

# Declaración del Proyecto Tecnológico

El proyecto busca implementar Open MetaData dentro de una empresa, que maneje grandes volúmenes de información sobre datos, producción, logística, exportación y distribución a nivel nacional e internacional.

El problema principal es que los datos actualmente pueden estar dispersos en distintas fuentes (sistemas, control de calidad, ERP, sensores, registros, bases de datos de exportación, hojas de excel)

Esto genera:

* Duplicación de información.
* Datos inconsistentes.
* Dificultades para acceder rápidamente a la información crítica.
* Decisiones basadas en datos incompletos.

Transforma la gestión de datos en una ventaja competitiva internacional, garantizando que la empresa pueda crecer, competir con seguridad y mantener una gobernanza de datos.

**DataGov Lite** es una plataforma de gobierno de datos simplificada construida sobre Open Metadata, diseñada para que empresas medianas y PYMES puedan implementar gobernanza de datos sin necesidad de grandes inversiones en infraestructura o personal especializado.

**Problema que Resuelve**

Las empresas hoy generan datos constantemente pero:

* No saben qué datos tienen ni dónde están
* Los equipos trabajan con versiones contradictorias de los mismos datos
* No existe documentación consistente sobre el significado de los datos
* Los errores en datos se detectan demasiado tarde

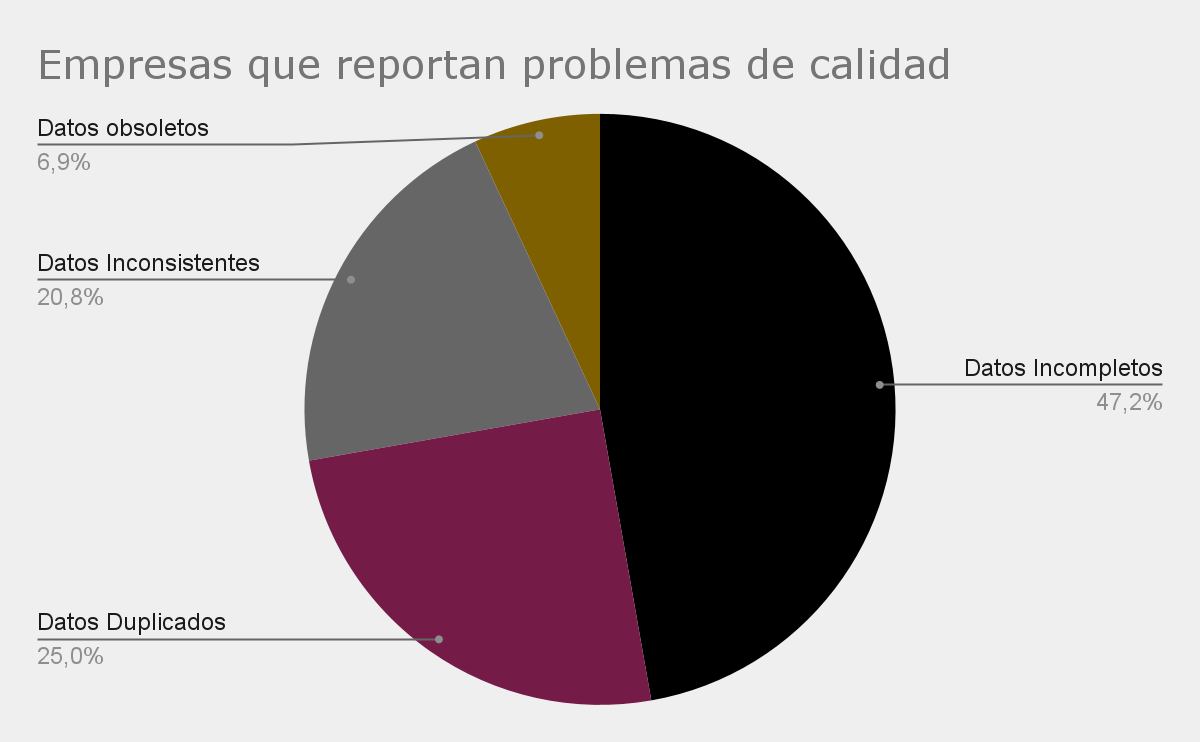
**Propósito Fundamental**

democratizar el acceso al gobierno de datos, haciendo que cualquier organización pueda:

* Descubrir qué datos tiene y dónde están localizados
* Entender el significado y contexto de sus datos
* Confiar en la calidad y confiabilidad de su información
* Colaborar alrededor de los activos de datos compartidos

# Definir la problemática a solucionar o necesidad a satisfacer

Problemas de calidad de datos:



El mundo sufre una crisis silenciosa de gobernabilidad de datos. Las empresas generan datos masivamente pero navegan en un caos donde nadie sabe qué datos tienen, dónde están ni si son confiables. Esta anarquía data cuesta billones en pérdidas anuales por decisiones equivocadas, oportunidades perdidas y multas regulatorias.

Los equipos técnicos y de negocio trabajan en silos, usando versiones contradictorias de la misma información, mientras el 50% del tiempo de los analistas se pierde buscando y limpiando datos en lugar de generar valor. Las soluciones existentes son complejas, costosas y requieren implementaciones de meses, haciendo que el gobierno de datos sea inaccesible para la mayoría de las empresas.

Esta problemática crea una brecha digital donde sólo las grandes empresas con recursos abundantes pueden permitirse el lujo de gobernar sus datos, mientras las PYMES se quedan atrás, tomando decisiones a ciegas en un mundo cada vez más data-driven. La consecuencia es una economía menos eficiente, innovación más lenta y servicios que podrían ser mejores si estuvieran basados en datos confiables.

# 

# Relación del proyecto APT con las competencias del perfil de egreso.

La relación del proyecto con las competencias del perfil de duoc son en base a los criterios pedidos, se seguirá la rúbrica y las normativas de lo que debería contener el proyecto. Basándose en el diseño de pruebas de validación tanto de los productos como de los procesos utilizando buenas prácticas definidas por la industria. Aplicar pruebas de validación tanto de los productos como de los procesos utilizando buenas prácticas definidas por la industria. Desarrollar mejoras al producto en base al resultado de las mismas.

Se Gestionará el proyecto informáticos, ofreciendo alternativas para la toma de decisiones de acuerdo a los requerimientos de la organización. La competencia en este caso se llevará a cabo ofreciendo una solución con Open MetaData intentando crear un sistema para empresas con grandes flujos de datos. Para darle relevancia al proyecto y para ver la capacidad del equipo de trabajo.

También como lo indica el perfil se construirán modelos de datos para soportar los requerimientos de la organización de acuerdo a un diseño definido y escalable en el tiempo. Se integrarán bases de datos funcionales para el prototipo del sistema que se llevará a cabo para las pruebas.

Por último da paso al desarrollo de una solución de software utilizando técnicas que permitan sistematizar el proceso de desarrollo y mantenimiento, asegurando el logro de los objetivos.

No se le entregará una mayor relevancia a la mantención porque no es será necesario, sin embargo, se le dará más relevancia a las pruebas unitarias y la calidad del software o capacidad de funcionamiento más que de prototipo.

# 

# 

# 

# 

# 

# Relación del proyecto APT con tus intereses profesionales.

# 

En base a mis intereses propios se adecua medianamente, por el hecho de la protección de datos, en base a mis intereses el proyecto es viable y cercano a la mano de ciberseguridad o seguridad de la información, al ser Open MetaData una herramienta que busca resolver la pérdida y disrupción de meta datos en el ciberespacio empresarial.

**Potencial**: Se relaciona mutuamente en mi interés por la ciberseguridad y el manejo de la información.

# 

# 

# Argumento sobre la factibilidad del proyecto dentro de la asignatura.

Factibilidad de Adopción, Viabilidad Media

Se puede analizar la factibilidad con la gráfica que definen la problemática, ya que, existe globalmente, una demanda de baja calidad debido a la pérdida de datos y un gran desorden tanto, empresarial, como personal, solamente un 85% de los datos perdidos o inconsistentes se encuentran en PYMES, y actualmente solo un 15% de las empresas tienen una gestión de calidad de sus datos

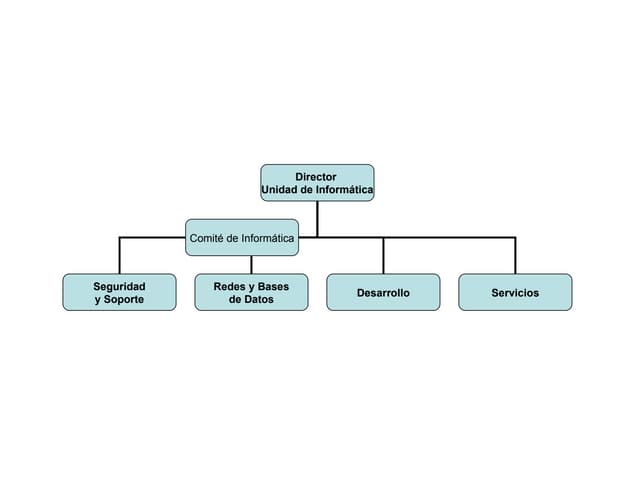
* Beneficio inmediato: Usuarios ven valor desde día 1
* Low-entry barrier: Interface simple y intuitiva
* Cambio gradual: No requiere transformación radical
* Training mínimo: Documentación embedded en la plataforma

Factibilidad Alumno, Viabilidad Positiva

El alumno conocerá una nueva herramienta y su forma de uso que es Open MetaData además de conocer su implementación y forma de uso. Además de las tecnologías estándar (docker, python, APIs REST), documentación extensa sobre O MD y recursos gratuitos que puede usar el alumno.

# Toma de roles

| **Nombre** | **Rol** | **Responsabilidad** | **Observación** |
| --- | --- | --- | --- |
| Ramirez, Sebastian | Jefe Proyecto | Guiar equipo de trabajo |  |
| Vargas, Rodrigo. | Arquitectos de Cloud | Gestión servicios cloud |  |
| Vargas, Rodrigo. | Ingeniero Informático | Desarrollo y gestión de software |  |
| Ramirez, Sebastian | Ingeniero Devops | Encargado de seguridad y ciberseguridad en el desarrollo de software |  |
| Vargas, Rodrigo. | Q.A | Asegurar la calidad de los productos desarrollados |  |
| Vargas, Rodrigo. | Diseñador | Establecer patrones y diseños para el software a desarrollar |  |
| Ramirez, Sebastian | Especialista de base de datos | Diseño y gestión base de datos |  |



# 

# Definición del tiempo asociado al desarrollo del Proyecto Tecnológico.

**Fase 1:**

Se estima un tiempo de dos semanas de levantamiento funcional (Open Metadata y arquitectura del software), también se estiman 2.6 meses de desarrollo para la implementación del sistema, un mes de QA y una semana de marcha blanca.

* Desglose por tipo de trabajo:
* Backend development: 4 semanas.
* Frontend development: 3 semanas.
* Integraciones: 2 semanas.
* Testing & deployment: 1 semana.

**Fase 2:**

* Módulo principal: tiempo estimado 7 días.
* Módulo Metadata Catalog Module Interface: tiempo estimado 14 días.
* Módulo Data Quality Module Interface: tiempo estimado 11 días.
* Módulo Data Lineage Module Interface: tiempo estimado de 5 días.
* Módulo Governance Module Interface: tiempo estimado 8 días.

**Fase 3:**

En resumen en los módulos en general se necesitarán alrededor de 2 meses asociados a su correcta ejecución.

* Sprint 1: Setup + Dashboard básico.
* Sprint 2: Alertas + 2 conectores.
* Sprint 3: Lineage + Colaboración.
* Sprint 4: Calidad avanzada.
* Sprint 5: Optimización + Deployment.

# Gestionar los Interesados del Proyecto (gestionar la comunicación entre el cliente y el equipo de desarrollo del Proyecto Tecnológico - Canales de comunicación).

**Canales de comunicación**

Canales de comunicación informal:

Usado principalmente para la comunicación interna de la empresa y para emergencias. Dentro de esta categoría se encuentran:

* Llamadas telefónicas
* Mensajes de texto
* Redes sociales
* Videollamadas

Canales de comunicación formal:

Utilizado generalmente para la comunicación entre la empresa y el cliente, generalmente para establecer reuniones, resolver problemas, dejar constancia de lo hablado. El canal establecido es correo electrónico.

# 

# 

# Metodología implementada

En este caso más enfocado en analitica avanzada, es usar Scrum para el *desarrollo Técnico*, ***DAMA-DMBOK*** para la gobernanza de datos, y un *roadmap* de implementación en fases.

Metodología Ágil.

Para el desarrollo técnico (instalación, configuración, integración con fuentes de datos).

* Scrum ya que el equipo es pequeño y se puede trabajar en sprint de 2-3 semanas.
* Ventajas:
  + Entregas rápidas y visibles (semanalmente).
  + Permite ir probando Open MetaData en ambientes controlados antes de expandir.

Data Governance (***DAMA-DMBOK***).

* No solo es instalar OMD, se trata de definir roles de responsabilidad sobre los datos.
* DAMA-DMBOK
  + Calidad de datos
  + Seguridad y privacidad
  + Metadatos estándar

Metodología de implementación por fases (Roadmap de madurez de los datos)

* Fase 1: Piloto en un área → por ejemplo, logística de exportación.
* Fase 2: Ampliación del inventario y producción.
* Fase 3: Integración BI y reporte financiero.
* Fase 4: Gobierno de datos transversal de la empresa.

# BIBLIOGRAFÍA

Se adjuntará el material o fuentes que ocuparon, para dar a conocer de dónde viene la información del documento entregado “1.4\_AP T122 Constitución Fase1”:

* Lenguajes y librerias: <https://books.goalkicker.com/>
* Docs Oficiales:<https://docs.open-metadata.org/>
* Quick Start:<https://docs.open-metadata.org/quick-start>
* Docker Deployment:<https://docs.open-metadata.org/deployment/docker>
* Repositorio Principal: https://github.com/open-metadata/OpenMetadata
* SDK Python: https://github.com/open-metadata/openmetadata-sdk
* Ingestión Framework:https://github.com/open-metadata/ingestion-framework
* API REST Docs:<https://docs.open-metadata.org/api>
* Swagger UI: http://localhost:8585/swagger (cuando esté corriendo)
* Video Tutorials: https://www.youtube.com/c/OpenMetadata
* Blog Técnico: https://blog.open-metadata.org/
* Webinars: https://www.open-metadata.org/events
* Hands-on Labs: https://docs.open-metadata.org/tutorials