

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
VALPARAÍSO - CHILE



**“METODOLOGÍA PARA EL DISEÑO DE UN
MODELO DE GOBERNANZA DE DATOS EN
INSTITUCIONES EDUCATIVAS ESCOLARES”**

ANGELO CABRERA MONTENEGRO

TESINA PARA OPTAR AL GRADO
DE MAGÍSTER EN TECNOLOGÍAS DE
LA INFORMACIÓN

Profesor Guía: JOSE LUIS MARTÍ
Director de tesina: RAÚL MONGE



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☐ Memoria o trabajo de título ☒ Tesis de Postgrado

Título del trabajo: Metodología para el diseño de un modelo de Gobernanza de Datos en instituciones educativas escolares

Nombre del candidato(a): Angelo Cabrera Montenegro

Carrera / Grado: Magister en Tecnologías de la Información

Campus: San Joaquín **Departamento:** Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, José Luis Martí Lara, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

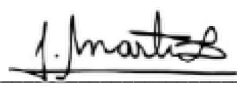
☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

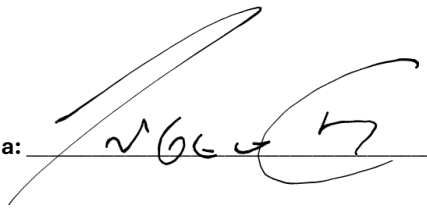
Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 20 / 10 / 2026 Firma: 

Estudiante o Candidato(a):

Fecha: 20/10/2026 Firma: 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.



Metodología para el diseño de un modelo de Gobernanza de Datos en instituciones educativas escolares

Angelo Cabrera Montenegro

The Butterfly 28, Villa alemana

angelo.cabrera.14@sansano.usm.cl

Resumen: Los procesos de transformación digital en instituciones educativas generan grandes volúmenes de datos que, sin una gobernanza de datos adecuada, suelen ser dispersos, inconsistentes y subutilizados, afectando la gestión escolar y el cumplimiento normativo. Esta tesina aborda dicho problema mediante Investigación Basada en Diseño (*Design Science Research*, DSR) y un caso de estudio en una fundación educativa chilena. El objetivo fue evaluar el impacto de la implementación piloto de un modelo de gobernanza de datos, diseñado a medida para el contexto escolar, sobre la satisfacción del equipo directivo y gerencial respecto a la disponibilidad y calidad de los datos utilizados para el seguimiento académico y formativo. Los resultados obtenidos demuestran que la aplicación del modelo generó una mejora sustancial, registrando un incremento del 66,53% en la satisfacción del equipo directivo. Esto valida el artefacto propuesto y entrega lineamientos prácticos para optimizar la gestión de datos en instituciones educativas escolares, considerando siempre la seguridad y privacidad.

Palabras Clave: Gobernanza de datos, calidad de datos, big data, educación, transformación digital

1 Introducción

Este capítulo presenta el contexto y la motivación de la investigación, delimita el problema abordado, formula los objetivos general y específicos junto con la hipótesis de trabajo, y describe la organización de la presente tesina.

1.1 Contexto y motivación

La incorporación de tecnología en las escuelas ha llevado a la digitalización de una gran cantidad de datos relativos al desempeño y aptitudes de sus estudiantes, así como también de los docentes y equipos administrativos. Dado esto, se vuelve cada vez más relevante que las instituciones educativas mejoren su gestión de datos y utilicen la información para tomar decisiones más informadas y efectivas.

En este contexto, al existir diversas plataformas tecnológicas que generan y almacenan sus datos, éstos generalmente se encuentran dispersos y en diferentes formatos dependiendo del origen del sistema (libro de clases, control de asistencia de administrativos, sistemas de contabilidad y finanzas, aplicaciones de evaluación académica, entre otros) y de cómo los dueños de cada proceso recopilan y analizan dichos datos. Esto lleva a que los datos se puedan encontrar incompletos, duplicados o en diversos formatos, generando errores o inconsistencias que dificultan que la información presentada a la comunidad educativa sea veraz y de calidad, afectando la toma de decisiones que permitan mejorar el desempeño de los estudiantes y de la institución en su totalidad.

1.2 Planteamiento del problema

Para poder responder a estas dificultades, se hace necesario contar con un modelo de gobernanza de datos que permita asegurar a las escuelas que los datos sean oportunos, confiables, seguros y protegidos contra el mal uso y amenazas de ciberseguridad, cumpliendo con las leyes y normas vigentes. De esta forma se puede desarrollar un programa que involucre políticas, estándares, definición y administración de datos, reportes y los procesos asociados.

El principal impacto que una buena implementación de gobernanza de datos puede tener en el ámbito educativo tiene directa relación con una mejora en el desempeño de los estudiantes. El poder contar con información relevante, de calidad y en forma oportuna permite a los docentes y directivos tomar medidas adecuadas que ayuden a su desarrollo. Por ejemplo, al tener datos precisos sobre el progreso individual de cada estudiante, los docentes pueden adaptar sus estrategias de enseñanza, brindar apoyo personalizado y realizar intervenciones tempranas cuando sea necesario. Además, la información confiable y precisa que se entrega a los apoderados y tutores facilita su participación en el proceso educativo. Finalmente, una buena gobernanza de datos permite una mejor gestión educativa en general, permitiendo a los equipos administrativos gestionar a la planta docente de forma eficiente.

1.3 Objetivos

El objetivo general del trabajo es diseñar, implementar y evaluar un modelo de gobernanza de datos adaptado a las condiciones organizacionales, normativas y de recursos de las instituciones educativas escolares chilenas, con el fin de mejorar la disponibilidad y la calidad de la información que sustenta el seguimiento académico y formativo de los estudiantes por parte de los equipos directivos y gerenciales.

En cuanto a los objetivos específicos, éstos son:

- Caracterizar el estado inicial de la gobernanza de datos en la institución caso de estudio, mediante un instrumento cuantitativo de percepción, a fin de establecer una línea base objetiva contra la cual contrastar el impacto de la intervención.
- Comparar los marcos de referencia, arquitecturas y metodologías disponibles en la literatura, a fin de seleccionar de manera fundamentada aquellos componentes aplicables a un contexto escolar con recursos limitados.
- Diseñar un modelo de gobernanza de datos que articule estructura organizativa, procesos y plataforma tecnológica, a fin de entregar a la institución un artefacto operable sin requerir la creación de nuevos cargos ni inversiones prohibitivas.
- Implementar el modelo mediante un piloto controlado en los dominios de seguimiento académico y formativo, a fin de demostrar su factibilidad técnica y organizacional en un entorno real.
- Evaluar el impacto del piloto sobre la satisfacción del equipo directivo y derivar lineamientos de replicabilidad, a fin de validar la hipótesis de investigación y permitir la transferencia del modelo a instituciones de características similares.

1.4 Hipótesis

Para validar la propuesta, se planteó la siguiente hipótesis:

La implementación piloto de un modelo de Gobernanza de Datos adaptado generará un incremento de al menos un 20% en la satisfacción del equipo directivo reportada respecto a la disponibilidad y calidad de los datos para el seguimiento académico estudiantil.

El umbral del 20% se definió como un criterio de significancia práctica que representa una mejora perceptible y operativamente relevante para el equipo directivo, evitando declarar válida la hipótesis ante variaciones marginales que podrían atribuirse al azar o al efecto novedad.

1.5 Estructura del informe

El presente documento se organiza en los siguientes capítulos: El capítulo 1 presenta el contexto, la problemática, los objetivos y la hipótesis de trabajo. El capítulo 2 desarrolla el marco conceptual, revisando los marcos de referencia de gobernanza de datos, modelos de madurez, arquitecturas analíticas y el contexto normativo chileno. El capítulo 3 detalla el diseño de la solución, abordando las tres primeras etapas de la metodología DSR: la selección fundamentada de componentes, la identificación del problema y línea base inicial, la definición de requerimientos y la propuesta del modelo GDME junto con su especificación técnica. El capítulo 4 describe las etapas de demostración y evaluación del piloto en el entorno real, presentando los resultados post-test, la verificación de la hipótesis y los lineamientos de replicabilidad. Finalmente se exponen las conclusiones del estudio, las limitaciones del trabajo y las líneas futuras de investigación, cerrando con las referencias bibliográficas y los anexos correspondientes.

2 Marco conceptual

Este capítulo revisa el cuerpo teórico y las alternativas metodológicas y tecnológicas disponibles para abordar un proyecto de gobernanza de datos. Para cada ámbito de decisión se presentan al menos dos o tres opciones vigentes en la literatura y la industria, junto con sus fortalezas y limitaciones. La selección fundamentada de las alternativas efectivamente utilizadas se realiza en el capítulo 3.

2.1 Gobernanza de datos: definición y delimitación conceptual

La gobernanza de datos se define como el ejercicio de autoridad y control sobre la gestión de activos de datos [1]. Si bien todas las organizaciones toman decisiones sobre los datos, aquellas que implementan un programa de gobernanza pueden aumentar el valor que obtienen de sus activos de datos.

Es necesario diferenciar los conceptos de gobernanza y gestión. El concepto de gobernanza hace referencia a las decisiones que se deben tomar para asegurar una gestión efectiva respecto al uso de tecnologías de información y comunicación (TIC) y a quién toma estas decisiones, a diferencia del concepto de gestión, que involucra la implementación de dichas decisiones [2]. En el caso específico de la gobernanza de datos, la gobernanza permite asegurar que la información se gestiona adecuadamente, mientras que la gestión consiste en la administración de los datos para lograr los objetivos de la organización.

2.2 Marco de referencia para la gobernanza de datos

Para gestionar adecuadamente la gobernanza de datos existen marcos de trabajo desarrollados específicamente para este propósito, que incluyen procesos, políticas, métodos y estructuras que permiten que quienes realizan procesos, optimizaciones, recolección, uso y publicación de datos puedan hacerlo adecuadamente sin errores y resguardando la privacidad de los datos [3]. A continuación se revisan cuatro alternativas de amplia adopción.

2.2.1 DAMA-DMBOK2

El *Data Management Body of Knowledge* (DAMA-DMBOK2), desarrollado por DAMA International, constituye el estándar de facto en gestión de datos a nivel global [1]. Organiza la disciplina en once áreas de conocimiento articuladas en torno a la *DAMA Wheel*:

- Gobernanza de Datos: Dirección y supervisión central.
- Arquitectura de Datos: Modelos y planos de datos integrados.
- Modelado y Diseño de Datos: Estructuras físicas y lógicas.
- Almacenamiento y Operaciones: Administración de repositorios de datos.
- Seguridad de Datos: Autenticación, autorización y privacidad.
- Integración e Interoperabilidad: Movimiento y transformación (ETL/ELT).

- Documentos y Contenido: Gestión de datos no estructurados.
- Datos Maestros y de Referencia: Unificación de entidades clave (MDM).
- Almacenamiento Analítico (Data Warehousing/BI): Soporte a la analítica.
- Gestión de Metadatos: Catálogo y linaje de datos.
- Calidad de Datos: Medición y mejora continua de dimensiones de calidad.

Su principal fortaleza es su exhaustividad conceptual. No obstante, DAMA-DMBOK2 es una guía descriptiva más que prescriptiva; no entrega una ruta operativa de implementación paso a paso y su envergadura puede resultar inabordable para organizaciones pequeñas o instituciones escolares con equipos reducidos.

2.2.2 COBIT

Desarrollado por ISACA, COBIT (*Control Objectives for Information and Related Technologies*) se enfoca en el gobierno corporativo de TI. Separa taxativamente el gobierno de la gestión. Ofrece un sólido alineamiento entre los objetivos de negocio y la gestión de riesgos tecnológicos, pero presenta un alto costo de adopción y una orientación hacia auditoría y cumplimiento normativo corporativo, resultando sobredimensionado para el alcance acotado de datos en escuelas [4].

2.2.3 DGI Data Governance Framework

Propuesto por el *Data Governance Institute*, se centra en la aplicación práctica de la gobernanza. Modela la iniciativa en 10 componentes universales clasificados en tres ejes: reglas (políticas y estándares), personas (comité, *data owners*, *data stewards*) y procesos (garantía de calidad, resolución de conflictos). Facilita la adopción ágil en etapas iniciales, pero cuenta con un menor respaldo académico y una cobertura acotada de la dimensión técnica de arquitectura [5].

2.2.4 ISO/IEC 38505-1

Norma internacional derivada del estándar ISO/IEC 38500 para el gobierno de TI. Orienta a los miembros del directorio y la alta dirección en la evaluación, dirección y monitoreo (EDM) del uso actual y futuro de los datos. Aplica los seis principios de gobierno (responsabilidad, estrategia, adquisición, rendimiento, conformidad y comportamiento humano) sobre las cinco etapas del ciclo de vida del dato (recolección, almacenamiento, distribución, reporte y disposición).

Es la única alternativa normativa certificable, ideal para auditorías de cumplimiento. No obstante, opera en un alto nivel de abstracción sin proporcionar metodologías operativas de implementación [6].

2.2.5 Síntesis comparativa

La tabla 1 compara los cuatro marcos según criterios relevantes para el contexto de este trabajo. Cada marco responde a una necesidad distinta; la elección depende del nivel de madurez, del tamaño de la organización y del objetivo perseguido.

2.3 Modelos de madurez en gestión de datos

Para establecer la línea base de capacidad organizacional en torno al manejo de datos, la literatura sugiere el uso de modelos de madurez. Estos marcos permiten evaluar el estado actual (*As-Is*) e identificar la brecha hacia el estado deseado (*To-Be*). Entre las alternativas de mayor adopción destacan:

CMMI Data Management Maturity Model: Evalúa 25 áreas de proceso agrupadas en 6 categorías sobre 5 niveles de madurez (Inicial, Gestionado, Definido, Medido y Optimizado) [7].

DCAM (Data Management Capability Assessment Model): Desarrollado por el EDM Council, enfocado prioritariamente en el sector financiero y estructurado en 8 componentes clave [8].

Stanford Data Governance Maturity Model: Diseñado específicamente para instituciones educativas, distingue entre dimensiones fundacionales (conciencia, estructura) y dimensiones operativas (calidad, metadatos) [9].

Mientras un modelo de madurez mide la capacidad procedimental y la sofisticación técnica de la organización, un instrumento de percepción mide el impacto directo en la satisfacción y usabilidad de los usuarios finales. Ambos enfoques son complementarios; sin embargo, en un entorno de recursos acotados, evaluar la percepción del equipo directivo entrega un indicador de valor de negocio inmediato para la validación de la hipótesis.

Tabla 1: Comparación de marcos de referencia de gobernanza de datos (fuente: elaboración propia)

Criterio	DAMA-DMBOK2	COBIT	DGI	ISO/IEC 38505-1
Nivel de abstracción	Operativo / Táctico	Estratégico corporativo	Táctico / Práctico	Estratégico normativo
Prescriptividad	Baja (descriptivo)	Media (guiado por riesgos)	Alta (enfocado en roles)	Muy baja (principios)
Cobertura del ciclo de vida	Exhaustiva (11 áreas)	Focalizada en TI / Riesgo	Enfocada en Reglas / Roles	Holística (estratégica)
Costo de adopción	Medio	Alto	Bajo	Alto (certificación)
Adecuación a recursos limitados	Moderada (requiere adaptar)	Baja	Alta	Baja

2.4 Pilares y dominios de la gobernanza de datos

Esta sección profundiza en los componentes fundamentales de un programa de gobernanza de datos, abordando tanto los pilares que lo desencadenan como las definiciones de roles, principios y etapas del ciclo de vida del dato necesarias para su efectiva implementación.

2.4.1 Pilares que desencadenan programas de gobernanza

Las tres áreas que requieren y generalmente desencadenan programas de gobernanza de datos son [10]:

Gestión de datos maestros (MDM, Master Data Management): Representa el proceso de gestionar datos a través de una organización. La gobernanza de datos da soporte al MDM de la siguiente forma:

- Asegurando que se definen, mantienen y cumplen estándares.
- Asegurando que el MDM está alineado con las necesidades del negocio y no sólo con la tecnología.
- Asegurando que la calidad, procesos de cambio y otras actividades en el MDM son aceptadas y adoptadas por la organización.

Inteligencia de Negocios (BI, Business Intelligence): La inteligencia de negocios corresponde al uso de información para conseguir las metas organizacionales. La gobernanza de datos mejora el BI a través de:

- Asegurar que las actividades de BI están alineadas con las actividades del negocio.
- Asegurar que la calidad de los datos da soporte al BI.
- Asegurar la consistencia en el tratamiento de los datos, es decir, que las métricas utilizadas por distintas áreas de la organización tengan el mismo significado.
- Apoya en el cumplimiento de una arquitectura de BI definida, evitando la redundancia de reportes y su crecimiento descontrolado.

Calidad de los datos: Para que los datos sean considerados un activo confiable, deben cumplir con una serie de características conocidas como dimensiones de la calidad. Un modelo de gobernanza robusto establece métricas y controles para asegurar el cumplimiento de estas dimensiones a lo largo de todo el ciclo de vida del dato. Las más relevantes para el contexto educativo son la exactitud, completitud, consistencia, oportunidad, unicidad y validez:

- Exactitud se refiere al grado en que los datos representan el objeto o evento que describen. En el contexto educativo, por ejemplo, la calificación registrada en el libro de clases digital en la evaluación de un estudiante debe ser exactamente la misma nota que el profesor puso en la evaluación física.
- Completitud mide si todos los datos requeridos esenciales están presentes. Por ejemplo, la ficha de matrícula del alumno debe contener obligatoriamente el nombre y teléfono de contacto de al menos un apoderado para tener en caso de emergencia.
- La consistencia asegura que los datos no presenten contradicciones y sean los mismos a través de diferentes sistemas o plataformas. En el caso de las instituciones educativas, por ejemplo, el registro de la asistencia y notas debe ser el mismo tanto en el libro de clases digital, como en los sistemas de registro del Ministerio de Educación.
- La oportunidad se refiere a que los datos estén disponibles en el momento que se necesitan para la toma de decisiones. Por ejemplo, el registro de asistencia debe realizarse de forma diaria para que los equipos directivos puedan identificar patrones de ausentismo crónico a tiempo para activar los protocolos correspondientes.
- Unicidad garantiza que no exista más de un registro para la misma entidad en una base de datos. En el ámbito educativo, cada alumno debe existir solo una vez en la base de datos de estudiantes, identificado por su RUT o Pasaporte.
- Finalmente, la validez asegura que los datos se ajusten a las reglas de negocio, formato o rangos definidos. Por ejemplo, el promedio final para un alumno en un ciclo lectivo definido, solo puede aceptar valores numéricos con un decimal dentro del rango [1,0-7,0].

Un programa de gobernanza requiere también una clara definición de responsabilidades operativas e institucionales:

Chief Data Officer (CDO): Ejecutivo de nivel directivo responsable del valor estratégico de los datos.

Propietario de datos (Data owner): Rol directivo o gerencial que posee la autoridad sobre un dominio de datos específico (ej. académico o formativo), define las reglas de negocio y autoriza los permisos de acceso.

Custodio de datos (Data steward): Rol operativo y táctico experto en el dominio de datos, responsable de vigilar la calidad, aplicar las reglas de negocio y gestionar los datos en el día a día. Se subdivide en custodios de negocio, técnicos y de dominio.

Custodio técnico (Data custodian): Responsable de la infraestructura física, bases de datos y seguridad técnica donde residen los datos.

La literatura contrapone dos esquemas de asignación: roles dedicados (creación de nuevos cargos exclusivos) frente a roles federados (asignación de responsabilidades sobre cargos preexistentes).

2.4.2 Principios de la gobernanza de datos

Se identifican cuatro principios de la gobernanza de datos [11]:

Organización: La gobernanza de datos tiene una dimensión organizacional, en la cual el modelo de gobernanza debe ser institucionalizado a través de una estructura formal que se adapte a la organización en particular.

Alineamiento: La gobernanza de datos debe asegurar que los datos cumplan las necesidades del negocio. Un programa de gobernanza debe demostrar valor de negocio, demostrando que los datos son útiles [12].

Cumplimiento: La gobernanza de datos debe tener una autoridad claramente definida que pueda crear y hacer cumplir las políticas de datos y procedimientos.

Entendimiento común: La gobernanza de datos sólo se podrá implementar de manera apropiada si se comprende de manera clara qué significa que los datos estén gestionados, y el por qué es relevante para la organización.

2.4.3 Ciclo de vida de los datos

El ciclo de vida de los datos comprende la generación, adquisición, conservación y procesamiento de datos, especificando las políticas de gestión para cada etapa, y la relación entre cada una de ellas [13].

Un modelo de gobernanza de datos aplica políticas y controles específicos en cada una de estas etapas para garantizar la integridad, calidad y seguridad de la información a lo largo de todo su recorrido. Por ejemplo, en la etapa de generación, la gobernanza define estándares de calidad para asegurar que los datos se capturen de forma precisa y completa. Durante su conservación y uso, establece controles de acceso y normativas de seguridad para proteger los datos contra usos no autorizados. Finalmente, en la etapa de procesamiento y eventual eliminación, dicta las políticas de retención y los procedimientos seguros de borrado para cumplir con las normativas de privacidad.

2.5 Arquitecturas de datos analíticos

La materialización de un modelo de gobernanza de datos requiere una plataforma tecnológica que soporte la consolidación, almacenamiento y explotación de la información. Las decisiones de arquitectura condicionan la auditabilidad, el control de acceso, la trazabilidad y la medición de la calidad del dato. A continuación, se revisan sintéticamente las alternativas de referencia en tres ejes fundamentales:

2.5.1 Paradigmas de repositorio analítico

Para materializar la infraestructura de datos analíticos, existen diversos paradigmas de repositorios. Dentro de ellos se puede mencionar:

Data Warehouse. Repositorio centralizado estructurado, integrado, no volátil y que varía en el tiempo para apoyar la toma de decisiones. Involucra gobernanza por diseño, pues exige modelar y validar los datos antes de su carga. Su rigidez frente a datos no estructurados se compensa con un alto desempeño en consultas analíticas y control de acceso granular [14].

Data Lake. El término corresponde a un repositorio que almacena datos en su formato nativo, sin transformación previa, difiriendo la interpretación al momento de la consulta (*schema-on-read*) [15]. Se apoya en almacenamiento de objetos de bajo costo y formatos abiertos, y admite datos estructurados, semiestructurados y no estructurados.

Lakehouse. Corresponde a una arquitectura que busca converger ambos paradigmas: incorpora una capa transaccional sobre almacenamiento de objetos en formatos columnares abiertos, aportando propiedades ACID, control de versiones, consulta de estados históricos y validación de esquema, sin renunciar a la flexibilidad ni al bajo costo del almacenamiento [16]. Su promesa es evitar la duplicación de datos entre un lago destinado a ciencia de datos y un almacén destinado a inteligencia de negocios.

Data Mesh. Corresponde a un cuarto enfoque que aplica la arquitectura de microservicios al dominio de la gestión de datos. Busca distribuir el consumo de datos a través de dominios específicos utilizando una capa de interoperabilidad universal [17].

2.5.2 Estrategias de integración y modelado

El segundo eje de decisión corresponde a cómo los datos transitan desde los sistemas de origen hacia el repositorio analítico. La literatura distingue dos dimensiones: el momento en que ocurre la transformación y la latencia con que se propagan los cambios.

ETL vs. ELT: El enfoque tradicional ETL (*Extract, Transform, Load*) transforma los datos antes de cargarlos, eliminando el dato original. En contraste, el enfoque ELT (*Extract, Load, Transform*) aprovecha el poder de cómputo de repositorios en la nube para cargar los datos crudos y transformarlos internamente en SQL. Esto preserva el dato original, habilita el linaje, la trazabilidad y la re-ejecución de reglas de negocio, atributos esenciales para la gobernanza.

Enfoques de modelado: Destacan el modelado dimensional de Kimball (tablas de hechos y dimensiones, ideal para la usabilidad en herramientas de BI y tratamiento de dimensiones lentamente cambiantes), el modelo normalizado de Inmon (3FN corporativo) y *Data Vault* (orientado a auditabilidad extrema)[18].

2.5.3 Criterios para la decisión arquitectónica

De la revisión anterior se desprende que no existe una combinación arquitectónica universalmente superior, sino configuraciones más o menos pertinentes según el contexto. Los criterios que la literatura y la práctica sugieren considerar son los siguientes:

Naturaleza y volumen de las fuentes: predominio de datos estructurados provenientes de sistemas transaccionales frente a datos heterogéneos o no estructurados.

Estabilidad de las preguntas analíticas: preguntas conocidas y recurrentes frente a exploración de interrogantes no anticipadas.

Latencia requerida: derivada del ciclo de decisión del proceso de negocio y no de la capacidad técnica disponible.

Competencias del equipo y sostenibilidad: capacidad de operar y mantener la solución una vez concluido el proyecto, sin dependencia de quien la implementó.

Costo total de propiedad: licencias, almacenamiento, cómputo y, especialmente, esfuerzo humano de operación.

Exigencias de trazabilidad y auditoría: derivadas del marco normativo aplicable y de la sensibilidad de los datos tratados.

Capacidad de evolución: posibilidad de incorporar nuevas fuentes y casos de uso sin rediseñar la solución.

La aplicación de estos criterios al contexto de la institución caso de estudio, y la combinación resultante, se desarrollan en la sección 3.1.

2.6 Metodologías de investigación aplicada en tecnologías de información

Un trabajo que produce una solución para un problema organizacional real enfrenta una decisión metodológica previa: determinar bajo qué marco de investigación se diseña esa solución y, sobre todo, bajo qué criterios se evalúa. En el ámbito de los Sistemas de Información (SI), se pueden mencionar los siguientes enfoques:

Investigación Basada en Diseño (*Design Science Research*, DSR): Se enfoca en la creación y evaluación de un artefacto innovador (modelos, métodos, instanciaciones) diseñado para resolver un problema práctico identificado en un entorno real. Se estructura el proceso en seis etapas iterativas: (1) identificación del problema y motivación, (2) definición de los objetivos de la solución, (3) diseño y desarrollo del artefacto, (4) demostración, (5) evaluación y (6) comunicación [19].

Investigación-Acción (*Action Research*): Enfocada en la transformación socio-organizacional mediante ciclos de planificación, acción, observación y reflexión donde el investigador actúa como agente de cambio. Su fortaleza radica en el cambio cultural, pero su capacidad de producir artefactos tecnológicos transferibles y replicables es menor [20].

Estudio de Caso: Estrategia de investigación empírica centrada en comprender un fenómeno contemporáneo en su contexto real. Es eminentemente descriptiva o explicativa, resultando más pertinente como método de evaluación del artefacto que como marco de diseño metodológico [21].

2.7 Instrumentos de medición de percepción en sistemas de información

Para evaluar la efectividad del artefacto propuesto sobre la gestión directiva, se requiere un marco conceptual de evaluación de impacto percibido. A continuación, se presentan los instrumentos que se utilizan en dicha evaluación:

Modelo de Éxito de Sistemas de Información: Constituye el marco de referencia consolidado para medir el impacto de las tecnologías de información en las organizaciones. El modelo establece que la Satisfacción del Usuario y el Impacto Neto están directamente determinados por la Calidad de la Información (exactitud, completitud, consistencia) y la Calidad del Sistema (disponibilidad, oportunidad, acceso). En esta tesina, el cuestionario se estructura operativamente sobre estas dos dimensiones de calidad para cuantificar el valor percibido por el equipo directivo [22].

Escala de Tipo Likert y Diseños Pre-Test / Post-Test: Para medir constructos de percepción se utilizan escalas ordinales Likert de 5 puntos (1 = Muy insatisfecho / En desacuerdo; 5 = Muy satisfecho / De acuerdo). El impacto de la intervención se evalúa mediante un diseño pre-experimental de grupo único pre-test/post-test, calculando la variación porcentual de las medias aritméticas como indicador de mejora práctica. Este diseño es idóneo para pilotos controlados en censos acotados, permitiendo medir la evolución directa sobre la línea base [23].

2.8 Gobernanza de datos en el contexto educativo

En esta sección se examina la aplicación de los principios de gobernanza de datos en el sector educativo, considerando tanto las exigencias normativas vigentes como los retos operativos intrínsecos a las instituciones escolares.

2.8.1 Marco normativo chileno aplicable

De acuerdo con lo indicado a la ley 21.430 sobre garantías de la niñez y adolescencia [24] en su artículo 33, se consagra el derecho a la vida privada y protección de datos personales de todo niño, niña y adolescente. Específicamente indica que: “los funcionarios públicos, las organizaciones de la sociedad civil que se relacionen con la niñez y su personal deberán guardar reserva y confidencialidad sobre los datos personales de niños, niñas y adolescentes”. Esto es especialmente relevante en educación, donde se almacena y trata la mayor cantidad de datos personales de menores de edad. De acuerdo a cifras del Mineduc, el año 2023 hubo 3.891.058 estudiantes matriculados, agrupados en nivel de Enseñanza Básica (56,1%), Enseñanza Media (20,0%) y Parvularia (9,4%) [25].

La promulgación de la Ley N° 21.719 [26] moderniza sustancialmente el régimen de la Ley N° 19.628 [27], alineando la legislación chilena con estándares internacionales como el GDPR europeo y creando la Agencia de Protección de Datos Personales. Esta ley consagra ocho principios rectores: licitud, finalidad, proporcionalidad, calidad, responsabilidad, seguridad, transparencia y confidencialidad. En el ámbito educativo, su aplicación es crítica debido al tratamiento masivo de datos sensibles de niños, niñas y adolescentes. La ley exige que el tratamiento se funde en bases de licitud estrictas (cumplimiento legal o consentimiento explícito) y consagra los derechos de acceso, rectificación, cancelación, oposición y portabilidad. Asimismo, impone el principio de responsabilidad proactiva (*accountability*), obligando a las instituciones a mantener un registro de actividades de tratamiento, realizar evaluaciones de impacto y notificar brechas de seguridad.

En el marco de la implementación de la Ley N° 21.180 de Transformación Digital del Estado, la norma dispone que los procedimientos administrativos de los órganos de la Administración del Estado se deben realizar en formato electrónico, debiendo implementarse de forma progresiva desde el año 2022 al 2027 [28]. Esto ha llevado a definir una propuesta de gestión y gobernanza de datos del Estado [29], lo cual, a su vez, ha hecho que el Ministerio de Educación comience a trabajar en una política de gobernanza de datos [30]. Esto puede impactar en la forma en que las instituciones educativas en Chile gestionan sus propios datos.

Cada exigencia legal se traduce directamente en un componente del modelo propuesto: el principio de seguridad se operacionaliza mediante la matriz de accesos por roles (IAM); el principio de proporcionalidad guía el protocolo de minimización de datos; y el principio de responsabilidad proactiva se formaliza a través de los roles de custodia asignados.

2.8.2 Desafíos para el sector educativo

La implementación de un modelo de gobernanza de datos en el sector educativo enfrenta un conjunto de desafíos particulares que lo diferencian significativamente del entorno corporativo. Estos obstáculos que deben ser comprendidos para diseñar soluciones que sean pertinentes y sostenibles en dicho contexto se presentan a continuación:

Fragmentación tecnológica y silos de información: El ecosistema digital de las instituciones educativas se caracteriza por una marcada heterogeneidad, con una multiplicidad de plataformas y sistemas de información que, por lo general, no operan de manera integrada. Esta falta de interoperabilidad genera silos de información, donde los datos residen de forma aislada, duplicada e inconsistente, lo que impide obtener una visión unificada y holística del estudiante y de la gestión escolar.

Recursos limitados: A diferencia del sector privado, las instituciones educativas operan con restricciones presupuestarias crónicas y una escasez de capital humano especializado en tecnologías y gestión de datos. Esta limitación de recursos dificulta la inversión en plataformas tecnológicas avanzadas y en la contratación de personal con las competencias necesarias para liderar proyectos complejos de gobernanza de datos.

Cultura de datos incipiente: Más allá de las barreras técnicas, existe un desafío cultural relacionado con la alfabetización de datos (*data literacy*). A menudo, existe una brecha en las competencias de los equipos directivos y docentes para no solo acceder a la información, sino también para analizarla críticamente y transformarla en decisiones pedagógicas y estratégicas efectivas. En el día a día, se evidencia también que en los colegios acostumbra a trabajar con información presentada de forma estática, en formatos de documentos con análisis y gráficos estáticos. El desarrollo de una cultura de datos, que valore y utilice la evidencia para la mejora continua, es un proceso gradual y complejo.

Naturaleza compleja y sensible del dato: Finalmente, el dato educativo posee una complejidad inherente superior a la de otros sectores. No se limita a registros académicos, sino que abarca un espectro multidimensional que incluye información socioemocional, de salud, de convivencia y familiar. Este carácter sensible, protegido por normativas estrictas sobre datos personales de menores, eleva significativamente las exigencias de seguridad, privacidad y gestión ética, haciendo indispensable un marco de gobernanza que garantice su tratamiento responsable.

2.9 Trabajos relacionados y brecha identificada

Existen investigaciones que han abordado la importancia de contar con un plan de gobernanza de datos en instituciones educativas. En [31] se investigó la importancia de la calidad de los datos en instituciones universitarias, aplicando un estudio en la Universidad Cooperativa de Colombia. También en el ámbito universitario, en [32] se identifican los problemas relacionados con no contar con un plan de gobernanza de datos, y se propone una plataforma de gobernanza de datos para universidades. La implementación de gobernanza de datos como un habilitador de la modernización de la gobernanza educacional se muestra a través de un caso de estudio en [33].

Por otro lado, entidades gubernamentales de diversos países han abordado la gobernanza de datos en instituciones escolares con el fin de dar lineamientos a las escuelas sobre el uso de los datos que generan, el resguardo con la privacidad y las responsabilidades involucradas. Es el caso de la Superintendencia de Instrucción Pública de Washington, que desarrolló políticas y procedimientos para un sistema de gobernanza de datos en instituciones K12, escuelas que imparten educación desde el jardín (kindergarten) hasta el 12° grado. [34], enfocado principalmente en la calidad de los datos y la infraestructura requerida. El departamento de educación de Estados Unidos por su parte, a través de su Centro de Asistencia Técnica para la Privacidad

(PTAC), publicó una guía sobre la gestión de sistemas complejos de datos a través de la gobernanza [35]. Junto con ello, también han disponibilizado un *checklist* para verificar la implementación de un programa de gobernanza de datos en las escuelas K-12 [36].

Otras instituciones, como The London School of Economics and Political Science (LSE), en conjunto con la ONG 5Rights Foundation, han desarrollado a través de la Digital Futures Commission (DFC), un plan para el manejo de los datos educativos en el Reino Unido [37], donde se destaca la importancia de respetar los derechos de los menores de edad y el involucramiento de todos los actores que participan en la generación y uso de estos datos, desde las escuelas y sus docentes, padres y tutores, los proveedores de tecnologías para la educación, y las instituciones estatales.

Estudios en Reino Unido han analizado la relación que existe entre los proveedores de servicios tecnológicos de educación y las escuelas, y el uso que estos proveedores hacen de los datos recolectados [38], identificando los riesgos que existen al no haber una regulación clara en el uso de los datos. Por otro lado, se destaca el aporte del uso de datos en la educación, a través del análisis predictivo, prescriptivo, y la visualización [39], y en cómo *big data* se ha vuelto relevante en el ámbito educativo [40].

De la revisión de la literatura se concluye que la investigación académica previa en gobernanza de datos educativos ([31], [32], [33]) se ha concentrado casi exclusivamente en instituciones de educación superior - entidades que disponen de presupuestos, áreas de TI especializadas y estructuras de gobierno incomparables con una escuela. Por otro lado, las iniciativas orientadas al sector escolar K-12 ([34], [35], [36], [37]) corresponden mayoritariamente a guías gubernamentales o marcos normativos teóricos que carecen de una instanciación técnica evaluada empíricamente.

Se identifica, por lo tanto, la siguiente brecha de investigación: no existe en la literatura un modelo de gobernanza de datos diseñado, instanciado técnicamente mediante arquitecturas de bajo costo y evaluado cuantitativamente en su impacto operacional dentro de instituciones escolares K-12 con recursos limitados en el contexto normativo chileno. El modelo de Gobernanza de Datos para la Mejora Escolar (GDME) desarrollado en esta tesina aborda explícitamente dicha brecha.

3 Diseño de la solución

Este capítulo desarrolla el diseño del modelo de Gobernanza de Datos para la Mejora Escolar (GDME). Comienza estableciendo las decisiones metodológicas, de marcos y tecnológicas adoptadas, seleccionadas de entre las alternativas revisadas en el capítulo 2 y justificadas según las restricciones y contexto de la institución caso de estudio. El capítulo se organiza de acuerdo con las tres primeras etapas de la metodología DSR: identificación del problema y motivación, definición de los objetivos de la solución, y diseño y desarrollo del artefacto.

3.1 Selección de metodología, marcos y herramientas

La elección de los componentes del modelo responde a las restricciones del contexto escolar: equipo de TI reducido, presupuesto acotado, necesidad de resultados en un ciclo corto y capacidad de mantención posterior sin dependencia del investigador. La tabla 2 sintetiza la selección fundamentada de las alternativas teóricas presentadas en el capítulo 2.

3.2 Etapa 1 - Identificación del problema

Esta sección recopila el diagnóstico cuantitativo y la definición del alcance.

Tabla 2: Selección de metodología, marcos y herramientas (fuente: elaboración propia)

Dimensión	Alternativas revisadas	Opción seleccionada	Justificación
Metodología de la investigación	DSR vs. Investigación-Acción vs. Estudio de Caso	DSR	Permite diseñar, instanciar y evaluar empíricamente un artefacto tecnológico (GDME) orientado a resolver un problema práctico.
Marco de gobernanza	DAMA-DMBOK2 vs. COBIT vs. DGI vs. ISO/IEC 38505-1	DAMA-DMBOK2	Ofrece el mapa de disciplinas más completo (MDM, BI, Calidad) para construir la gestión de datos como activo
Estructura de roles	Roles dedicados vs. Roles federados	Roles federados (Existentes)	Asigna responsabilidades de <i>Data Owner</i> y <i>Data Steward</i> a cargos preexistentes evitando sobrecostos
Repositorio analítico	<i>Data Warehouse</i> vs. <i>Data Lake</i> vs. <i>Lakehouse</i> vs. <i>Data Mesh</i>	<i>Data Warehouse</i> (Cloud)	Asegura esquema estructurado, gobernanza por diseño y control de acceso granular sin sobrecarga técnica.
Integración y modelado	ETL vs. ELT / Kimball vs. Inmon	ELT / Modelado Dimensional (Kimball)	ELT preserva el dato crudo para trazabilidad; Kimball optimiza usabilidad en BI y manejo histórico de estudiantes.
Herramientas de plataforma	Diversos proveedores y herramientas	Google BigQuery + Looker Studio + Python	Solución <i>serverless</i> de bajo costo, sin administración de servidores, altamente integrable y costo-efectiva

3.2.1 Delimitación del alcance

Para asegurar la viabilidad y el foco de la investigación, el alcance del diagnóstico se acotó a dos dominios críticos para la gestión educativa: el seguimiento académico y el seguimiento formativo. Esta delimitación responde directamente al objetivo de la tesina, que buscó medir el impacto de la gobernanza en la satisfacción de los equipos directivo y gerencial respecto a la disponibilidad y calidad de los datos en estas áreas para la toma de decisiones.

Seguimiento académico: comprende los datos y procesos vinculados al rendimiento de los estudiantes, tales como calificaciones, porcentajes de logro, tasas de aprobación y la generación de informes académicos consolidados.

Seguimiento formativo: incluye los datos y procesos relacionados con el desarrollo integral del estudiante, como los registros de asistencia, atrasos y anotaciones (positivas y negativas).

3.2.2 Definición de la población objetivo

La población objetivo para la recolección de datos se definió como el equipo directivo y gerencial de la institución educativa. Este grupo fue seleccionado por ser el principal consumidor de la información consolidada para la toma de decisiones estratégicas y tácticas. La hipótesis de la investigación se centra en su nivel de satisfacción, por lo que su percepción es la variable fundamental a medir.

3.2.3 Diseño del instrumento de medición

Dado el ámbito del estudio se optó por un enfoque cuantitativo. Para ello, se diseña un instrumento único de recolección de datos: un cuestionario estructurado. El cuestionario (ver Anexo I) fue desarrollado para medir la percepción de la población objetivo a través de dos dimensiones clave, que se presentan en la tabla 3.

El instrumento utiliza una escala de tipo Likert de 1 a 5, donde 1 representa “Muy en desacuerdo” o “Muy insatisfecho” y 5 representa “Muy de acuerdo” o “Muy satisfecho”. Este formato permite cuantificar las percepciones y calcular un puntaje de satisfacción basal, que sirve como punto de comparación contra el cual se mide el impacto de la intervención propuesta.

Tabla 3: Dimensiones del cuestionario a aplicar (fuente: elaboración propia)

Dimensión	Descripción
Disponibilidad y acceso a la información	Evalúa la facilidad y oportunidad con que los directivos pueden acceder a los datos académicos y formativos que requieren
Calidad y confianza en los datos	Mide el grado de confianza de los usuarios en la exactitud, completitud y consistencia en la información presentada

3.2.4 Protocolo de aplicación y recolección

Se estableció un protocolo formal para la aplicación del instrumento, con el fin de asegurar la máxima tasa de respuesta y la calidad de los datos recolectados, considerando comunicación oficial, distribución del instrumento, y plazo y seguimiento.

La recolección de datos se llevó a cabo siguiendo estrictamente el protocolo definido. El cuestionario estructurado fue distribuido en formato digital (Google Forms) al universo total del equipo directivo y gerencial de la fundación educativa. Se estableció un periodo de recolección de siete días hábiles, durante el cual se monitoreó la tasa de respuesta. Para asegurar la máxima participación, se envió un recordatorio formal a la población objetivo 48 horas antes de la finalización del plazo. Se garantizó en todo momento el anonimato y la confidencialidad de los participantes a través de la configuración de la plataforma de formularios en línea, que no recopiló información personal identificable. Al cierre del periodo se obtuvo el conjunto de datos brutos correspondiente a las respuestas completas del cuestionario.

3.2.5 Procesamiento y análisis

Finalizada la recolección, los datos brutos fueron exportados a una hoja de cálculo para su limpieza, procesamiento y tabulación. Esta etapa preparatoria fue fundamental para asegurar la integridad y la correcta estructuración de la información previa al análisis. El proceso consistió en la revisión de la completitud de cada cuestionario para asegurar que no hubiera datos omitidos. Posteriormente, las respuestas de la escala Likert fueron codificadas y asignadas a sus valores numéricos correspondientes (donde 1 = Muy insatisfecho/Muy en desacuerdo y 5 = Muy satisfecho/Muy de acuerdo). Los datos fueron organizados y estructurados en una matriz, agrupando las respuestas según las dos dimensiones definidas en el instrumento.

Para el análisis de los datos tabulados se empleó estadística descriptiva. Este enfoque permitió resumir y describir las características principales del conjunto de datos, transformando las respuestas individuales en indicadores agregados que reflejan la percepción colectiva de la población objetivo. Para cada una de las preguntas del cuestionario y para cada una de las dos dimensiones se calculó la media aritmética, obteniendo el valor promedio de satisfacción en cada ítem y dimensión. Adicionalmente, se calculó la desviación estándar para cada pregunta con el fin de observar el grado de dispersión o consenso en las percepciones de los directivos. Un valor bajo en la desviación estándar indica un alto nivel de acuerdo, mientras que un valor alto señala una diversidad de opiniones.

3.2.6 Resultados del diagnóstico y línea base

El resultado final de esta fase fue la consolidación de los análisis en un diagnóstico cuantitativo del estado inicial, o "*As-Is*", de la gobernanza de datos en la institución. A partir de los promedios calculados, se estableció el indicador clave de la investigación: la línea base de satisfacción del equipo directivo.

Se calculó un puntaje promedio general de satisfacción, así como los puntajes promedio específicos para cada una de las dos dimensiones evaluadas. Este puntaje basal representa el punto de partida numérico y objetivo antes de la implementación de cualquier mejora. Sirve como el punto de referencia fundamental contra el cual se contrastan los resultados de una medición posterior, permitiendo así la validación o refutación de la hipótesis central de esta tesina, que postula un incremento de al menos un 20% sobre esta línea base como resultado de la intervención propuesta.

Tras aplicar el instrumento a una población de 9 directivos, el análisis establece una línea base general de 2,36, reflejando una insatisfacción de base. Los directivos logran acceder a los datos, pero a costa de navegar entre múltiples sistemas desconectados, lo que se refleja en evaluaciones críticas y puntos de dolor muy específicos como el análisis histórico (DISP4: 1,89).

Al analizar el promedio por percepción de disponibilidad, se puede apreciar que la percepción de que es fácil acceder a datos históricos (de años anteriores) de los estudiantes cuando es necesario, es aquella con la menor evaluación (DISP4).

En cuanto a la dimensión de calidad, los indicadores con menor promedio son aquellos que hacen referencia a que los datos de calificaciones registrados son exactos y libres de errores (CAL1), y que los datos de un mismo estudiante son consistentes a través de diferentes sistemas o reportes (CAL3).

En síntesis, los resultados del diagnóstico inicial cuantitativo evidencian que el núcleo de la insatisfacción directiva radica en la imposibilidad de consolidar análisis históricos (promedio 1,89), en los errores en los registros de calificaciones, y en la falta de consistencia de datos a través de diversos sistemas y reportes.

3.3 Etapa 2 - Definición de los objetivos de la solución

A partir de los hallazgos del diagnóstico inicial, se tradujeron las necesidades identificadas en requerimientos funcionales y operacionales que el Modelo GDME debía satisfacer. La tabla 4 detalla esta correspondencia.

Tabla 4: Traducción de hallazgos del diagnóstico en objetivos y requerimientos de la solución (fuente: elaboración propia)

Hallazgo	Requerimiento de la solución	Componente del modelo que lo resuelve
Baja evaluación en acceso a datos históricos (DISP4: 1,89)	Centralizar datos históricos en un repositorio único con retención de cambios.	<i>Data Warehouse</i> (BigQuery) con modelo dimensional Kimball.
Errores en registros de calificaciones e inconsistencia de datos (CAL1 y CAL3: 2,22)	Definir reglas de negocio automatizadas y monitorear inconsistencias.	Proceso de calidad de datos y tablero de calidad en Looker Studio

3.4 Etapa 3 - Diseño y desarrollo del artefacto

A partir de los hallazgos obtenidos en la fase de diagnóstico y en concordancia con la metodología DSR, en esta sección se presenta el diseño del artefacto central de esta investigación: el modelo GDME.

El GDME está concebido como un marco de trabajo adaptado a la realidad y a los recursos de las instituciones educativas, cuyo propósito es solucionar las problemáticas de dispersión, inconsistencia y subutilización de la información detectadas. El modelo busca establecer una base sólida para que los datos se transformen en un activo estratégico, mejorando la calidad y disponibilidad de la información para el seguimiento académico y formativo.

3.4.1 Principios rectores

Principio de dato único y confiable: La institución debe operar sobre una única fuente para sus datos críticos, especialmente los datos maestros de los estudiantes. Esto elimina la ambigüedad y asegura la consistencia de la información a través de todos los sistemas y reportes.

Principio de seguridad y ética por diseño: El manejo de datos, en particular los datos sensibles de los estudiantes, debe cumplir con la normativa vigente (ley 21.430, ley 21.719). La seguridad, la privacidad y el uso ético de la información son componentes integrales del modelo y no un añadido posterior.

Principio de utilidad pedagógica: El fin último de la gobernanza de datos en el contexto escolar es la mejora de los procesos de enseñanza-aprendizaje. Por tanto, todo esfuerzo en la gestión de datos debe estar orientado a facilitar la toma de decisiones informadas por parte de directivos, coordinadores y docentes.

3.4.2 Estructura organizativa y roles

El GDME se estructura en tres componentes interdependientes que deben ser desarrollados en conjunto para asegurar el éxito de la gobernanza de datos: la estructura organizativa y roles, los procesos clave de gobernanza y la plataforma tecnológica habilitadora.

Se propone una estructura organizativa que no implica la creación de nuevos cargos, sino la asignación de responsabilidades de datos a roles ya existentes dentro de la institución.

Comité de Gobernanza de Datos (nivel estratégico)

Misión: constituye la máxima autoridad en la toma de decisiones estratégicas sobre los activos de datos de la institución. Es responsable de aprobar políticas, priorizar iniciativas y resolver conflictos relacionados con la información.

Composición: para asegurar la alineación entre la estrategia de la fundación y la realidad operativa, el comité integra roles transversales y roles locales del establecimiento piloto:

- Gerente de Control de Gestión y Planificación de la institución (nivel central/sostenedor): lidera la visión estratégica y estandarización de datos para toda la red de colegios.
- Encargado(a) de Infraestructura Tecnológica de la institución (nivel central/sostenedor): garantiza la seguridad y arquitectura tecnológica común.
- Director del establecimiento (nivel local): representa la realidad operativa y valida la factibilidad de las políticas en el colegio.
- Subdirector Académico del establecimiento (nivel local): actúa como propietario del dominio de datos académicos.
- Encargado de Convivencia Escolar del establecimiento (nivel local): actúa como propietario del dominio de datos formativos.

Responsabilidades clave:

- Aprobar la definición de los datos maestros.
- Validar las reglas de calidad de datos.
- Revisar periódicamente los indicadores de calidad y acceso a la información.

Roles de custodia de datos (nivel táctico)

Propietario de Datos (Data Owner): rol de alta dirección (Subdirector Académico, Encargado de Convivencia) con la máxima responsabilidad sobre un dominio de datos específico (académico, formativo). Define las reglas de negocio y autoriza el acceso a la información bajo su custodia.

Custodio de Datos (Data Steward): rol táctico, experto en un área de datos particular, responsable de la gestión y calidad de los datos en el día a día. Se asignó esta responsabilidad a roles como Coordinadores de Ciclo (para datos de rendimiento de sus cursos), Secretaria Académica (para datos de matrícula) y Orientador(a) (para datos de seguimiento psicoeducativo).

3.4.3 Procesos clave de gobernanza

Proceso de gestión de datos maestros (MDM)

Este proceso se centra en la creación de una “Ficha Única de Estudiante”. Los pasos para su implementación son:

1. Definición: El comité de gobernanza define los campos que constituyen el dato maestro del estudiante.
2. Identificación de la fuente fidedigna: Se determina qué sistema o registro es la fuente oficial para cada campo definido.
3. Centralización: Se implementa un repositorio central que consolida esta Ficha Única.
4. Protocolo de mantenimiento: Se establece un procedimiento formal, a cargo del custodio designado, para la creación, actualización y eliminación de registros, asegurando que todos los demás sistemas consuman la información desde esta fuente central.

Proceso de gestión de la calidad de los datos

Se establece un ciclo de mejora continua para la calidad de la información:

1. Definición de las reglas: Los custodios de datos proponen reglas de negocio para validar la calidad de los datos de su dominio (Por ejemplo: “Todo estudiante debe tener su RUT con formato y dígito verificador válido”).
2. Medición y monitoreo: Se implementa un tablero de calidad de datos que mide automáticamente el cumplimiento de estas reglas.
3. Remediación: El comité revisa periódicamente este tablero y asigna a los custodios la tarea de corregir los errores detectados en sus fuentes.

Proceso de gestión de acceso y seguridad

Para garantizar el cumplimiento normativo y la seguridad, se establece el siguiente proceso:

1. Definición de matriz de acceso: Se crea una matriz que define qué rol en la organización puede consultar, crear, modificar o eliminar qué tipo de datos.
2. Protocolo de solicitudes: Se formaliza un canal para que los usuarios soliciten acceso a la información, el cual debe ser aprobado por el propietario del dato correspondiente.

3.4.4 Plataforma tecnológica habilitadora

La implementación del modelo se apoya en una arquitectura tecnológica centralizada. Se implementa un repositorio central de datos (*Data Warehouse*) que consolida la información clave de los distintos sistemas transaccionales (SIGE, libro de clases digital). Este repositorio alberga la “Ficha Única del Estudiante” y los principales indicadores académicos y formativos.

Sobre el repositorio central, se implementa una herramienta de BI que permite la creación de los tableros de calidad de datos y seguimiento para el equipo directivo, facilitando una visualización intuitiva y actualizada de la información.

3.4.5 Especificación técnica de la arquitectura

Esta sección detalla la arquitectura y los componentes técnicos definidos para la plataforma tecnológica que da soporte al GDME. La solución se diseñó bajo los principios de escalabilidad, bajo costo y mínima sobrecarga administrativa, considerando las restricciones y el contexto de una institución educativa.

Se adoptó una arquitectura de datos en la nube que sigue un flujo lógico de tres etapas: ingesta y transformación (ELT), almacenamiento y modelado (*data warehouse*), y visualización. Esta arquitectura centraliza la información de los sistemas de origen dispersos en un único repositorio confiable para su posterior análisis, y se presenta en la figura 1.

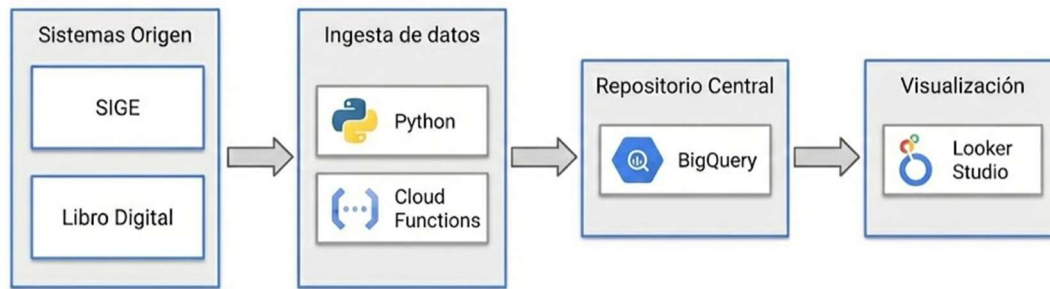


Figura 1: Diagrama de arquitectura de alto nivel

Componente 1: Sistemas de origen. Representa el conjunto de plataformas y registros transaccionales desde los cuales se extrae la información primaria para el modelo de gobernanza. Estos sistemas es donde se registra la mayor cantidad de información de los estudiantes y sus trayectorias dentro de los colegios.

Componente 2: Ingesta y transformación (ELT). Este componente es responsable de extraer los datos desde sus sistemas de origen, cargarlos en el repositorio central y transformarlos para que sean aptos para el análisis. La solución se conecta a la plataforma institucional del libro de clases digital, que contiene la información académica y formativa de los estudiantes. Para la extracción y carga de datos se utilizaron rutinas en Python empleando librerías estándar como Pandas para la manipulación de datos y SQLAlchemy para la conexión a bases de datos. Estas rutinas se orquestan y ejecutan de forma automatizada y sin servidor a través de Google Cloud Functions. Este enfoque no requiere la gestión de infraestructura y su costo se basa en el tiempo de ejecución.

Componente 3: Repositorio central de datos. En este repositorio residen los datos unificados y constituye la fuente confiable de información. Se utilizó Google BigQuery como plataforma de almacenamiento y procesamiento, por las siguientes características:

- Modelo sin servidor: Google gestiona toda la infraestructura subyacente, eliminando la necesidad de administración por parte del equipo de TI de la institución.
- Costo basado en uso: el modelo de precios se basa en la cantidad de datos almacenados y procesados en las consultas, con una capa gratuita que permite iniciar un proyecto de esta escala.
- Escalabilidad y rendimiento: BigQuery maneja grandes volúmenes de datos y ejecuta consultas complejas en segundos.

Dentro de BigQuery, los datos se organizan siguiendo un modelo dimensional en estrella, conforme a la decisión de modelado justificada en 3.1.

Componente 4: Visualización y explotación de datos. Este componente corresponde a la capa final de la solución, que entrega la información procesada a los usuarios finales. Como herramienta de BI se utilizó Google Looker Studio, por tres motivos: es una herramienta gratuita; posee un conector nativo y de alto rendimiento con Google BigQuery, lo que simplifica la conexión a los datos; y su interfaz es intuitiva, permitiendo que usuarios no técnicos exploren los datos y construyan reportes básicos.

Los entregables iniciales desarrollados en Looker Studio corresponden a dos tableros principales, diseñados para responder a las necesidades detectadas en el diagnóstico:

- Tablero de calidad de datos: permite visualizar el cumplimiento de las reglas de calidad definidas en el modelo de gobernanza, permitiendo al comité identificar y gestionar activamente las inconsistencias.
- Tablero de seguimiento académico y formativo: ofrece una vista consolidada del rendimiento de los estudiantes, tasas de asistencia y tendencias de comportamiento, permitiendo al equipo directivo realizar análisis comparativos y tomar decisiones estratégicas informadas.

En la tabla 5 se resumen los componentes del GDME:

Tabla 5: Detalle de los componentes del modelo

Componente	Característica principal	Descripción específica (Adaptada al contexto escolar)
Principios	Adaptabilidad y ahorro	Diseñado para operar con recursos (humanos y financieros) limitados, priorizando herramientas de bajo costo y asignando responsabilidades a roles existentes.
	Utilidad pedagógica	El fin último no es el control del dato en sí mismo, sino facilitar la toma de decisiones que impacten positivamente en el seguimiento académico y formativo.
	Seguridad por diseño	Cumplimiento nativo de la normativa para la protección de datos sensibles de menores.
Estructura organizativa	Comité de gobernanza	(Nivel estratégico) Liderado por el Gerente de Control de Gestión y Planificación de la institución. Aprueba políticas y prioriza iniciativas.
	Propietarios (<i>Data Owners</i>)	(Nivel táctico) Roles directivos con máxima autoridad sobre un dominio de datos.
	Custodios (<i>Data Stewards</i>)	(Nivel operativo) Roles expertos responsables de la calidad del dato en el día a día.
Procesos clave	Gestión de datos maestros (MDM)	Creación de una "Ficha Única del Estudiante" para eliminar duplicidades y establecer una "fuente única de la verdad".
	Gestión de calidad de datos	Ciclo de mejora continua basado en la medición (tablero de calidad de datos) y remediación de datos que no cumplen las reglas de negocio.
	Gestión de acceso	Protocolo formal basado en una matriz de roles y permisos para asegurar que solo el personal autorizado acceda a datos sensibles.
Plataforma tecnológica	Repositorio central (<i>Data Warehouse</i>)	Google BigQuery: Plataforma <i>serverless</i> (sin gestión de servidores) y de bajo costo para centralizar la información.
	Ingesta de datos (ELT)	Google Cloud Functions + Python: <i>Scripts</i> automatizados y económicos para extraer datos de los sistemas de origen (silos).
	Visualización (BI)	Google Looker Studio: Herramienta gratuita para la creación de tableros de seguimiento (para directivos) y de calidad de datos (para el Comité).

4 Implementación y evaluación

Este capítulo describe las etapas de demostración y evaluación de la metodología DSR. Se presenta la instanciación del modelo GDME en el entorno real de la institución educativa y, a continuación, la medición de su impacto mediante la aplicación del instrumento utilizado en el diagnóstico. El capítulo concluye verificando el nivel de cumplimiento de la hipótesis y de cada uno de los objetivos específicos planteados, y analizando las condiciones de replicabilidad del modelo en instituciones de características similares.

4.1 Etapa 4 - Demostración: puesta en marcha del piloto

Para validar la utilidad y viabilidad del artefacto diseñado, esta sección detalla la ejecución de la etapa de demostración, describiendo los pasos seguidos para poner en marcha el modelo GDME en un escenario real dentro de la institución.

4.1.1 Planificación y preparación

Esta subsección describe los elementos estratégicos y las definiciones operativas que permitieron estructurar la implementación piloto, garantizando su alineación con los objetivos de la investigación.

Definición del alcance. Se definió un alcance acotado pero representativo para la prueba piloto. La implementación se centró exclusivamente en los dominios de seguimiento académico y formativo, utilizando como fuentes de datos iniciales el libro de clases digital y la información de SIGE. El piloto tuvo una duración de un mes para permitir la generación y el uso de los datos.

Constitución del comité de gobernanza. Como primer paso de la implementación del modelo, se constituyó formalmente el Comité de Gobernanza de Datos, compuesto por los roles directivos definidos en 3.4.2. En una sesión inicial se presentaron los objetivos del piloto, se validaron las responsabilidades de cada miembro y se establecieron las reglas de calidad de datos iniciales a monitorear.

Consideraciones de seguridad. Dado que el piloto involucra el tratamiento de datos personales de menores de edad (categorizados como datos sensibles según la normativa vigente), se adoptaron medidas concretas para garantizar su resguardo durante toda la ejecución, en coherencia con el principio de seguridad y ética por diseño del modelo GDME.

En primer lugar, se contó con la aprobación de la dirección de la institución para el tratamiento de los datos con fines del piloto, acordando que la información no saldría del entorno *cloud* institucional y que su uso se limitaría exclusivamente al seguimiento académico y formativo. En segundo lugar, se aplicó el principio de minimización de datos: se cargaron únicamente los campos estrictamente necesarios para los indicadores definidos, excluyendo información no requerida para el análisis.

En cuanto al control de acceso, se implementó la matriz de roles y permisos definida en el modelo mediante los mecanismos nativos de Google Cloud (IAM en BigQuery y permisos a nivel de reporte en Looker Studio), de modo que cada miembro del comité y cada usuario accediera únicamente a los datos correspondientes a su dominio y nivel de responsabilidad. El acceso al repositorio central quedó restringido a los roles autorizados, y la visualización se entregó a los directivos a través de tableros con permisos gestionados individualmente. Finalmente, respecto al instrumento de evaluación, se garantizó el anonimato de los participantes y se recabó su consentimiento informado (ver Anexo I), sin recopilar datos identificables de los directivos encuestados.

4.1.2 Desarrollo e implementación técnica

Esta subfase consistió en la construcción de la plataforma tecnológica habilitadora, siguiendo la especificación descrita en 3.4.5.

Desarrollo de procesos ELT. Se desarrollaron y probaron las rutinas en Python para la extracción, carga y transformación de los datos desde las fuentes de origen (libro de clases digital y SIGE) hacia el repositorio central.

Construcción del repositorio central. Se configuró el entorno en Google BigQuery, se crearon las tablas y se programó la ejecución periódica de los procesos ELT para mantener los datos actualizados.

Diseño de tableros de BI. Se construyeron en Google Looker Studio los dos artefactos de visualización principales: el tablero de calidad de datos y el tablero de seguimiento académico y formativo. Su diseño se realizó de forma iterativa, con sesiones de validación junto al Comité de Gobernanza para asegurar que la visualización respondiera a las necesidades de gestión.

4.1.3 Puesta en marcha y capacitación

Una vez desarrollada la plataforma se procedió con su despliegue y la capacitación de los usuarios finales.

Sesiones de capacitación. Se realizaron sesiones de formación dirigidas a todo el equipo directivo y gerencial. La capacitación se centró en la interpretación de los nuevos tableros, la navegación en la información y su uso para fundamentar las discusiones en reuniones de gestión y consejos de profesores.

Operación asistida. Durante el primer mes de operación se mantuvo un monitoreo constante del funcionamiento de la plataforma y se proporcionó soporte directo a los usuarios para resolver dudas y fomentar la adopción de las nuevas herramientas. El Comité de Gobernanza celebró su primera reunión operativa para revisar el tablero de calidad de datos y asignar las primeras tareas de remediación de información.

4.2 Etapa 5 - Evaluación

Tras la puesta en marcha del piloto, el siguiente paso metodológico consiste en evaluar los resultados obtenidos, utilizando el mismo instrumento diagnóstico para cuantificar el impacto del modelo GDME en las dimensiones de disponibilidad y calidad de los datos.

4.2.1 Contexto de la evaluación

La evaluación se llevó a cabo tras la finalización del período piloto de un mes, aplicando el mismo instrumento de medición utilizado en el diagnóstico (cuestionario estructurado en escala Likert de 1 a 5). El instrumento fue respondido por la totalidad de la población objetivo definida: el equipo directivo y gerencial ($N=9$), asegurando así la comparabilidad directa con la línea base establecida en 3.2.6.

4.2.2 Comparativo general: pre-test vs. post-test

El análisis global de los datos revela un cambio sustancial en la percepción de los usuarios tras la implementación de la plataforma centralizada y los procesos de gobernanza. Como se observa en la tabla 6, el promedio general de satisfacción aumentó considerablemente, pasando de una percepción baja a una satisfactoria.

Dado el tamaño acotado de la población ($N=9$), el análisis se mantuvo en el plano descriptivo; no se aplicaron pruebas de significancia inferencial, cuya validez es limitada con muestras de este tamaño. El aumento de la desviación estándar en la medición posterior indica que, si bien la satisfacción general se incrementó de forma marcada, el impacto no fue percibido de manera homogénea por todos los directivos: la dispersión de las respuestas creció, lo que sugiere que la adopción y valoración de la nueva plataforma varió entre los miembros del equipo.

Los resultados evidencian que el impacto fue prácticamente uniforme entre ambas dimensiones (+66,52% y +66,80%), lo que valida la efectividad de la arquitectura tecnológica implementada, que resolvió el problema de la fragmentación de datos al centralizar la información en una vista única.

4.2.3 Análisis detallado por dimensión

Impacto en la disponibilidad y acceso. Esta dimensión fue la más crítica durante el diagnóstico. Los directivos reportaban perder tiempo navegando entre distintas plataformas (libro de clases, SIGE, planillas de cálculo) para obtener una visión completa del estudiante. Tras el piloto, los ítems mejor evaluados fueron:

Acceso oportuno: Pasó de 2,44 a 4,11. Los usuarios destacaron que la actualización automática (proceso diario) del tablero en Looker Studio eliminó la necesidad de solicitar reportes manuales a los coordinadores.

Acceso a históricos: Pasó de 1,89 a 3,11. Si bien sigue siendo el ítem con menor puntaje absoluto de la dimensión, la centralización en el repositorio permite consultar el comportamiento de asistencia de años anteriores de forma directa, algo que anteriormente requería revisar libros físicos o archivos históricos. Su nivel aún moderado es coherente con el carácter progresivo de la remediación histórica.

Impacto en la calidad y confianza. Si bien la tecnología facilita el acceso, la gobernanza asegura la calidad. El incremento en esta dimensión (de 2,44 a 4,07) refleja el trabajo de los nuevos roles definidos, específicamente los custodios de datos.

El tablero de calidad de datos permitió visibilizar errores que antes pasaban desapercibidos, como RUTs duplicados o inconsistencias en los registros de matrícula. Aunque la percepción de confianza aumentó, no alcanzó el máximo de la escala debido a que el proceso evidenció que la limpieza de datos históricos es una tarea progresiva que requiere tiempo humano de remediación, un hallazgo consistente con la literatura sobre la madurez de la gobernanza.

Tabla 6: Comparativa General de Satisfacción

Dimensión Evaluada	Línea Base (Pre-Piloto)	Desviación estándar (Pre-Piloto)	Resultado (Post-Piloto)	Desviación estándar (Post-Piloto)	Variación Absoluta	Variación Porcentual
Disponibilidad y Acceso	2,27	0,61	3,78	1,09	+1,51	+66,52%
Calidad y Confianza	2,44	0,63	4,07	0,76	+1,63	+66,80%
Promedio General	2,36		3,93		+1,57	+66,53%

Nota: La variación porcentual se calcula mediante la fórmula estándar de tasa de crecimiento.

4.3 Verificación de la hipótesis

La hipótesis planteada en esta investigación postulaba lo siguiente:

"La implementación piloto de un modelo de Gobernanza de Datos adaptado generará un incremento de al menos un 20% en la satisfacción del equipo directivo reportada respecto a la disponibilidad y calidad de los datos para el seguimiento académico estudiantil".

Considerando los resultados cuantitativos obtenidos:

1. **Línea Base (Pre):** 2,36
2. **Resultado Final (Post):** 3,93
3. **Variación Real:** 66,53%

y que $66,53\% > 20\%$, se declara que la **hipótesis ha sido validada** sobre la satisfacción general. El modelo GDME demostró ser capaz de superar ampliamente las expectativas mínimas de mejora, transformando la experiencia de gestión de datos de los directivos en un ciclo corto de tiempo.

4.4 Verificación del cumplimiento de objetivos

Además de la validación de la hipótesis, corresponde evaluar en qué medida la investigación satisfizo los objetivos específicos formulados en la sección 1.3. La tabla 7 presenta esta verificación, indicando para cada objetivo la evidencia producida, la sección del informe donde consta y el nivel de cumplimiento alcanzado. El objetivo general -diseñar, implementar y evaluar un modelo de gobernanza de datos adaptado al contexto escolar chileno- se considera alcanzado: el artefacto fue construido, instanciado en un entorno real y evaluado con un instrumento aplicado a la totalidad de la población objetivo, verificándose una mejora sustancial en la disponibilidad y calidad percibidas de la información para el seguimiento académico y formativo. Cabe precisar dos alcances: la mejora en la calidad del dato se validó en el plano de la percepción directiva y no mediante métricas de conformidad extraídas del propio tablero de calidad; y los lineamientos de replicabilidad se derivan analíticamente del diseño y de la experiencia del piloto, sin haber sido puestos a prueba en una segunda institución, verificación que se retoma como línea futura.

Tabla 7: Verificación del cumplimiento de los objetivos específicos

Objetivo específico	Evidencia producida	Sección	Cumplimiento
Caracterizar el estado inicial de la gobernanza de datos	Línea base de 2,36 sobre 5,0, con desagregación por dimensión e ítem, sobre el censo directivo (N = 9)	3.2.6	Cumplido
Comparar marcos, arquitecturas y metodologías, y seleccionar los componentes aplicables	Revisión comparativa de cuatro marcos de gobernanza, tres modelos de madurez, cuatro paradigmas de repositorio y tres enfoques de modelado, con matriz de decisión fundamentada según las restricciones del contexto	2.2 a 2.7; 3.1	Cumplido
Diseñar un modelo que articule estructura, procesos y plataforma	Modelo GDME con tres principios rectores, estructura de roles federados sobre cargos existentes, tres procesos de gobernanza y arquitectura tecnológica especificada	3.4	Cumplido
Implementar el modelo mediante un piloto controlado	Comité constituido, rutinas ELT operativas, repositorio central poblado, dos tableros publicados y equipo directivo capacitado, sin creación de cargos ni licencias adicionales	4.1	Cumplido
Evaluar el impacto y derivar lineamientos de replicabilidad	Medición posterior con el mismo instrumento, incremento de 66,53% sobre la línea base, y distinción entre núcleo transferible y componentes contingentes con guía de adopción	4.2, 4.3, 4.6	Cumplido

4.5 Hallazgos cualitativos y observaciones

Más allá de las cifras, la ejecución del piloto arrojó hallazgos cualitativos relevantes durante las sesiones de retroalimentación con el Comité de Gobernanza:

1. **Cambio en la conversación directiva:** Antes del piloto, gran parte del tiempo en las reuniones se perdía discutiendo la validez del dato ("¿Este promedio está actualizado?", "¿Esta es la asistencia real?"). Con la "Ficha Única del Estudiante" operando como fuente de verdad, la discusión se trasladó hacia la toma de decisiones pedagógicas basadas en la evidencia.

2. **Empoderamiento de los Data Stewards:** Asignar la responsabilidad de la calidad de datos a roles existentes (como la Secretaría Académica) inicialmente generó resistencia por la carga laboral percibida. Sin embargo, al utilizar el tablero de calidad de datos para identificar errores masivos, estos roles reportaron una disminución en el retrabajo operativo a largo plazo.
3. **Viabilidad de la solución Low-Code/Serverless:** La arquitectura basada en Google Cloud Functions y BigQuery funcionó de manera transparente, sin requerir mantenimiento de servidores por parte del encargado de TI, confirmando que es una solución sostenible para instituciones con recursos técnicos limitados.

Estos hallazgos no provienen de un instrumento estructurado, sino del registro sistemático de la interacción con la institución durante el piloto. Su fuente son los registros de acuerdos de las sesiones del Comité de Gobernanza de Datos, las notas de campo tomadas durante las sesiones de capacitación y el período de operación asistida, y las consultas recibidas por el canal de soporte a usuarios. Dado que el investigador formó parte de la institución y participó activamente en la implementación, estas observaciones corresponden a observación participante y deben interpretarse como evidencia complementaria de carácter exploratorio, útil para contextualizar los resultados cuantitativos y orientar investigaciones posteriores, pero no como hallazgos con el mismo estatus de validez que los obtenidos mediante el instrumento aplicado al equipo directivo.

4.6 Replicabilidad del modelo

Uno de los propósitos de esta investigación es que el modelo GDME pueda ser adoptado por instituciones escolares distintas de la estudiada. Ello exige distinguir qué componentes del modelo son de aplicación general y cuáles responden a decisiones contingentes al caso, establecer las condiciones mínimas que una institución debe reunir para adoptarlo, y ofrecer una secuencia de implementación que un tercero pueda seguir sin necesidad de reconstruir el proceso completo.

4.6.1 Núcleo transferible vs. componentes contingentes

El GDME se compone de elementos con distinto grado de dependencia respecto del contexto en que fue instanciado. La tabla 8 clasifica cada componente según su transferibilidad y señala la adaptación que requeriría en otra institución.

De esta clasificación se desprende que el núcleo del modelo -principios, estructura de roles, procesos y patrón arquitectónico- es transferible sin modificaciones sustantivas, y que la adaptación se concentra en la capa de conexión con los sistemas de origen y en la parametrización de las reglas de negocio. Las decisiones de producto, en particular, no forman parte del modelo: fueron seleccionadas por su costo y su disponibilidad en la institución, y podrían reemplazarse sin alterar la lógica de la solución.

4.6.2 Condiciones mínimas de adopción

La aplicación del modelo en otra institución supone reunir cuatro condiciones. La primera es el patrocinio explícito de la dirección, sin el cual la asignación de responsabilidades de datos a cargos existentes carece de la autoridad necesaria para sostenerse frente a la carga operativa cotidiana. La segunda es disponer de al menos un rol técnico con capacidad para desarrollar y mantener las rutinas de carga; el modelo reduce esa exigencia al mínimo mediante una arquitectura sin servidores, pero no la elimina. La tercera es el acceso efectivo a los datos de los sistemas de origen, ya sea mediante conexión directa a la base de datos o mediante exportaciones periódicas acordadas con el proveedor, condición que en la práctica suele ser la principal barrera de entrada. La cuarta es un nivel mínimo de alfabetización de datos en el equipo directivo, suficiente para interpretar los tableros y sostener discusiones basadas en evidencia; su ausencia no impide la implementación técnica, pero limita la apropiación del artefacto y, con ello, el impacto que este puede alcanzar.

4.6.3 Guía de adopción por etapas

La secuencia de implementación seguida en el piloto, depurada de las particularidades del caso, se resume en seis etapas:

1. **Diagnosticar.** Aplicar el instrumento del Anexo I al equipo directivo para establecer la línea base e identificar los puntos críticos que orientarán las prioridades de diseño.
2. **Constituir el comité.** Designar formalmente los roles del comité y los propietarios y custodios de cada dominio de datos, y acordar en la sesión inicial las primeras reglas de calidad a monitorear.
3. **Definir el dato maestro.** Determinar los campos que integran la ficha única del estudiante y establecer la fuente fidedigna de cada uno, resolviendo los conflictos entre sistemas antes de cualquier desarrollo.
4. **Implementar el repositorio.** Construir las rutinas de extracción y carga, modelar los datos en el repositorio central y programar su actualización periódica según el ciclo de decisión de la institución.
5. **Publicar los tableros.** Desarrollar el tablero de seguimiento académico y formativo y el tablero de calidad de datos, validando iterativamente con el comité, y capacitar a los usuarios finales en su interpretación.
6. **Iterar la calidad.** Establecer una cadencia de revisión del tablero de calidad, asignar las tareas de remediación a los custodios correspondientes y medir periódicamente el avance.

Las etapas 1 a 3 son organizacionales y anteceden a cualquier desarrollo técnico; omitirlas o posponerlas conduce a construir una plataforma sobre definiciones de datos no acordadas, que es precisamente el origen de las inconsistencias que el modelo busca resolver.

Tabla 8: Transferibilidad de los componentes del modelo GDME

Componente	Transferibilidad	Adaptación requerida
Principios rectores	Alta	Ninguna, son independientes del tamaño y tipo de establecimiento
Estructura del comité y roles federados	Alta	Mapear los roles a la nomenclatura de cargos de cada institución; en establecimientos sin nivel central, el comité se constituye íntegramente con roles locales
Procesos de datos maestros, calidad y acceso	Alta	Redefinir el conjunto de campos de la ficha única y las reglas de negocio según los sistemas y protocolos propios
Patrón arquitectónico	Alta	Ninguna a nivel de patrón; es independiente del proveedor
Productos específicos (BigQuery, Google Cloud Functions, Looker Studio)	Media	Sustituibles por equivalentes de otros proveedores o por alternativas de código abierto, conservando el patrón
Rutinas de extracción de datos desde las fuentes	Baja	Dependen del sistema de origen; SIGE es común a todo el sistema escolar chileno, pero el libro de clases digital varía entre proveedores
Instrumento de medición y línea base	Alta el instrumento, nula la línea base	El cuestionario es reutilizable; los valores de referencia deben levantarse en cada institución

Conclusiones

Esta investigación abordó la dispersión de datos y la subutilización de la información en instituciones educativas escolares. Mediante la metodología DSR se diseñó, implementó y evaluó el modelo de Gobernanza de Datos para la Mejora Escolar (GDME) en una fundación educativa chilena.

El diagnóstico inicial confirmó cuantitativamente la brecha existente, estableciendo una línea base de satisfacción de 2,36 sobre 5,0, especialmente crítica en el acceso a información histórica. Tras la intervención, la satisfacción promedio se incrementó a 3,93, equivalente a un aumento de 66,53%, superando ampliamente el umbral de 20% postulado en la hipótesis, la cual quedó validada. La mejora fue prácticamente equivalente en ambas dimensiones evaluadas, lo que indica que la centralización tecnológica constituye la palanca más efectiva en el corto plazo, si bien la calidad del dato demostró ser un desafío que requiere, además de tecnología, un cambio cultural sostenido y tiempo humano de remediación.

La principal contribución de esta tesina es la formalización del GDME como un artefacto factible y de bajo costo para el contexto educativo escolar chileno. El modelo demuestra que es posible establecer una fuente única de la verdad mediante una arquitectura *serverless* sin incurrir en inversiones prohibitivas en infraestructura ni licencias. Adicionalmente, el estudio aporta una hoja de ruta organizacional al evidenciar que la asignación de roles de gobernanza a cargos existentes es una estrategia viable para instalar una cultura de datos, siempre que se les dote de herramientas que simplifiquen su labor operativa.

En síntesis, este trabajo demuestra que la gobernanza de datos es una herramienta estratégica y accesible que, mediante un diseño ajustado al contexto, permite a las escuelas chilenas dejar de buscar datos para comenzar a tomar decisiones que mejoren la educación.

En cuanto a las limitaciones identificadas durante la implementación piloto del modelo GDME, las principales son:

Generalización. La validación se realizó mediante un estudio de caso único en una fundación particular. Si bien los problemas son comunes al sector, la aplicación en establecimientos con menor madurez digital podría presentar barreras de entrada mayores.

Tiempo de maduración. El piloto tuvo una duración de un mes. Si bien fue suficiente para validar la disponibilidad, la mejora profunda de la calidad del dato es un proceso iterativo de largo plazo que no alcanzó a ser medido en su totalidad.

Foco de la métrica. La investigación se centró en la satisfacción percibida del usuario directivo. No se midió el impacto directo en el rendimiento académico de los estudiantes (por ejemplo, resultados SIMCE o tasas de aprobación), ya que dicho efecto es una variable rezagada que requiere estudios longitudinales de varios años.

Finalmente, sobre líneas futuras, se propone lo siguiente:

Replicación en el sector público. Aplicar el modelo GDME en establecimientos municipales para evaluar su adaptabilidad en entornos con normativas administrativas más rígidas.

Medición de impacto pedagógico. Realizar un estudio longitudinal (2 a 3 años) que correlacione la madurez de la gobernanza de datos con indicadores duros de mejora escolar, como la retención escolar y los resultados de aprendizaje.

Integración de analítica predictiva. Implementar modelos de aprendizaje automático sobre los datos históricos para predecir la deserción escolar temprana, transitando de una gestión descriptiva a una predictiva.

Adecuación al nuevo marco normativo. Extender el modelo con los instrumentos de cumplimiento que exige la Ley N° 21.719, convirtiendo al GDME en un modelo no solo de gestión sino también de cumplimiento normativo verificable.

Referencias

- [1] S. Earley, DAMA-DMBOK : Data Management Body of knowledge, 2nd ed. Basket Ridge, New Jersey: Technics Publications, 2017.
- [2] V. Khatri and C. V. Brown, "Designing Data Governance," Communications of the ACM, vol. 53, no. 1, p. 148, Jan. 2010, doi: <https://doi.org/bbx5dx>.
- [3] S. Ahmadi, M. M. Tavana, S. Shokouhyar, and M. Dortaj, "A new fuzzy approach for managing data governance implementation relevant activities," TQM Journal, 2021, doi: 10.1108/TQM-01-2021-0015.
- [4] G. Mangalaraj, A. Singh, y A. Taneja, «IT Governance Frameworks and COBIT -a Literature Review Completed Research Paper». <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1262&context=amcis2014>
- [5] D. Acev et al., "Systematic analysis of data governance frameworks and their relevance to data trusts," Management Review Quarterly, July 2025, doi: 10.1007/s11301-025-00545-1.
- [6] ISO/IEC 38505-1 Information technology - Governance of data - Part 1: Application of ISO/IEC 38500 to the governance of data, 2026, <https://www.iso.org/standard/87195.html>
- [7] S. N. Asih, R. Nabila, I. H. Ismed, W. R. Fitriani, A. N. Hidayanto and S. B. Yudhoatmojo, "Evaluation of Data Operations Management Maturity Level using CMMI in a State-Owned Enterprise," 2019 5th International Conference on Computing Engineering and Design (ICCED), Singapore, 2019, pp. 1-6, doi: 10.1109/ICCED46541.2019.9161117.
- [8] EDM Council. DCAM (Data Management Capability Assessment Model), Version 2.2, 2020.
- [9] D. H. Kurniawan, Y. Ruldeviyani, M. R. Adrian, S. Handayani, M. R. Pohan and R. K. T, "Data Governance Maturity Assessment: A Case Study in IT Bureau of Audit Board," 2019 International Conference on Information Management and Technology (ICIMTech), Jakarta/Bali, Indonesia, 2019, pp. 629-634, doi: 10.1109/ICIMTech.2019.8843742.
- [10] J. Ladley, DATA GOVERNANCE : how to design, deploy and sustain an effective data governance program. S.L.: Elsevier Academic Press, 2019.
- [11] P. Brous, M. Janssen, and R. Vilminko-Heikkinen, "Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles," Lecture Notes in Computer Science, pp. 115–125, 2016, doi: <https://doi.org/ghn6t2>.
- [12] S. S. Dawes, "Stewardship and usefulness: Policy principles for information-based transparency," Government Information Quarterly, vol. 27, no. 4, pp. 377–383, Oct. 2010, doi: <https://doi.org/cq3xjk>.
- [13] A. V. Levitin and T. C. Redman, "A model of the data (life) cycles with application to quality," Information and Software Technology, vol. 35, no. 4, pp. 217–223, Apr. 1993, doi: <https://doi.org/cgtwrp>.
- [14] T. Li, Y. Chui and P. S. Ren, "Analysis on Data Warehouse Technology and Its Development Situation," 2008 International Symposium on Information Science and Engineering, Shanghai, China, 2008, pp. 486-488, doi: 10.1109/ISISE.2008.30.
- [15] E. Zagan and M. Danubianu, "Cloud DATA LAKE: The new trend of data storage," 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 2021, pp. 1-4, doi: 10.1109/HORA52670.2021.9461293.
- [16] D. Orešćanin and T. Hlupić, "Data Lakehouse - a Novel Step in Analytics Architecture," 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2021, pp. 1242-1246, doi: 10.23919/MIPRO52101.2021.9597091.
- [17] G. K. Aydın, S. G. Köse, C. Ulus and M. F. Akay, "Development Of A Self-Serve Data Mesh Infrastructure," 2025 9th International Symposium on Innovative Approaches in Smart Technologies (ISAS), Gaziantep, Türkiye, 2025, pp. 1-4, doi: 10.1109/ISAS66241.2025.11101756.
- [18] L. Yessad and A. Labiod, "Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault," 2016 International Conference on System Reliability and Science (ICSRS), Paris, France, 2016, pp. 95-99, doi: 10.1109/ICSRS.2016.7815845.
- [19] P. Johannesson and E. Perjons, Introduction To Design Science. Springer International Pu, 2016.
- [20] F. Lau, "Toward a framework for action research in information systems studies," Information Technology & People, vol. 12, no. 2, pp. 148–176, June 1999, doi: 10.1108/09593849910267206.
- [21] T. Hollweck, "Robert K. Yin. (2014). Case Study Research Design and Methods (5Th Ed.). Thousand Oaks, CA: Sage. 282 Pages."

- [22] W. H. Delone y E. R. McLean, “The DeLone and McLean Model of Information Systems Success: A Ten-Year Update”, *Journal of Management Information Systems*, vol. 19, n.o 4, pp. 9-30, abr. 2003.
- [23] W.-E. Sosa-Gonzalez, J. Bautista-Ortega, y A. Santillán Fernández, “Escala De Likert: Una Alternativa Para Elaborar E Interpretar Un Instrumento De Percepción Social”.
- [24] B. del C. Nacional, “Biblioteca del Congreso Nacional | Ley Chile,” www.bcn.cl/leychile, Mar. 15, 2022. <https://www.bcn.cl/leychile/navegar?idNorma=1173643>
- [25] “Datos Abiertos Mineduc,” <https://datosabiertos.mineduc.cl>.
- [26] Congreso, “Biblioteca del Congreso Nacional | Ley Chile,” www.bcn.cl/leychile, 2025. <https://www.bcn.cl/leychile/navegar?idNorma=1209272>
- [27] B. del C. Nacional, “Biblioteca del Congreso Nacional | Ley Chile,” www.bcn.cl/leychile, Aug. 28, 1999. <https://www.bcn.cl/leychile/navegar?idNorma=141599&idVersion=2023-05-09&idParte=8642700>.
- [28] “Ley de Transformación Digital,” *Ley de Transformación Digital*. <https://digital.gob.cl/transformacion-digital/ley-de-transformacion-digital/>
- [29] “Estrategia de Gestión de Datos”, 2023. https://participacion.digital.gob.cl/uploads/7383448d-a601-4a6e-9e4e-b9d1dcd92b50/phase_file/file/4dbf82b3-bb1c-4943-b62b-b68cafe66b25/Estrategia_de_Gesti%C3%B3n_de_Datos_2023.pdf
- [30] “Gobernanza y gestión de los datos,” Mineduc.cl, 2023. <https://www.innovacion.mineduc.cl/iniciativas/transformaci%C3%B3n-digital/gobernanza-y-gesti%C3%B3n-de-los-datos>.
- [31] M. Osorio, Mayda Patricia González, and C. Andrés, “La gobernabilidad de datos como apoyo en la gestión de datos de instituciones de educación superior,” *ESPACIOS*, vol. 38, no. 51, Nov. 2017.
- [32] H. Jiang and W. Yan, “The Application of Data Governance in Universities,” vol. 26, pp. 54–59, May 2023, doi: <https://doi.org/10.1109/icaibd57115.2023.10206371>.
- [33] Z. Wang and R. Huang, “Research on the implementation path and practice of data driven university governance modernization-Taking Shandong Youth College of Political Science as an example,” 2021 11th International Conference on Information Technology in Medicine and Education (ITME), vol. 40, pp. 500–504, Nov. 2021, doi: <https://doi.org/10.1109/itme53901.2021.00106>.
- [34] E. Rang, “Data Governance System for K-12 Data”, 2020: <https://ospi.k12.wa.us/sites/default/files/2023-08/2020-06-data-governance-manual-final.pdf>
- [35] “Data Governance and Stewardship About PTAC.”, 2015: https://studentprivacy.ed.gov/sites/default/files/resource_document/file/Data_Governance_and_Stewardship_0.pdf
- [36] “Data Governance Checklist About PTAC.”, 2015: https://studentprivacy.ed.gov/sites/default/files/resource_document/file/Data%20Governance%20Checklist_0.pdf
- [37] “A Blueprint for Education Data” 2023: <https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/A-Blueprint-for-Education-Data-FINAL-Online.pdf>
- [38] E. Day, K. Pothong, A. Atabey, and S. Livingstone, “Who controls children’s education data? A socio-legal analysis of the UK governance regimes for schools and EdTech,” *Learning, Media and Technology*, pp. 1–15, Dec. 2022, doi: <https://doi.org/10.1080/17439884.2022.2152838>.
- [39] B. Williamson, “Digital education governance: data visualization, predictive analytics, and ‘real-time’ policy instruments,” *Journal of Education Policy*, vol. 31, no. 2, pp. 123–141, Apr. 2015, doi: <https://doi.org/10.1080/02680939.2015.1035758>.
- [40] X. Yang, Zhu Xiang-rong, and D. Chen, “Discourses regarding education governance in the digital age at K-12 level: Possibilities, risks, and strategies,” *Teaching and Teacher Education*, vol. 132, pp. 104261–104261, Oct. 2023, doi: <https://doi.org/10.1016/j.tate.2023.104261>.

Anexos

Anexo I: Formulario de aplicación de encuesta

1. Introducción y Consentimiento Informado:
 - a. Breve explicación del propósito del estudio
 - b. Indicación de duración estimada (ej. 10-15 minutos).
 - c. Datos de contacto del investigador.
 - d. Casilla de verificación para aceptar participar.
2. Datos Demográficos / Contexto:
 - a. Cargo/Rol principal en la institución/colegio: (Lista desplegable: Director, Subdirector, Gerente TI, Gerente Control Gestión, Gerente General)
 - b. Años de experiencia en su cargo actual
3. Percepción sobre DISPONIBILIDAD de Datos (para seguimiento académico/formativo):
 - a. *Instrucción:* "A continuación, evalúe su nivel de acuerdo con las siguientes afirmaciones sobre qué tan disponibles están los datos que usted necesita para realizar el seguimiento académico y formativo de los estudiantes. Use la escala: 1 = Totalmente en desacuerdo, 2 = En desacuerdo, 3 = Ni de acuerdo ni en desacuerdo, 4 = De acuerdo, 5 = Totalmente de acuerdo."
 - b. *Preguntas (Idénticas Pre y Post):*
 - i. DISP1: Encuentro fácilmente la información sobre calificaciones y evaluaciones de los estudiantes que necesito. (Escala 1-5)
 - ii. DISP2: Puedo acceder de manera oportuna a los registros de asistencia de los estudiantes. (Escala 1-5)
 - iii. DISP3: Los datos necesarios para identificar tempranamente a estudiantes en riesgo académico o con necesidades de apoyo están accesibles a tiempo. (Escala 1-5)
 - iv. DISP4: Es fácil acceder a datos históricos (de años anteriores) de los estudiantes cuando es necesario. (Escala 1-5)
 - v. DISP5: En general, estoy satisfecho/a con la disponibilidad (facilidad y oportunidad de acceso) de los datos para el seguimiento estudiantil. (Escala 1-5)
(Pregunta clave resumen de disponibilidad)
4. Percepción sobre CALIDAD de Datos (para seguimiento académico/formativo):
 - a. *Instrucción:* "Ahora, piense en la calidad de los datos que utiliza para el seguimiento. Indique su nivel de acuerdo con las siguientes afirmaciones. Use la misma escala: 1 = Totalmente en desacuerdo, ..., 5 = Totalmente de acuerdo."
 - b. *Preguntas (Idénticas Pre y Post):*
 - i. CAL1: Considero que los datos de calificaciones registrados son exactos y libres de errores. (Escala 1-5)
 - ii. CAL2: La información de asistencia registrada suele ser completa y precisa. (Escala 1-5)
 - iii. CAL3: Los datos de un mismo estudiante son consistentes a través de diferentes sistemas o reportes (ej. datos personales, curso asignado). (Escala 1-5)
 - iv. CAL4: Los datos están suficientemente actualizados para ser útiles en el seguimiento. (Escala 1-5)
 - v. CAL5: Confío en la calidad general de los datos que utilizo para tomar decisiones sobre los estudiantes. (Escala 1-5)
 - vi. CAL6: En general, estoy satisfecho/a con la calidad (exactitud, completitud, consistencia, actualidad) de los datos para el seguimiento estudiantil. (Escala 1-5)
(Pregunta clave resumen de calidad)
5. Cierre:
 - a. Agradecimiento por el tiempo y la colaboración.

Anexo II: Detalle de resultados de encuestas

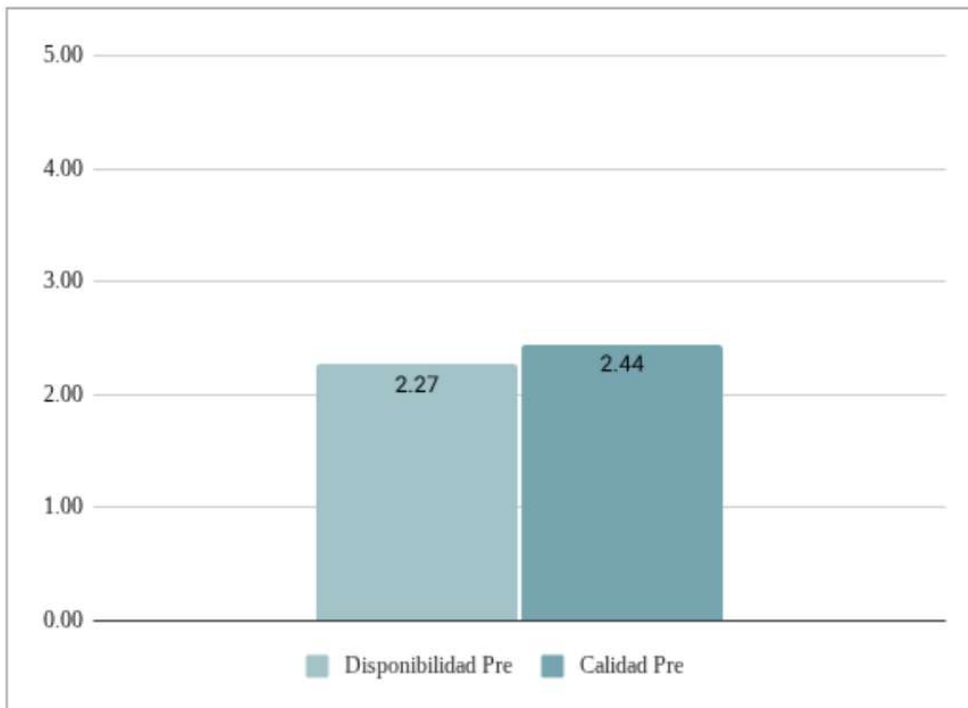


Figura 2: Promedio de evaluación diagnóstica por dimensión

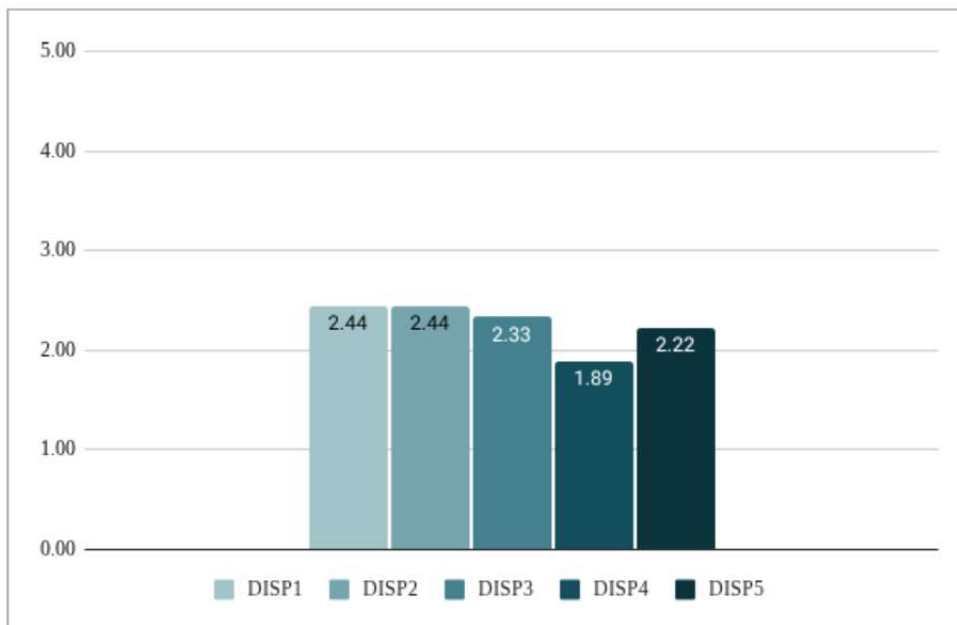


Figura 3: Promedio de evaluación diagnóstica de percepción de Disponibilidad

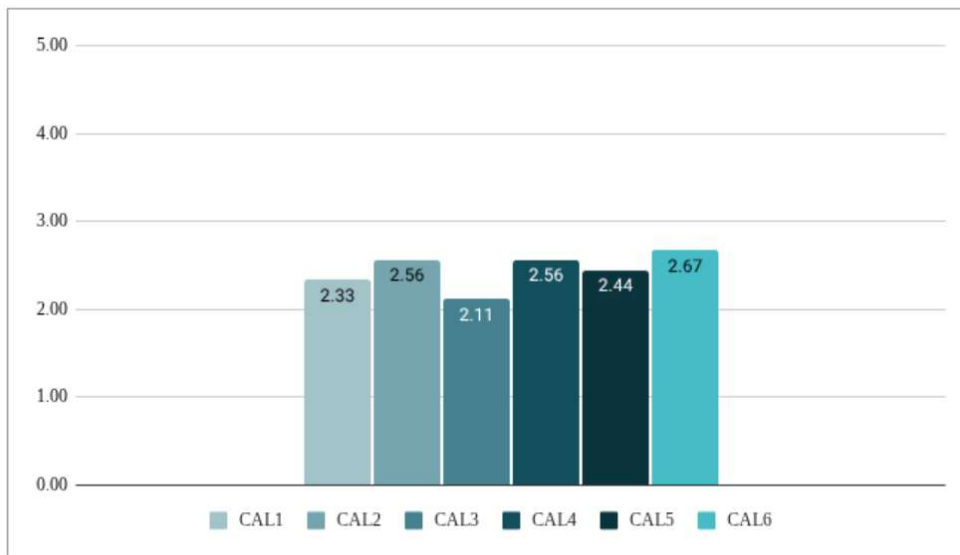


Figura 4: Promedio de evaluación diagnóstica de percepción de Calidad