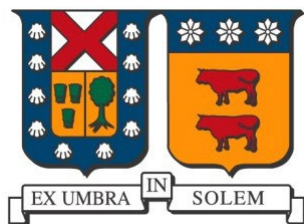


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
VALPARAÍSO - CHILE



**“ARQUITECTURA DE REFERENCIA PARA LA GESTIÓN
SEGURA DE DATOS MASIVOS EN SERVICIOS PÚBLICOS:
UN ENFOQUE INTEGRADO DE DATA LAKE Y
PROTECCIÓN DE DATOS PERSONALES”**

MARCO ANTONIO MENDOZA PARADA

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

Profesor Guía: DANIEL HORMAZABAL



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: Arquitectura de Referencia para la Gestión Segura de Datos Masivos en Servicios Públicos: Un Enfoque Integrado de Data Lake y Protección de Datos Personales

Nombre del candidato(a): Marco Antonio Mendoza Parada

Carrera / Grado: Magíster en Tecnologías de la Información

Campus: San Joaquín Departamento: Departamento de Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Daniel Mauricio Hormazábal Ocampo, 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: 26-10-2026

Firma:

Estudiante o Candidato(a):

Fecha: 26-10-2026

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.



Arquitectura de Referencia para la Gestión Segura de Datos Masivos en Servicios Públicos: Un Enfoque Integrado de Data Lake y Protección de Datos Personales

Marco Antonio Mendoza Parada
ma.mendoza.p@gmail.com

Resumen: Las instituciones públicas chilenas enfrentan el desafío de modernizar la gestión de datos masivos bajo la Ley de Transformación Digital del Estado (N°21.180), la Ley Marco de Ciberseguridad (N°21.663) y la Ley de Protección de Datos Personales (N°21.719). Este trabajo propone un Framework de Arquitectura de Referencia (FAR) que integra Data Lake (Medallion), seguridad Zero Trust y Privacy by Design, con un mapeo explícito a la normativa. La validación se realiza bajo el paradigma de Design Science Research (DSR) sobre una prueba de concepto real en AWS, aplicada al proceso de Bonificación de Salud (BRSALUD) del Instituto de Previsión Social (IPS). El FAR se formaliza como un blueprint replicable, adoptable por cualquier servicio público mediante parametrización, y se demuestra su independencia tecnológica mediante un comparativo de despliegue equivalente en AWS y Azure. El impacto esperado es un marco reutilizable que mejora la experiencia ciudadana, reduce los riesgos de seguridad y habilita el cumplimiento normativo, transferible a otros organismos del Estado y a países de la región con regulaciones similares.

Palabras Clave: Arquitectura de Referencia, Data Lake, Zero Trust, Privacy by Design, Protección de Datos Personales.

1. Introducción y planteamiento del problema

1.1 Contexto y motivación

Las instituciones del sector público chileno atraviesan una transformación digital sin precedentes, impulsada por la Ley N°21.180 de Transformación Digital del Estado y por la creciente demanda ciudadana de servicios digitales eficientes, seguros y accesibles. En este escenario, los organismos administran volúmenes masivos y heterogéneos de datos provenientes de múltiples fuentes —registros civiles, sistemas de salud, beneficios sociales, plataformas de trámites en línea y datos operacionales internos—, cuyo tratamiento adecuado es determinante para la calidad del servicio y para la confianza ciudadana. La infraestructura tecnológica predominante, sin embargo, no fue concebida para esta escala ni para las exigencias normativas actuales, lo que abre una brecha entre las capacidades instaladas y los requisitos de eficiencia, seguridad y cumplimiento.

Esta brecha tiene consecuencias concretas: retrasa la modernización del Estado, eleva los costos operacionales, dificulta la interoperabilidad entre organismos y expone datos sensibles de millones de ciudadanos. La promulgación de la Ley Marco de Ciberseguridad (N°21.663) y de la Ley de Protección de Datos Personales (N°21.719) eleva aún más el estándar exigible, estableciendo obligaciones de seguridad, trazabilidad y protección de datos para las cuales los sistemas legados no están preparados. Resolver esta brecha de forma estandarizada y replicable constituye, por

tanto, una necesidad estratégica del Estado, y motiva el desarrollo de una arquitectura de referencia que articule, en un único marco, tecnología, seguridad y cumplimiento normativo.

1.2 Definición técnica del problema (en el dominio de TI)

El ecosistema tecnológico del sector público chileno se caracteriza por arquitecturas monolíticas y sistemas legados (legacy) que operan en silos de información, con incapacidad estructural para gestionar las “Tres V” del Big Data (Volumen, Velocidad y Variedad). El problema se descompone en tres dimensiones críticas:

- 1. Fragmentación y rigidez de datos:** los datos residen en bases relacionales aisladas y con esquemas rígidos, lo que impide la ingesta y el procesamiento eficiente de datos no estructurados (documentos, imágenes, logs) que constituyen gran parte de la interacción ciudadana actual. La ausencia de una “Fuente Única de Verdad” deriva en duplicidad de información e inconsistencias entre servicios, e impide la generación de inteligencia consolidada.
- 2. Obsolescencia en seguridad y control de acceso:** los modelos de seguridad perimetrales resultan insuficientes ante las amenazas avanzadas del entorno digital. Se observan deficiencias en el cifrado en reposo y en tránsito, así como una gestión de identidades fragmentada que dificulta la trazabilidad y la auditoría de accesos exigidas por la Ley N°21.663 [2], especialmente en escenarios con múltiples actores institucionales.
- 3. Escalabilidad limitada:** la infraestructura on-premise carece de la elasticidad necesaria para soportar picos de demanda ciudadana, lo que provoca latencias inaceptables o denegación de servicio en periodos críticos y compromete la continuidad operacional exigida por la Ley N°21.180 [1].

1.3 Estrategia o enfoque de la solución propuesta

La estrategia consiste en el diseño y validación de un Framework de Arquitectura de Referencia para Big Data, basado en el paradigma de Data Lake sobre infraestructura de nube, con un enfoque transversal de “Security & Privacy by Design” [6]. La propuesta se estructura como una arquitectura multicapa desacoplada, en la que cada capa puede evolucionar de forma independiente: una capa de ingesta híbrida que conecta tanto sistemas legacy como fuentes modernas en modos batch y streaming; una capa de almacenamiento organizada en zonas progresivas de calidad; una capa de procesamiento analítico con capacidades de transformación, enriquecimiento y aprendizaje automático; una capa transversal de gobernanza y seguridad que materializa el control de acceso, el cifrado, el catálogo y el cumplimiento normativo; y una capa de servicios y consumo que expone la información de forma segura e interoperable. La Figura 1 presenta la arquitectura de referencia resultante.

Esta estrategia busca transformar el modelo actual de “silos inseguros” en una “plataforma de datos integrada y gobernable”, en la que la seguridad y la protección de datos no son una capa añadida a posteriori, sino una propiedad estructural del sistema, mapeada de forma explícita a los requisitos legales vigentes.

1.4 Objetivos del trabajo

Objetivo general: diseñar y validar un Framework de Arquitectura de Referencia (FAR) para la gestión segura de datos masivos en servicios públicos chilenos, que integre Data Lake, seguridad Zero Trust y Privacy by Design con cumplimiento normativo por diseño, y que sea replicable como blueprint estandarizado. Los objetivos específicos son:

- OE1.** Diagnosticar el estado actual de la gestión de datos en una institución pública piloto, estableciendo la línea base de cumplimiento normativo, rendimiento, incidentes de seguridad y calidad de datos.
- OE2.** Diseñar el FAR con sus cinco capas funcionales y la matriz de mapeo entre componentes arquitectónicos y artículos de las Leyes N°21.663, N°21.180 y N°21.719.
- OE3.** Modelar y demostrar el rendimiento esperado del FAR sobre una prueba de concepto real, con indicadores de latencia, throughput y calidad de datos.
- OE4.** Validar el FAR mediante juicio de expertos y evaluación comparativa entre la línea base y los resultados proyectados.
- OE5.** Documentar el FAR como blueprint replicable y elaborar el informe final con conclusiones y recomendaciones de adopción.

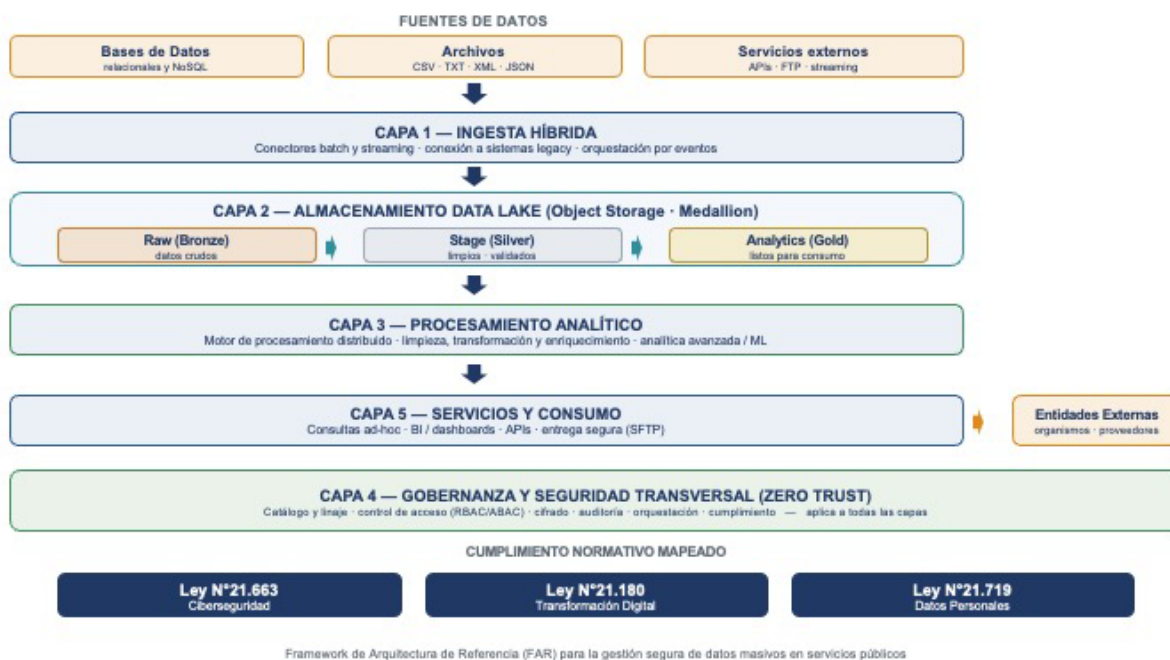


Figura 1. Arquitectura de Referencia (FAR) — visión de capas y componentes

2. Marco teórico y estado del arte

La fundamentación teórica de este trabajo se construye sobre la convergencia de cuatro pilares disciplinares: la arquitectura de Big Data, la gobernanza de datos [5], la ciberseguridad y la protección de datos personales. La gestión de Big Data se caracteriza por las denominadas “Tres V” —Volumen, Velocidad y Variedad—, a las que la literatura suele añadir la Veracidad y el Valor; estas propiedades exigen paradigmas arquitectónicos distintos de las bases de datos relacionales

tradicionales, orientados a la flexibilidad de esquema, el procesamiento distribuido y la escalabilidad horizontal.

2.1 Conceptos fundamentales

Data Lake y modelo de zonas (Medallion) [4]. A diferencia del Data Warehouse tradicional, que requiere un esquema definido antes de la carga (Schema-on-Write), el Data Lake se fundamenta en el paradigma Schema-on-Read, permitiendo la ingesta de datos en su formato nativo y postergando la imposición de estructura hasta el momento del consumo. Para evitar que el repositorio degenera en un “pantano de datos” (Data Swamp), se adopta la Arquitectura de Zonas o Medallion, que segrega la información en tres niveles progresivos de calidad: Bronze (datos crudos e inmutables tal como llegan de la fuente), Silver (datos limpios, deduplicados y validados) y Gold (datos agregados y listos para el consumo de negocio). Este modelo habilita la gobernanza incremental de la calidad y la trazabilidad completa del linaje desde el origen hasta el consumo.

Modelo de seguridad Zero Trust [15]. Formalizado en la publicación NIST SP 800-207, el modelo Zero Trust elimina el concepto de perímetro de red confiable y establece como principio que ninguna entidad —usuario, dispositivo o sistema— debe ser confiable por defecto, con independencia de su ubicación. Sus pilares fundamentales —verificación continua de identidad, acceso de mínimo privilegio y microsegmentación— resultan especialmente pertinentes para arquitecturas de datos gubernamentales, donde múltiples actores institucionales acceden a información sensible desde canales y ubicaciones diversas. Este enfoque es vital para cumplir la exigencia de la Ley N°21.663 respecto de la contención de amenazas laterales.

Privacidad desde el diseño (Privacy by Design) [6]. El marco de Ann Cavoukian establece que la protección de datos debe incorporarse en el diseño de los sistemas desde su concepción y no como una adición posterior. Sus siete principios fundacionales —entre ellos la privacidad proactiva, la privacidad como configuración predeterminada y la seguridad de extremo a extremo— se traducen, en el contexto de Big Data público, en técnicas de seudonimización y anonimización dinámica aplicadas en la capa de procesamiento antes de que los datos sean expuestos a los analistas.

Protección de Datos Personales y la Ley N°21.719 [3]. El cuarto pilar lo constituye el régimen chileno de protección de datos, modernizado por la Ley N°21.719, publicada el 13 de diciembre de 2024, que reemplaza el régimen de la Ley N°19.628 y entrará en plena vigencia el 1 de diciembre de 2026. La ley homologa el estándar nacional con el Reglamento General de Protección de Datos europeo (GDPR) y crea la Agencia de Protección de Datos Personales, órgano autónomo con facultades fiscalizadoras y sancionatorias. El tratamiento de datos por órganos públicos queda regido por los ocho principios del artículo 3° —licitud y lealtad, finalidad, proporcionalidad, calidad, responsabilidad proactiva, transparencia, confidencialidad y seguridad— e introduce instrumentos operativos que la arquitectura debe materializar: el Registro de Actividades de Tratamiento (RAT), el Delegado de Protección de Datos (DPO), la Evaluación de Impacto en Protección de Datos (EIPD) para tratamientos de alto riesgo, la notificación obligatoria de brechas y el ejercicio efectivo de los derechos del titular (acceso, rectificación, supresión, oposición, portabilidad y bloqueo). Esta ley transforma el principio de Privacy by Design, de un marco conceptual, en un mandato normativo trazable.

2.2 Estado del arte y casos de referencia

La revisión abarca experiencias internacionales de gobierno digital y arquitecturas de Big Data del sector público; para cada caso se contrastan su aporte (similitud) y sus límites respecto de esta propuesta (diferencia).

X-Road (Estonia) [9]. Capa nacional de interoperabilidad que conecta de forma segura más de 1.700 organizaciones públicas y privadas, con cifrado de extremo a extremo, firmas digitales distribuidas y el principio “una sola vez” (once-only). **Similitud:** el énfasis en la interoperabilidad segura entre instituciones y en la trazabilidad de los intercambios. **Diferencia:** X-Road es una arquitectura distribuida de transmisión de datos y no un Data Lake; no aborda el almacenamiento ni el procesamiento analítico de datos masivos, ni incorpora Privacy by Design sobre datos sensibles.

CityOS — Barcelona [10]. Plataforma de Big Data sobre Hadoop/Spark que integra datos de sensores IoT y de sistemas municipales para la toma de decisiones en una ciudad inteligente. **Similitud:** demuestra la viabilidad técnica de gestionar grandes volúmenes a escala en el sector público. **Diferencia:** su dominio se acota a la eficiencia operativa urbana y no define controles Zero Trust ni un modelo formal de protección de datos personales sensibles.

GDS Data Ethics Framework — Reino Unido [11]. Estándares para el uso ético y responsable de datos en el sector público, con listas de verificación éticas y legales a lo largo del ciclo de vida del dato. **Similitud:** responsabilidad, transparencia y uso legítimo de los datos ciudadanos. **Diferencia:** es un marco conceptual y ético, sin arquitectura técnica implementable ni componentes de seguridad o procesamiento.

Brecha identificada. En el sector privado existen implementaciones maduras de Data Lakes (banca y retail) y, en el ámbito público, casos avanzados de gobierno digital; no obstante, ninguno integra de manera simultánea (i) un modelo de gestión de datos masivos basado en Data Lake, (ii) seguridad Zero Trust transversal, (iii) Privacy by Design como principio estructural y (iv) un mapeo explícito a la normativa local chilena. A ello se suma que la mayoría de las soluciones desplegadas en Chile son implementaciones ad-hoc, sin una arquitectura de referencia estandarizada, lo que eleva los costos y los riesgos de seguridad. Esta brecha justifica y delimita el ámbito de la propuesta.

2.3 Gobernanza de datos y estándares de referencia

La gobernanza de datos constituye el sustrato que evita que un Data Lake degenere en un repositorio inservible y que da soporte al cumplimiento normativo. El cuerpo de conocimiento DAMA-DMBOK [5] organiza esta disciplina en áreas como la calidad, la arquitectura, la seguridad, los metadatos y el ciclo de vida del dato, todas ellas materializadas en la capa transversal del FAR. La arquitectura de referencia se apoya, además, en un conjunto de estándares internacionales que aportan rigor y comparabilidad:

- **TOGAF y ArchiMate [14]:** método y notación para el diseño de arquitectura empresarial y la documentación de los bloques de construcción del FAR.
- **NIST Big Data Reference Architecture (SP 1500-6) [4]:** modelo conceptual de referencia para sistemas de datos masivos que orienta la descomposición en capas.
- **NIST SP 800-207 (Zero Trust) [15] e ISO/IEC 27001:** marcos de seguridad de la información que fundamentan los controles de la capa transversal.

- **ISO/IEC 27701:** extensión de privacidad que articula la gestión de datos personales con los requisitos de la Ley N°21.719.

La adopción de estándares reconocidos no solo eleva la calidad del diseño, sino que facilita la auditoría, la certificación y la transferibilidad del blueprint entre instituciones, objetivos centrales de este trabajo.

3. Propuesta de solución e hipótesis

3.1 Propuesta técnica de solución

El FAR opera como un blueprint reutilizable, escalable y agnóstico a la tecnología, diseñado bajo el principio integrador “Security & Privacy by Design”. Se estructura en cinco capas funcionales y una capa transversal, cuya función se describe a continuación:

Capa 1 — Ingesta híbrida. recolecta datos desde fuentes heterogéneas (bases de datos legadas, plataformas transaccionales, APIs y archivos) en modos batch y streaming, con clasificación de la sensibilidad del dato en el origen para habilitar su tratamiento diferenciado aguas abajo.

Capa 2 — Almacenamiento Data Lake. organiza la información en zonas Bronze/Silver/Gold sobre almacenamiento de objetos, con versionado, políticas de ciclo de vida y formato columnar optimizado, garantizando durabilidad, resiliencia y trazabilidad.

Capa 3 — Procesamiento analítico. transforma y enriquece los datos, aplica reglas de negocio y habilita capacidades de aprendizaje automático, incorporando la seudonimización dinámica de datos personales antes de su exposición.

Capa 4 — Gobernanza y seguridad transversal. núcleo diferenciador de la propuesta; define el control de acceso granular (RBAC/ABAC), el cifrado en reposo y en tránsito, el catálogo y el linaje, la auditoría y el módulo de cumplimiento mapeado a la normativa, aplicando el modelo Zero Trust a todas las capas.

Capa 5 — Servicios y consumo. expone la información procesada mediante APIs estandarizadas, consultas analíticas, dashboards y entrega segura de archivos, habilitando la interoperabilidad con otras agencias del Estado.

Las innovaciones clave de la propuesta son: (i) el mapeo explícito entre componentes arquitectónicos y requisitos normativos; (ii) la modularidad que permite una adopción incremental según el nivel de madurez de cada institución; y (iii) la naturaleza agnóstica a la tecnología, que evita la dependencia de un proveedor específico.

3.2 Matriz de mapeo normativo (capas del FAR × leyes N°21.663 / N°21.180 / N°21.719)

La matriz operacionaliza la contribución central del trabajo, vinculando cada capa funcional con los requisitos normativos aplicables. Constituye el insumo del módulo de cumplimiento (Capa 4) y del gap analysis de la fase de validación, y se reutiliza como referencia normativa a lo largo del documento. En lo relativo a la Ley N°21.180, el requisito de interoperabilidad no se satisface únicamente mediante la exposición de APIs propias: su cumplimiento efectivo en el Estado de Chile se materializa mediante los habilitadores de gobierno digital ya operativos, en particular la Plataforma Integrada de Servicios Electrónicos del Estado (PISEE) como capa de intercambio interinstitucional, Clave Única como proveedor de identidad ciudadana —empleada en el módulo de ejercicio de derechos de la Sección 3.3— y FirmaGob para la suscripción electrónica de actos

administrativos. La Capa 5 del FAR se diseña, en consecuencia, para integrarse con dichos habilitadores en lugar de duplicar sus funciones, criterio que el panel de expertos respaldó de forma convergente (Anexo A) por evitar la proliferación de integraciones punto a punto entre organismos.

Capa del FAR	Ley N°21.663 — Ciberseguridad	Ley N°21.180 — Transf. Digital	Ley N°21.719 — Datos Personales
Capa 1 — Ingesta Híbrida	Art. 2°: autenticación de origen. Art. 3°: seguridad; cifrado en tránsito.	Interoperabilidad y neutralidad tecnológica: conectores con sistemas legacy.	Art. 3°: licitud, finalidad y minimización; clasificación de PII en el origen.
Capa 2 — Almacenamiento (Medallion)	Art. 3°: confidencialidad, integridad y disponibilidad; cifrado en reposo.	Integridad y conservación del expediente electrónico.	Art. 3°: seguridad y calidad; seudonimización; protección de datos sensibles.
Capa 3 — Procesamiento Analítico	Art. 2°: integridad y trazabilidad de procesos.	Calidad del dato para procedimientos electrónicos.	Anonimización/seudonimización dinámica; EIPD para alto riesgo.
Capa 4 — Gobernanza y Seguridad (Zero Trust)	Art. 7°: deberes generales. Art. 8°: SGSI y continuidad. Art. 9°: reporte al CSIRT.	Seguridad de procedimientos electrónicos.	Art. 3°: responsabilidad y confidencialidad; RAT, DPO, brechas; RBAC/ABAC.
Capa 5 — Servicios y Consumo	Art. 2°: autenticación y confidencialidad en APIs.	Interoperabilidad: exposición a otras agencias.	Derechos ARCO+; transparencia; transferencias reguladas.

3.3 Mecanismo de ejercicio de derechos del titular (ARCO+)

El cumplimiento de la Ley N°21.719 no se agota en el mapeo normativo: exige que la arquitectura provea mecanismos concretos para que el titular ejerza sus derechos de acceso, rectificación, cancelación (supresión), oposición, portabilidad y bloqueo (ARCO+). Dado que el caso BRSALUD trata datos de salud —la categoría más sensible bajo la ley—, la Capa 5 incorpora un Módulo de Ejercicio de Derechos del Titular con el siguiente flujo de diseño:

- **Autenticación e ingreso de la solicitud:** el ciudadano se autentica (por ejemplo, mediante Clave Única o el proveedor de identidad institucional) y registra su solicitud a través de un portal o API dedicada, quedando trazada con sello de tiempo.
- **Localización de datos vía catálogo y linaje:** la solicitud se resuelve consultando el catálogo de datos (Glue Data Catalog / Purview) y el linaje, que permiten identificar todas las instancias del dato del titular a lo largo de las zonas Raw, Stage y Analytics.
- **Resolución por tipo de derecho:** acceso → exportación del conjunto de datos del titular; rectificación → corrección propagada por el linaje; cancelación/supresión → borrado o anonimización respetando las excepciones de retención legal; oposición → marcado de exclusión del tratamiento; portabilidad → entrega en formato interoperable (JSON/CSV); y bloqueo → restricción temporal de acceso.
- **Gobernanza del proceso:** cada solicitud se enruta al Delegado de Protección de Datos (DPO), se registra en el Registro de Actividades de Tratamiento (RAT), se audita (CloudTrail / Azure Monitor) y se resuelve dentro de los plazos legales.

De este modo, los derechos ARCO+ que la matriz asigna a la Capa 5 dejan de ser una declaración normativa para convertirse en un componente arquitectónico verificable, cerrando la brecha entre lo que el diseño declara cumplir y lo que efectivamente implementa.

3.4 Hipótesis de trabajo y variables

Hipótesis principal: la implementación del FAR en una institución pública chilena incrementará el Índice de Cumplimiento Normativo (ICN) en al menos 30 puntos porcentuales respecto de la línea base previa a su adopción, alcanzando un $ICN \geq 90\%$ sobre los requisitos aplicables de las Leyes N°21.663, N°21.180 y N°21.719, sin degradación del rendimiento operativo del sistema de datos.

Hipótesis secundaria: el FAR es replicable en instituciones públicas de diferente tamaño y dominio con adaptaciones que no superen el 30% de sus componentes originales, lo que lo hace viable como estándar transferible entre organismos del Estado.

Operacionalización del ICN y umbral de contraste. El Índice de Cumplimiento Normativo se define como el cociente entre los requisitos normativos aplicables efectivamente satisfechos y el total de requisitos aplicables, expresado en porcentaje: $ICN = (\text{requisitos satisfechos} / \text{requisitos aplicables}) \times 100$. Los requisitos se derivan de la matriz de mapeo normativo (Sección 3.2) y se valoran mediante una rúbrica estructurada de forma graduada (0 = no cumple, 1 = cumple parcialmente, 2 = cumple), tanto para la línea base como para el FAR implementado. El umbral de “significativamente superior” se fija en un incremento ≥ 30 puntos porcentuales del ICN; complementariamente, la validez del cumplimiento declarado se contrasta con el juicio experto exigiendo el consenso del panel de expertos. Dado que el panel es reducido (3 a 4 especialistas, $n \leq 5$), se adopta el criterio de Lynn (1986), que exige unanimidad (índice de validez de contenido $I-CVI = 1,00$) para validar cada requisito, reservándose el umbral referencial de $\geq 80\%$ ($\geq 0,80$) para paneles de seis o más expertos. Estos umbrales constituyen el criterio de decisión que permite determinar, de forma no ambigua, la aceptación o el rechazo de la hipótesis. Refinamiento del instrumento. El juicio experto reportado en el Anexo A identificó de forma convergente —los tres evaluadores coincidieron en el ítem H1.2— dos líneas de robustecimiento del ICN que se recogen aquí: en primer lugar, ponderar los requisitos por criticidad, de modo que aquellos cuyo incumplimiento comporta un riesgo mayor —por ejemplo, los asociados al tratamiento de datos sensibles o a la notificación de brechas— tengan un peso superior al de requisitos de carácter formal, evitando que la agregación simple compense un incumplimiento crítico con varios cumplimientos menores; y en segundo lugar, documentar de forma explícita la evidencia que sustenta la valoración de cada requisito, de manera que el índice sea auditable y reproducible por un tercero. Ambos refinamientos se incorporan como línea de evolución del instrumento y se declaran, asimismo, entre las limitaciones del trabajo (Sección 7).

Variables. La variable independiente es la implementación del FAR (integración de Data Lake, Zero Trust y Privacy by Design). Las variables dependientes son el índice de cumplimiento normativo (ICN, escala porcentual sobre los requisitos aplicables), el rendimiento operativo (tiempos de consulta y procesamiento en milisegundos) y el nivel de incidentes de seguridad (número de vulnerabilidades identificadas en auditoría).

3.5 Resultados esperados e impacto

El principal resultado esperado del trabajo es un Framework de Arquitectura de Referencia validado y documentado como blueprint replicable, acompañado de la matriz de mapeo normativo, los diagramas de arquitectura y despliegue, y la evidencia empírica de su aplicación sobre un proceso real del Estado. A nivel técnico, se espera demostrar que el modelo permite reproducir fielmente la lógica de negocio de un proceso crítico, sostener su ventana de ejecución dentro de los plazos regulatorios y elevar el nivel de cumplimiento normativo respecto de la línea base, sin degradación del rendimiento.

A nivel de impacto, los resultados tienen un interés público directo. Para la institución piloto, el FAR habilita una plataforma de datos gobernada que reduce los riesgos de seguridad y de incumplimiento, y que mejora la trazabilidad y la calidad de los datos que sustentan beneficios ciudadanos. Para el conjunto del Estado, el carácter replicable del blueprint permite que cualquier servicio público adopte el marco con mínimas modificaciones, evitando el costo y el riesgo de diseñar arquitecturas ad-hoc. A nivel nacional, la propuesta contribuye a la modernización del Estado y a la implementación efectiva de las Leyes N°21.663, N°21.180 y N°21.719; y, a nivel regional, es transferible a países de América Latina con marcos de protección de datos emergentes (Colombia, Argentina y Perú), lo que amplía su potencial de adopción más allá de las fronteras nacionales.

En síntesis, el trabajo no solo aporta un artefacto técnico, sino un instrumento de política pública digital: un estándar de referencia que articula eficiencia operativa, seguridad y cumplimiento normativo en un único marco accionable.

4. Metodología de validación: Design Science Research (DSR)

La validación se enmarca metodológicamente en el paradigma Design Science Research (DSR), formalizado por Hevner et al. [7] y operacionalizado por Peffers et al. en el Design Science Research Methodology (DSRM) [8]. A diferencia de la investigación orientada a explicar fenómenos existentes, la DSR es un paradigma orientado a la resolución de problemas cuyo resultado es la creación y evaluación de un artefacto —en este caso, el FAR— que aborda una necesidad organizacional identificada. En este trabajo, el artefacto se construye y evalúa sobre una prueba de concepto (PoC) real implementada en AWS, cuyas métricas observadas se complementan con juicio experto, lo que legitima la evaluación sin requerir un despliegue productivo y mantiene la coherencia con los objetivos específicos OE1–OE5 y con las tres fases de validación.

4.1 Caso de referencia: el proceso BRSALUD del Instituto de Previsión Social

Las seis actividades del DSRM se instancian en un proceso de negocio representativo del Instituto de Previsión Social (IPS): la Bonificación de Salud (BRSALUD), que otorga la exención del pago de la cotización de salud a los pensionados que cumplen los requisitos legales [12]. El proceso construye mensualmente el universo de beneficiarios a partir de múltiples regímenes (D.L. N°3.500, sistema de reparto de las ex cajas administradas por el IPS, beneficiarios PGU y Bono por Hijo, y pensiones de las Leyes N°16.744 y N°19.234) y aplica las reglas del beneficio: ser pensionado, haber cumplido 65 años, pertenecer a los cuatro primeros quintiles según el Puntaje de Focalización Previsional (PFP) y cumplir el requisito de residencia. El resultado son nóminas en archivos planos que deben entregarse el último día hábil de cada mes a las entidades pagadoras (AFP, IPS, ISL, compañías de seguros y mutuales).

Su carácter representativo radica en que integra alrededor de cincuenta fuentes heterogéneas (con un volumen inicial estimado de 10 TB), trata datos personales sensibles (salud, datos socioeconómicos del Registro Social de Hogares y focalización) y está sujeto a una obligación de continuidad regulatoria, lo que lo convierte en un banco de pruebas idóneo para el FAR [12], [13].

4.2 Resultado de las seis actividades del DSRM

Actividad 1 — Identificación del problema y motivación

El diagnóstico del proceso BRSALUD evidencia tres dimensiones críticas. Primero, una fragmentación e ingesta heterogénea: el proceso integra bases Oracle institucionales (esquemas

SPS/CONCESIO), archivos CSV y TXT de ancho fijo, una declaración jurada obtenida desde un sistema offline (URL → XML → TXT) y archivos transferidos por FTP desde organismos externos (Registro Social de Hogares del MINDESIO e Informe de Residencia de la PDI), con parte de la ingesta realizada de forma manual. Segundo, el tratamiento de datos personales sensibles —RUN, nombres, tipo de pensión, condición de invalidez, datos socioeconómicos y focalización— sin controles centrados en el dato, cifrado granular ni trazabilidad. Tercero, una obligación de continuidad regulatoria: la entrega de las nóminas el último día hábil (norma SUPEN), cuyo incumplimiento afectaría la exención de salud de cientos de miles de pensionados. Esta actividad constituye el levantamiento de la línea base del caso (OE1, Fase 1).

Actividad 2 — Definición de objetivos de la solución

A partir del problema y de los requisitos normativos se definen objetivos funcionales y no funcionales medibles. Los objetivos funcionales son centralizar las ~50 fuentes en un Data Lake de tres capas; automatizar el pipeline de extremo a extremo (del disparo del último día hábil a la entrega de nóminas); reproducir fielmente la lógica de negocio (construcción del universo, exclusiones por BENREG/APS/PBS, validación de residencia por tres criterios, cotizaciones ≥ 240 y PFP ≤ 1876); generar los archivos BRSALUD, de extinciones y TAPS con su estructura exacta en ZIP con MD5; y habilitar los informes de gestión IG-BRSALUD-001 a 006. Los objetivos no funcionales y normativos comprenden el cifrado en reposo y tránsito, el control de acceso de mínimo privilegio, el versionado, el linaje y el catálogo, la seudonimización de PII, la auditoría y la continuidad, derivados de los principios de la Ley N°21.719, del deber de seguridad y reporte de la Ley N°21.663 y de la interoperabilidad de la Ley N°21.180. Los criterios de aceptación del proyecto exigen Raw 100% ingerida, Stage 99,5% en Parquet catalogado y Analytics consultable [13].

Actividad 3 — Diseño y desarrollo del artefacto

El artefacto instancia el FAR para BRSALUD sobre servicios de AWS, documentado con notación ArchiMate. La siguiente tabla resume la materialización de cada capa:

Capa del FAR	Servicios AWS	Instanciación en el caso BRSALUD
Capa 1 — Ingesta Híbrida	AWS Glue, DataSync, Lambda	Conectores a bases Oracle institucionales (esquemas SPS/CONCESIO), archivos CSV/TXT/XML y FTP (RSH, PDI). Disparo automático el último día hábil vía Glue Workflow.
Capa 2 — Almacenamiento (Medallion)	Amazon S3, S3 Versioning, ciclo de vida	Zonas Raw → Stage → Analytics (Bronze/Silver/Gold). Parquet particionado por institución pagadora (extafo).
Capa 3 — Procesamiento Analítico	AWS Glue Jobs (PySpark), Glue DataBrew	Job de construcción de universo (edad ≥ 65 , exclusiones, residencia y PFP ≤ 1876) y job de generación de nóminas (línea de ancho fijo, TXT → ZIP → MD5).
Capa 4 — Gobernanza y Seguridad	Lake Formation, Glue Data Catalog, IAM, KMS, CloudTrail, Step Functions	Catálogo y linaje, mínimo privilegio (RBAC), cifrado, auditoría y orquestación. Cumplimiento mapeado a las Leyes N°21.663, N°21.180 y N°21.719.
Capa 5 — Servicios y Consumo	Amazon Athena, QuickSight, Filesync/SFTP	Consultas ad-hoc, dashboards de gestión (IG-BRSALUD-001 a 006) y entrega interoperable de nóminas a las entidades pagadoras.

El pipeline opera de forma desacoplada: un disparador del último día hábil inicia un Glue Workflow que lee la zona Raw, ejecuta la unión y el filtrado de RUN válidos, escribe un Parquet particionado por institución pagadora en la zona Stage, procesa en paralelo un job por institución y, finalmente, consolida los archivos planos, los comprime y calcula su MD5 en la zona Analytics. Dos componentes concentran la lógica de negocio: el job de construcción del universo (que unifica los regímenes, elimina fallecidos, calcula la edad desde el RSH, aplica exclusiones y valida residencia

y focalización) y el job de generación de nóminas (que arma la línea de ancho fijo por beneficiario y produce el archivo por entidad con su ZIP y MD5) [12].

Actividad 4 — Demostración

El artefacto se demuestra ejecutando corridas mensuales reales sobre una prueba de concepto implementada en AWS (buckets poc-bd-raw, poc-bd-stage y poc-bd-analytics). La ejecución completa del Glue Workflow se realiza en torno a cuatro minutos: la ingesta desde la zona Raw toma cerca de un minuto con cuarenta segundos y la consolidación final cerca de un minuto con veinte segundos. La corrida produce las nóminas con el patrón de nombre esperado (por ejemplo, brsalud{aaaamm}.ext), particionadas por entidad pagadora y cargadas en la zona Analytics con su MD5. En la corrida mensual de la PoC, la zona Raw ingestó 28 fuentes de datos y 24,8 millones de registros (12,5 GB) en 18 minutos; la zona Stage validó la totalidad de los registros sin rechazos (calidad del 100%, superando el criterio de 99,5%) en 9,6 minutos; y la zona Analytics, consultable, produjo 90.117 registros BRSALUD, 1.500.000 extinciones y 55 archivos TAPS, con una latencia de consulta cercana a 200 ms. El pipeline completo se ejecutó de extremo a extremo en 36 minutos, por debajo de la ventana objetivo, con un throughput aproximado de 48.000 registros por segundo y un costo cercano a 18 USD por corrida. La verificación funcional por reconciliación arrojó una coincidencia del 73,3% entre las nóminas generadas por el FAR (90.117 beneficiarios) y la nómina oficial del período (123.000 registros); las 32.883 diferencias se explican por la inconsistencia de los insumos externos de la PDI, que el proceso vigente resuelve mediante manipulación manual. La siguiente tabla contrasta los criterios de aceptación definidos en la Actividad 2 con los resultados obtenidos en la PoC:

Zona / criterio	Meta	Resultado (PoC)
Raw — ingesta de fuentes	100%	28 de ~50 fuentes (24,8 M registros; 18 min)
Stage — calidad / conformidad	≥ 99,5%	100% (0 rechazos; 9,6 min)
Analytics — disponibilidad	Consultable	Consultable (90.117 BRSALUD; 4,5 min)
Rendimiento y reconciliación	≤ 45 min; ≥ 70%	End-to-end 36 min; coincidencia 73,3%

Nota: los valores corresponden a la corrida mensual observada en la PoC. La completitud de fuentes (28 ingestadas) debe interpretarse frente al universo total del proceso (~50 fuentes), cuya cobertura restante constituye trabajo de integración pendiente; y el índice de coincidencia funcional (73,3%) está condicionado por la calidad de los insumos externos (PDI).

Actividad 5 — Evaluación

La evaluación del artefacto se conduce conforme a los criterios de Hevner et al. —relevancia, rigor y utilidad— mediante la triangulación de cuatro fuentes de evidencia: (i) verificación funcional, comparando las nóminas generadas por el FAR con las del proceso heredado para un mismo período; (ii) verificación de cumplimiento, mediante el cálculo del Índice de Cumplimiento Normativo (ICN) definido en la Sección 3.4 sobre la matriz de mapeo, donde el tratamiento de datos sensibles activa la exigencia de EIPD, seudonimización y RAT; (iii) contraste de las métricas reales de la PoC con los criterios de aceptación (Raw 100%, Stage 99,5%, Analytics consultable); y (iv) juicio de expertos, mediante un panel de 3 a 4 especialistas del sector público TI, la academia y los organismos reguladores. Los resultados del juicio experto se reportan de forma íntegra en el Anexo A. El instrumento, compuesto por 59 ítems cerrados, se aplicó a un panel de tres especialistas con

perfiles complementarios (dirección de tecnología institucional, arquitectura de plataformas transversales del Estado y arquitectura empresarial e interoperabilidad). Sobre 177 respuestas válidas, el 78,5% correspondió al valor máximo de la escala y no se registró ninguna respuesta en el tramo desfavorable, con una media global de 4,79 sobre 5. La totalidad de los ítems alcanzó un I-CVI de 1,00, de modo que ambas hipótesis obtuvieron un S-CVI/Ave de 1,00 y satisfacen el criterio de Lynn (1986) exigible a paneles reducidos: H1 se considera validada, y H2 validada en sus dimensiones de diseño y factibilidad, con la precisión unánime de que la confirmación cuantitativa del umbral de adaptación requiere una segunda instanciación. El panel formuló asimismo seis áreas de mejora convergentes —ninguna calificada como impedimento— que se sistematizan en el Anexo A. La evaluación incorpora además el tratamiento de los riesgos del proyecto: la mitigación del “pantano de datos” mediante catálogo y gobernanza, la calidad de datos de origen mediante jobs de validación en la zona Stage, y el riesgo de seguridad mediante control de acceso basado en roles, cifrado y auditorías periódicas [13].

El resultado de la verificación de cumplimiento se expresa mediante el ICN, contrastando la línea base con el FAR para cada ley. Aplicando la rúbrica graduada (0 = no cumple, 1 = cumple parcialmente, 2 = cumple) sobre los 25 requisitos aplicables derivados de la matriz de mapeo (Sección 3.2), el ICN global ponderado se eleva de un 18% en la línea base a un 92% con el FAR, un incremento de +74 puntos porcentuales. Desagregado por ley, el ICN con el FAR alcanza 92,9% en la Ley N°21.663 (línea base 14,3%; incremento de +78,6 p.p.), 96,2% en la Ley N°21.719 (línea base 11,5%; +84,6 p.p.) y 80,0% en la Ley N°21.180 (línea base 40,0%; +40,0 p.p.). Dado que el incremento global (+74 p.p.) supera el umbral de 30 p.p. y el ICN con el FAR (92%) supera el 90% exigido, se verifica la hipótesis principal (H1). Los requisitos de naturaleza organizacional —en particular la notificación de brechas y de incidentes a la autoridad— se valoraron de forma conservadora como cumplimiento parcial, dado que el FAR aporta el soporte técnico pero su cumplimiento pleno exige medidas institucionales. El detalle por ley se resume en la siguiente tabla:

Ley	ICN línea base	ICN con FAR	Incremento (Δ)
N°21.663 — Ciberseguridad	14,3%	92,9%	+78,6 p.p.
N°21.180 — Transf. Digital	40,0%	80,0%	+40,0 p.p.
N°21.719 — Datos Personales	11,5%	96,2%	+84,6 p.p.
Global (ponderado)	18,0%	92,0%	+74,0 p.p.

Nota: los valores de línea base provienen del diagnóstico (OE1) y los del FAR de la evaluación sobre la PoC; el ICN global (92%) se pondera por el número de requisitos aplicables de cada ley (13, 7 y 5 para las Leyes N°21.719, N°21.663 y N°21.180, respectivamente). Los requisitos organizacionales de notificación a la autoridad se valoran como cumplimiento parcial por tratarse de medidas que exceden el alcance técnico de la PoC.

Actividad 6 — Comunicación

El artefacto se comunica como un blueprint replicable (Sección 5) a través de los entregables del proyecto —documento de arquitectura, código de los pipelines, catálogo de datos, manual de gobernanza, dashboards y transferencia de conocimiento— y, en el plano académico, mediante el informe final de tesina y su defensa. La transferibilidad es directa a otros procesos del IPS (PGU, APS), a otros organismos del Estado y a países de la región con regulaciones emergentes similares.

4.3 Metodología de trabajo e instrumentos

El proyecto se enmarca en una investigación aplicada con enfoque mixto (cuantitativo-cualitativo), desarrollada en cuatro etapas y articulada con las seis actividades del DSRM. La recolección y el análisis de datos se apoyan en los siguientes métodos e instrumentos:

- **Revisión sistemática de la literatura:** búsqueda estructurada en IEEE Xplore, ACM Digital Library y Scopus (criterios: últimos años, sector público, Big Data, Zero Trust y Privacy by Design) para fundamentar el marco teórico y el estado del arte.
- **Análisis documental normativo:** estudio de las Leyes N°21.663, N°21.180 y N°21.719 para construir la matriz de requisitos aplicables y el módulo de cumplimiento.
- **Entrevistas semiestructuradas:** levantamiento de la arquitectura actual, los flujos de datos y la línea base de métricas con los responsables de TI de la institución piloto.
- **Diseño y modelado con notación ArchiMate:** construcción iterativa de las cinco capas del FAR, con revisiones periódicas con el profesor guía.
- **Pruebas y simulaciones:** corridas reales sobre la prueba de concepto en nube y simulaciones de escenarios de carga representativos del volumen institucional.
- **Juicio de expertos:** entrevistas estructuradas a un panel de 3 a 4 especialistas del sector público TI, la academia y los organismos reguladores, con triangulación frente a las métricas cuantitativas. El panel efectivamente constituido reunió a tres especialistas y sus resultados agregados —concordancia, índices de validez de contenido y observaciones cualitativas— se reportan en el Anexo A.

La rigurosidad de la evaluación se sustenta, así, en la triangulación de tres fuentes de evidencia —métricas cuantitativas del modelado y la PoC, juicio experto estructurado y verificación de cumplimiento mediante la matriz de mapeo—, lo que fortalece la validez de constructo y mitiga el sesgo de una única fuente de validación.

5. Blueprint replicable del FAR

Uno de los entregables centrales del trabajo es la formalización del FAR como un blueprint en formato estándar, alineado a TOGAF/ArchiMate, a la NIST Big Data Reference Architecture [4], al AWS Well-Architected Framework [16], al modelo Zero Trust NIST SP 800-207 [15] y a Privacy by Design, de modo que cualquier servicio público pueda adoptarlo con mínimas modificaciones. El principio de replicabilidad es directo: la estructura de cinco capas (Figura 1), los componentes transversales de gobernanza y seguridad, y el mapeo normativo (Sección 3) permanecen fijos; la institución adoptante solo configura un conjunto acotado de parámetros propios de su dominio. Así, el esfuerzo de adopción se reduce a la parametrización y a la implementación incremental, no al rediseño arquitectónico.

5.1 Ficha del blueprint

Atributo	Detalle
Nombre	Blueprint FAR — Gestión Segura de Datos Masivos en Servicios Públicos
Versión	1.0 (2026)
Propósito	Marco de referencia replicable para implementar un Data Lake gobernado, seguro y conforme a la normativa chilena, adoptable mediante parametrización
Audiencia	Equipos de TI, arquitectura, seguridad y datos de cualquier servicio público chileno

Atributo	Detalle
Marcos base	TOGAF / ArchiMate 3.2 · NIST Big Data Reference Architecture (SP 1500-6) · AWS Well-Architected · Zero Trust (NIST SP 800-207) · Privacy by Design
Normativa	Leyes N°21.663, N°21.180 y N°21.719 · ISO/IEC 27001 y 27701
Modelo de adopción	Incremental por capas (Raw → Stage → Analytics), según nivel de madurez
Tecnología	Agnóstica; implementación de referencia sobre nube (AWS / Azure)

5.2 Principios, modelo de adopción y bloques de construcción

El blueprint se rige por seis principios rectores: Security & Privacy by Design (la seguridad y la privacidad son intrínsecas a cada capa), agnosticismo tecnológico (los componentes se definen funcionalmente y la implementación es sustituible), modularidad e incrementalidad (cada capa se adopta de forma independiente), mínimo privilegio y Zero Trust (ningún acceso es confiable por defecto), cumplimiento por diseño (cada componente se mapea a artículos de las leyes) e interoperabilidad (exposición estandarizada de datos para el intercambio entre organismos). La adopción se realiza en tres pasos —adoptar la estructura fija, parametrizar (sección 5.4) e implementar incrementalmente— y se organiza por niveles de madurez que habilitan capas de forma progresiva sin rehacer lo construido:

Nivel	Alcance	Capas habilitadas
Inicial	Ingesta y almacenamiento gobernados; cifrado, catálogo y control de acceso básicos	Capas 1–2 + base de Capa 4
Intermedio	Procesamiento Medallion y consumo analítico (consultas y dashboards)	+ Capas 3 y 5
Avanzado	Analítica avanzada/ML, automatización completa y cumplimiento certificable	Todas las capas + optimización continua

Cada capa es un bloque de construcción reutilizable, definido por su función y sus componentes agnósticos, con una implementación de referencia sobre AWS (la implementación equivalente en Azure se detalla en la Sección 6). Estos bloques no se modifican al adoptar el blueprint:

Capa del FAR	Componentes (agnóstico)	Implementación de referencia (AWS)
Capa 1 — Ingesta Híbrida	Conectores batch y streaming, conexión a legacy, clasificación en origen	AWS Glue · DataSync · Lambda · Glue Workflow
Capa 2 — Almacenamiento (Medallion)	Object storage por zonas Bronze/Silver/Gold, versionado, ciclo de vida, Parquet	Amazon S3 · S3 Versioning · S3 Lifecycle
Capa 3 — Procesamiento	Limpieza, validación, transformación, enriquecimiento y ML	AWS Glue Jobs (PySpark) · Glue DataBrew · SageMaker (opcional)
Capa 4 — Gobernanza y Seguridad	Catálogo y linaje, RBAC/ABAC, cifrado, auditoría, orquestación y cumplimiento	Lake Formation · Glue Data Catalog · IAM · KMS · CloudTrail · Step Functions
Capa 5 — Servicios y Consumo	Consulta ad-hoc, BI/dashboards, APIs y entrega segura	Amazon Athena · QuickSight · Filesync/SFTP · API Gateway

5.3 Componentes transversales

La capa transversal materializa tres dominios de control que aplican a todas las capas funcionales:

- **Gobernanza de datos:** catálogo, linaje, calidad y ciclo de vida del dato; designación de un Registro de Actividades de Tratamiento (RAT) y de un Delegado de Protección de Datos (DPO) cuando la escala o la sensibilidad del tratamiento lo exijan.

- **Seguridad Zero Trust:** verificación continua de identidad, microsegmentación por dominio institucional, cifrado en reposo y en tránsito, y auditoría continua de accesos.
- **Privacidad (Privacy by Design):** minimización, seudonimización y anonimización dinámica de datos personales, y Evaluación de Impacto (EIPD) para los tratamientos de alto riesgo.

5.4 Parámetros de adopción (lo que cada institución configura)

Esta es la sección clave para la replicabilidad: la arquitectura y los controles no cambian; solo se configuran los siguientes parámetros, propios del dominio de cada servicio. La columna “Ejemplo” ilustra cómo se instancian en el caso BRSALUD, lo que sirve de guía concreta para una nueva adopción. A los parámetros de dominio se añade un parámetro de integración con los habilitadores de gobierno digital del Estado: cada institución debe declarar qué servicios transversales consume y en qué capa los integra —típicamente la Plataforma Integrada de Servicios Electrónicos del Estado (PISEE) para el intercambio de información con otros organismos, Clave Única como proveedor de identidad en el módulo de ejercicio de derechos, y FirmaGob cuando la salida del proceso constituya un acto administrativo—, en lugar de construir integraciones punto a punto propias. Configurar este parámetro por la vía de los habilitadores existentes reduce el esfuerzo de adopción, evita duplicar capacidades ya provistas por el Estado y alinea la implementación con el requisito de interoperabilidad de la Ley N°21.180 mapeado en la Sección 3.2.

Parámetro	Descripción	Ejemplo (caso BRSALUD)	Dónde se configura
Inventario de fuentes	Catálogo de orígenes de datos	~50 fuentes (Oracle, CSV/TXT, FTP)	Capa 1 / catálogo
Clasificación de sensibilidad	Etiquetado de PII y datos sensibles por fuente	Salud, RSH, focalización (sensibles)	Capa 1 / gobernanza
Reglas de negocio	Filtros, exclusiones y validaciones del dominio	Edad ≥ 65, PFP ≤ 1876, residencia	Capa 3
Destinatarios y salidas	Formato y receptores de los entregables	Nóminas a AFP/IPS/ISL/CSV/mutuales	Capa 5
Frecuencia y disparador	Periodicidad y evento de ejecución	Mensual, último día hábil	Capa 1 / orquestación
Períodos de retención	Ciclo de vida del dato por zona y norma	Según política institucional y ley	Capa 2 (lifecycle)
Roles y accesos (RBAC)	Perfiles y mínimo privilegio por dominio	Roles por unidad del servicio	Capa 4 (IAM)
Llaves de cifrado	Gestión de claves por zona/dominio	Claves KMS por dominio	Capa 4 (KMS)
Informes y dashboards	Vistas de gestión requeridas	IG-BRSALUD-001 a 006	Capa 5 (BI)
Mapeo normativo específico	Artículos aplicables al servicio	21.663 / 21.180 / 21.719	Capa 4 (cumplimiento)

5.5 Roles (RACI) e indicadores de aceptación (KPIs)

La adopción requiere una asignación clara de responsabilidades entre los perfiles de ingeniero de datos, analista, arquitecto de soluciones y jefe de proyecto, que se formaliza en la siguiente matriz RACI:

Actividad / entregable	Ing. Datos	Analista	Arquitecto	PM
Diseño de arquitectura	C	I	A, R	I
Configuración de almacenamiento (zonas)	A, R	I	C	I
Pipelines de ingesta	A, R	I	C	I
Jobs de procesamiento	A, R	C	C	I
Catálogo de datos y linaje	R	C	A	I
Gobernanza y seguridad	C	I	A, R	I

Actividad / entregable	Ing. Datos	Analista	Arquitecto	PM
Dashboards e informes	C	A, R	C	I
Cumplimiento normativo (matriz / EIPD)	C	C	A, R	I
Transferencia de conocimiento	R	C	R	A

R = Responsable · A = Aprobador · C = Consultado · I = Informado.

El éxito de la adopción se verifica mediante indicadores objetivos y medibles, alineados con los criterios de aceptación del proyecto:

Indicador	Meta	Evidencia
Cobertura de ingesta	100% de las fuentes acordadas en Raw	Conteo de cargas vs. inventario
Calidad y optimización	≥ 99,5% de datos en Parquet catalogado (Stage)	Catálogo / jobs de validación
Disponibilidad analítica	Analytics consultable y archivos descargables	Pruebas de consulta / entrega
Cumplimiento normativo	Índice de conformidad superior a la línea base	Matriz de mapeo / gap analysis
Seguridad	Cifrado 100% y cero hallazgos críticos en auditoría	Auditorías periódicas
Rendimiento	Ventana de proceso dentro del SLA institucional	Métricas de ejecución (PoC)

5.6 Riesgos, mitigaciones y checklist de adopción

La gestión proactiva de riesgos es determinante para el éxito de la adopción. Los principales riesgos y sus mitigaciones son:

- **Pantano de datos (Data Swamp):** gobernanza estricta desde el inicio con catálogo y linaje documentados.
- **Calidad de datos de origen:** perfilado previo y jobs de validación en la zona Stage, con notificación a los equipos de origen.
- **Fallo de seguridad:** control de acceso de mínimo privilegio (RBAC), cifrado de extremo a extremo y auditorías periódicas.
- **Dependencia de proveedor:** definición agnóstica de los componentes e implementación sustituible, con formatos abiertos.

Checklist de adopción: (1) levantar el inventario de fuentes y clasificar la sensibilidad de los datos; (2) configurar las zonas de almacenamiento, el cifrado, el versionado y el ciclo de vida; (3) implementar los pipelines de ingesta con su disparador y periodicidad; (4) codificar las reglas de negocio del dominio en la capa de procesamiento; (5) habilitar la gobernanza y la seguridad (catálogo, linaje, RBAC, cifrado y auditoría); (6) completar el mapeo normativo y la EIPD cuando se traten datos sensibles; (7) configurar el consumo y la entrega de salidas a destinatarios; y (8) verificar los criterios de aceptación y los KPIs antes de pasar a producción; y (9) definir y probar el esquema de continuidad operacional del proceso, estableciendo los objetivos de tiempo de recuperación (RTO) y de punto de recuperación (RPO) admisibles, la estrategia de respaldo y restauración de las zonas del Data Lake, y la disponibilidad de un sitio o región alternativa que permita sostener la ventana regulatoria de entrega ante una indisponibilidad de la plataforma principal. Esta última verificación recoge una observación convergente del panel de expertos (Anexo A): en procesos con obligación de entrega en fecha fija, como BRSALUD, la continuidad no es un atributo deseable sino un requisito de cumplimiento, por lo que los valores de RTO y RPO deben derivarse del plazo regulatorio aplicable y no de una definición genérica de servicio.

5.7 Validación del blueprint

El blueprint fue ejercitado sobre la prueba de concepto real descrita en la Sección 4, aplicada al proceso BRSALUD del IPS. Este caso instancia las cinco capas, los componentes transversales y la mayoría de los parámetros de adopción de la sección 5.4, lo que demuestra la factibilidad de implementar íntegramente el marco en un proceso público con datos sensibles, fuentes heterogéneas y una obligación de continuidad mensual. No obstante, al tratarse de una única instanciación, este resultado evidencia la factibilidad del blueprint pero no confirma aún, por sí solo, el umbral de la hipótesis secundaria (adaptaciones $\leq 30\%$ de los componentes al replicar en otra institución): dicha verificación requiere una segunda instanciación. Como cierre del alcance restante se propone una segunda instanciación, al menos parcial, sobre otro proceso del IPS (por ejemplo, PGU) u otro organismo, que permita medir el porcentaje real de componentes adaptados y contrastarlo con el umbral comprometido. Esta lectura es coincidente con la del juicio experto reportado en el Anexo A: el panel validó la hipótesis secundaria (H2) por unanimidad —I-CVI = 1,00 en los cinco ítems, con una media de 4,80— y otorgó su valoración más alta precisamente al mecanismo que sustenta la replicabilidad: la separación entre la estructura invariante y el conjunto acotado de parámetros configurables, que los evaluadores identifican como la distinción formal entre una arquitectura de referencia y una arquitectura de solución. Dos de los tres evaluadores calificaron además el umbral de adaptación $\leq 30\%$ como probablemente conservador, fundándose en que la variabilidad entre instituciones se concentra en la capa de negocio, mientras que las capas de aplicación y tecnología y los controles transversales se conservan sin cambio. El panel fue, no obstante, unánime en que la confirmación cuantitativa del umbral requiere una segunda instanciación en un dominio institucional distinto, lo que refuerza la propuesta de cierre formulada en el párrafo precedente.

5.8 Gobierno de la arquitectura y prerequisites organizacionales

La adopción sostenible del FAR requiere, además de las condiciones técnicas descritas en las secciones precedentes, un conjunto de prerequisites organizacionales y un mecanismo de gobierno del propio marco. Esta subsección recoge una observación convergente del panel de expertos (Anexo A), cuyos integrantes coinciden —desde la dirección de tecnología institucional, la operación de plataformas transversales y la arquitectura empresarial— en que las dificultades de adopción de arquitecturas de referencia en el sector público son predominantemente organizacionales antes que tecnológicas.

Prerequisites organizacionales. Se identifican cuatro condiciones habilitantes: (i) la designación formal del Delegado de Protección de Datos, con atribuciones efectivas sobre las decisiones de tratamiento, condición sin la cual el módulo de ejercicio de derechos de la Sección 3.3 carece de responsable institucional; (ii) la constitución de una instancia de gobierno de datos con mandato institucional, capaz de resolver conflictos de propiedad y calidad del dato entre unidades; (iii) el patrocinio del jefe de servicio y no solo de la unidad de tecnología, dado que la adopción cruza fronteras organizacionales; y (iv) la formalización de convenios de intercambio con los organismos de origen, con niveles de servicio y contratos de datos comprometidos —esquema, frecuencia, criterios de calidad, tratamiento de rechazos y responsable de origen—, cuya ausencia explica buena parte de las fallas observadas en integraciones interinstitucionales y, en el caso BRSALUD, la brecha de reconciliación atribuida a los insumos externos.

Gobierno de la arquitectura. La conformidad de cada adopción con el marco requiere una instancia de revisión de arquitectura que verifique el cumplimiento de los bloques de construcción invariantes, criterios de conformidad verificables por nivel de madurez —de modo que una institución pueda declarar y demostrar el grado de adopción alcanzado— y un procedimiento de excepción documentada para los casos en que deba apartarse justificadamente del marco. Sin este mecanismo, la conformidad se erosiona ante la presión de plazos y cada implementación deriva

hacia una variante propia, con lo que el marco pierde su condición de estándar y se desvirtúa la replicabilidad postulada en la hipótesis secundaria.

Ciclo de vida del blueprint. Finalmente, el marco requiere una custodia institucional que asegure su versionado, su revisión periódica y su actualización ante cambios normativos —particularmente relevante ante la entrada en plena vigencia de la Ley N°21.719 en diciembre de 2026—, así como un mecanismo que permita a las instituciones adoptantes retroalimentar el estándar a partir de su experiencia de implementación. Se propone que dicha custodia resida en un organismo con rol rector en materia de gobierno digital, coherentemente con la recomendación unánime del panel de poner el blueprint y la matriz de mapeo normativo a disposición de la Secretaría de Gobierno Digital y de la red de arquitectos y jefaturas de TI del sector público.

6. Comparativo de servicios cloud: AWS vs. Microsoft Azure

Dado el carácter agnóstico del FAR, su implementación puede materializarse íntegramente en los dos principales proveedores de nube pública disponibles para el Estado de Chile: Amazon Web Services (AWS) y Microsoft Azure. Esta sección traslada el blueprint a una comparación concreta, identificando para cada capa y componente el servicio equivalente en ambas nubes y presentando los diagramas de despliegue de referencia por proveedor.

6.1 Contexto de residencia de datos en Chile

Ambos proveedores han establecido o anunciado regiones soberanas en Santiago, lo que es directamente relevante para el cumplimiento de la Ley N°21.719 en materia de residencia de datos. Microsoft inauguró su región Azure Chile Central en 2025, ya operativa, con residencia de datos local y certificaciones ISO 27001/27701 [18]. AWS anunció su región Sudamérica (Chile), con entrada en operación prevista para fines de 2026, y ya dispone de una AWS Local Zone en Santiago [17]. Cabe notar, además, que la Secretaría de Gobierno Digital (SGD) opera plataformas estatales críticas (Clave Única, FirmaGob, PISSE, DocDigital) sobre AWS, mientras que la integración nativa con Microsoft 365 favorece a Azure en instituciones que ya utilizan ese ecosistema.

6.2 Comparativo de servicios por capa del FAR

La siguiente tabla mapea cada componente del FAR a su servicio equivalente en AWS y en Azure, evidenciando que ambas plataformas cubren la totalidad de las capas.

Componente (capa del FAR)	AWS	Microsoft Azure
Ingesta / ETL (Capa 1)	AWS Glue · AWS DataSync	Azure Data Factory · Synapse Pipelines
Funciones serverless (Capa 1)	AWS Lambda	Azure Functions
Streaming y eventos (Capa 1)	Amazon Kinesis · EventBridge	Azure Event Hubs · Event Grid
Data Lake / object storage (Capa 2)	Amazon S3	Azure Data Lake Storage Gen2
Procesamiento Spark (Capa 3)	AWS Glue (PySpark) · Amazon EMR	Azure Synapse (Spark) · Azure Databricks
Machine Learning (Capa 3)	Amazon SageMaker	Azure Machine Learning
Catálogo y gobernanza (Capa 4)	Glue Data Catalog · Lake Formation	Microsoft Purview
Identidad y acceso (Capa 4)	AWS IAM	Microsoft Entra ID + Azure RBAC
Cifrado y llaves (Capa 4)	AWS KMS	Azure Key Vault
Auditoría y monitoreo (Capa 4)	AWS CloudTrail · CloudWatch	Azure Monitor · Log Analytics · Defender
Orquestación (Capa 4)	AWS Step Functions	Synapse Pipelines · Logic Apps
SQL serverless (Capa 5)	Amazon Athena	Azure Synapse Serverless SQL
BI y dashboards (Capa 5)	Amazon QuickSight	Microsoft Power BI
Entrega segura / SFTP (Capa 5)	AWS Transfer Family	Azure Blob Storage SFTP

6.3 Diagramas de despliegue por proveedor

Las Figuras 2 y 3 presentan el despliegue de referencia del FAR sobre AWS y sobre Azure, respectivamente, conservando la misma estructura de capas y la capa transversal de gobernanza y seguridad.

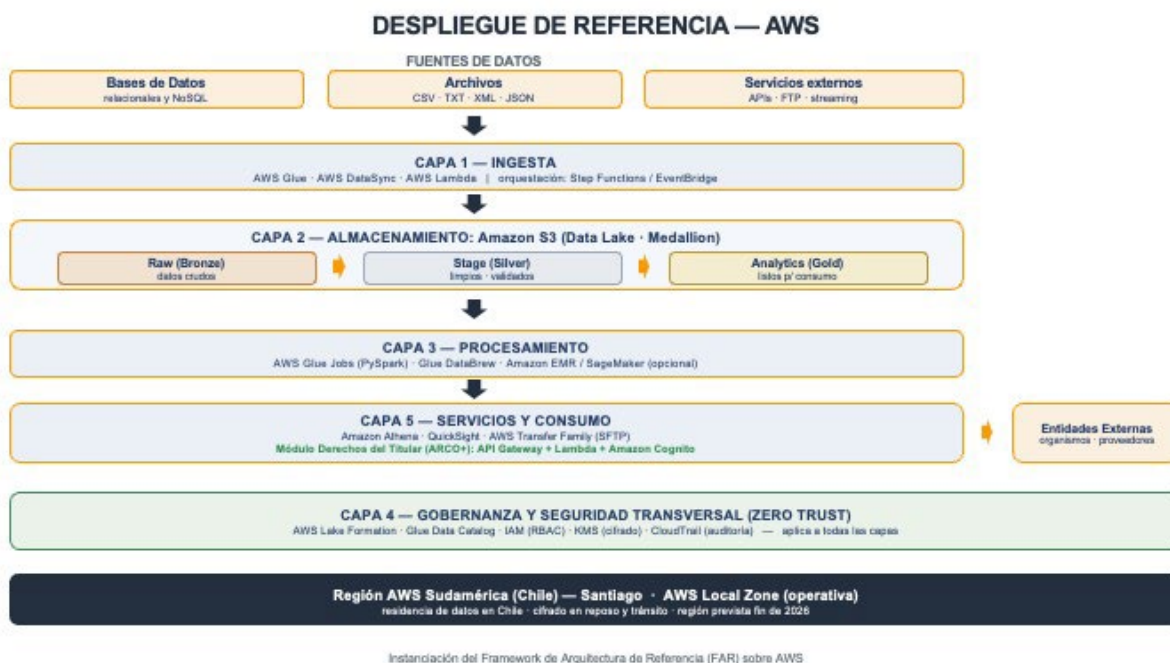


Figura 2. Diagrama de despliegue de referencia del FAR sobre AWS

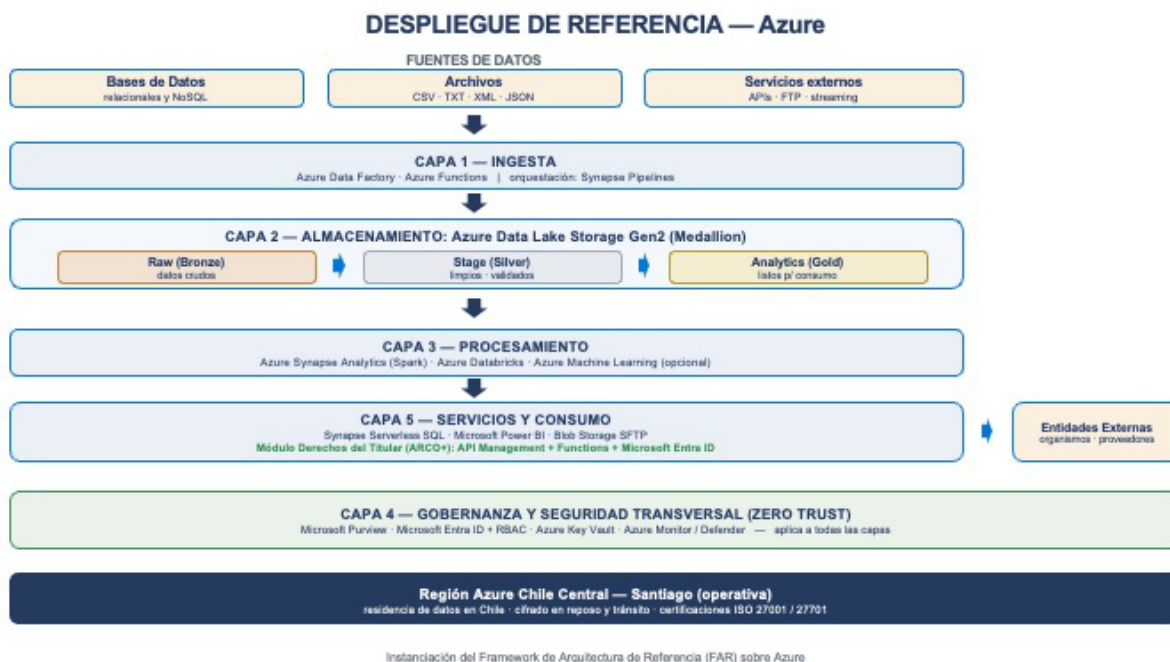


Figura 3. Diagrama de despliegue de referencia del FAR sobre Microsoft Azure

6.4 Comparación por dimensiones

A nivel de dimensiones clave, ambas nubes ofrecen capacidades equivalentes en almacenamiento, ETL, gobernanza, identidad, analítica y BI; las diferencias relevantes se concentran en la disponibilidad de región local y en la integración con el ecosistema Microsoft 365.

Dimensión	AWS	Microsoft Azure
Data Lake	Amazon S3	Azure Data Lake Storage Gen2
ETL y orquestación	AWS Glue + Step Functions	Data Factory / Synapse Pipelines
Catálogo y gobernanza	Glue Data Catalog + Lake Formation	Microsoft Purview
Identidad y acceso	AWS IAM	Microsoft Entra ID + RBAC
SQL serverless	Amazon Athena	Synapse Serverless SQL
BI / visualización	Amazon QuickSight	Microsoft Power BI
Región en Chile	Sudamérica (Chile): prevista fin de 2026; Local Zone Santiago operativa	Chile Central: operativa desde 2025
Residencia de datos	Sí (región / Local Zone)	Sí (Chile Central)
Integración con M365 / Power Platform	Limitada	Nativa (Power BI, Entra ID, Purview)
Certificaciones	ISO 27001/27701 · SOC 1/2/3	ISO 27001/27701 · SOC 1/2/3

6.5 Cumplimiento y criterios de selección

Ambas nubes ofrecen los controles necesarios para satisfacer el mapeo normativo del blueprint — cifrado en reposo y tránsito, control de acceso de mínimo privilegio, auditoría y trazabilidad, catálogo y gobernanza del dato, y residencia de datos en Chile—, por lo que son aptas frente a las Leyes N°21.663, N°21.180 y N°21.719. El diferenciador operativo relevante es la disponibilidad de región soberana local, hoy operativa en Azure y prevista para AWS hacia fines de 2026.

Dado que el FAR es agnóstico, la elección del proveedor depende de factores institucionales más que técnicos. Se recomienda ponderar: la residencia de datos y sus plazos; el ecosistema preexistente (las instituciones con fuerte adopción de Microsoft 365 obtienen sinergia con Azure, mientras que las plataformas estatales sobre AWS favorecen su continuidad); las competencias del equipo y la experiencia previa en cada plataforma; los convenios marco vigentes y los modelos de costo; y la portabilidad, que se preserva manteniendo la definición agnóstica del FAR mediante formatos abiertos como Parquet y lógica de procesamiento en PySpark.

6.6 Análisis comparativo por capa

Capa 1 — Ingesta. AWS Glue y AWS DataSync ofrecen un ecosistema maduro de conectores administrados, mientras que Azure Data Factory y Synapse Pipelines proveen una orquestación visual con amplia conectividad híbrida; ambas soportan ingesta batch y streaming (Kinesis/Event Hubs) y funciones serverless (Lambda/Functions). La paridad funcional es alta y la elección suele decidirse por la familiaridad del equipo con cada herramienta.

Capa 2 — Almacenamiento. Amazon S3 y Azure Data Lake Storage Gen2 son equivalentes en durabilidad, escalabilidad, versionado y políticas de ciclo de vida; ADLS Gen2 incorpora de forma nativa un espacio de nombres jerárquico orientado a analítica, mientras que S3 destaca por la amplitud de su ecosistema de integración. Ambos sustentan el modelo Medallion sin diferencias relevantes.

Capa 3 — Procesamiento. AWS Glue (Spark administrado) y Amazon EMR compiten con Azure Synapse (Spark) y Azure Databricks; este último, presente en ambas nubes, es un referente de

procesamiento analítico. La portabilidad de la lógica en PySpark hace que la migración entre proveedores sea de bajo costo en esta capa.

Capa 4 — Gobernanza y seguridad. Es la capa con mayor diferenciación de productos: AWS combina Lake Formation y Glue Data Catalog con IAM, KMS y CloudTrail, mientras que Azure consolida la gobernanza en Microsoft Purview e integra la identidad con Entra ID, el cifrado con Key Vault y el monitoreo con Azure Monitor y Defender. Purview ofrece una visión unificada de catálogo, linaje y clasificación que resulta atractiva para instituciones con ecosistema Microsoft.

Capa 5 — Servicios y consumo. Amazon Athena y Azure Synapse Serverless SQL proveen consulta ad-hoc sin servidor equivalente; en visualización, Amazon QuickSight compete con Microsoft Power BI, donde este último presenta una ventaja por su penetración en el sector público y su integración con Microsoft 365. La entrega segura de archivos se resuelve con AWS Transfer Family o con el endpoint SFTP de Azure Blob Storage.

7. Conclusiones y trabajo futuro

Este trabajo propuso y validó un Framework de Arquitectura de Referencia (FAR) para la gestión segura de datos masivos en servicios públicos chilenos, integrando Data Lake (Medallion), seguridad Zero Trust y Privacy by Design con un mapeo explícito a las Leyes N°21.663, N°21.180 y N°21.719. La validación se realizó mediante Design Science Research sobre una prueba de concepto real en AWS aplicada al proceso BRSALUD del IPS, demostrando la viabilidad técnica de la arquitectura y su capacidad para acreditar el cumplimiento normativo en un dominio con datos sensibles, fuentes heterogéneas y una obligación de continuidad mensual.

Las principales contribuciones son: (i) la arquitectura de referencia multicapa con cumplimiento por diseño; (ii) la matriz de mapeo normativo componente-por-ley; (iii) el blueprint replicable con su tabla de parámetros de adopción, que reduce el esfuerzo de adopción a la parametrización; y (iv) la validación empírica sobre un caso real del Estado, complementada con un comparativo de despliegue que demuestra la independencia tecnológica del marco entre AWS y Azure. La validación se sustenta en la triangulación de tres fuentes independientes de evidencia: la verificación de cumplimiento sobre la matriz de mapeo, que elevó el Índice de Cumplimiento Normativo desde un 18% de línea base hasta un 92% con el FAR (+74 p.p.); las métricas de la prueba de concepto, que acreditaron la ausencia de degradación del rendimiento (36 minutos de extremo a extremo, cerca de 48.000 registros por segundo y latencias de consulta próximas a 200 ms, dentro de la ventana regulatoria); y el juicio experto estructurado, aplicado a un panel de tres especialistas mediante un instrumento de 59 ítems, que arrojó una concordancia del 100% y un S-CVI/Ave de 1,00 para ambas hipótesis, satisfaciendo el criterio de Lynn (1986). En consecuencia, la hipótesis principal (H1) se considera validada, y la hipótesis secundaria (H2) validada en sus dimensiones de diseño y factibilidad, con la precisión —compartida por el panel y por el autor— de que la confirmación cuantitativa del umbral de adaptación $\leq 30\%$ requiere una segunda instanciación del blueprint en un dominio institucional distinto.

Limitaciones. La validación se realizó sobre un único proceso piloto y en un entorno de prueba de concepto, lo que delimita el alcance de las conclusiones en tres aspectos. Primero, la cobertura de la ingesta alcanzó 28 de las aproximadamente 50 fuentes del proceso y la reconciliación funcional arrojó una coincidencia del 73,3% respecto de la nómina oficial del período, diferencia atribuible a la calidad de los insumos externos; ambos constituyen trabajo de integración pendiente y no una limitación del diseño arquitectónico. Segundo, si bien el Índice de Cumplimiento Normativo cuenta

con evidencia empírica —92% con el FAR frente a un 18% de línea base—, su cálculo se sustenta en una rúbrica aplicada, por lo que su reproducibilidad se fortalecería mediante una doble valoración independiente y una ponderación de los requisitos por criticidad. Tercero, la hipótesis secundaria de replicabilidad, aun contando con el respaldo unánime del panel de expertos, no queda cuantitativamente confirmada con una sola instanciación y requiere un segundo caso en un dominio institucional distinto. A ello se añade que los números de artículo de la matriz normativa deben confirmarse contra los textos oficiales consolidados, considerando que la entrada en plena vigencia de la Ley N°21.719 en diciembre de 2026 podría introducir precisiones reglamentarias posteriores.

Trabajo futuro. Se proponen la extensión del FAR a otros procesos del IPS (PGU, APS) y a otros organismos del Estado; la automatización del módulo de cumplimiento normativo, que permita generar evidencia de conformidad de forma continua; la incorporación de capacidades de analítica avanzada y aprendizaje automático sobre la capa de consumo; y un estudio comparativo de costos totales de propiedad entre los despliegues en AWS y en Azure para distintos perfiles de institución.

Recomendaciones de adopción. Para los servicios públicos que adopten este marco se recomienda iniciar por el nivel de madurez Inicial sobre un proceso acotado pero representativo, priorizar la habilitación temprana de la capa transversal de gobernanza y seguridad —de modo que el cumplimiento se construya desde el primer dato ingerido y no como un esfuerzo posterior—, completar la Evaluación de Impacto en Protección de Datos antes de tratar información sensible, y formalizar la transferencia de conocimiento al equipo interno para asegurar la sostenibilidad de la plataforma. La adopción incremental, guiada por la tabla de parámetros y verificada con los KPIs definidos, reduce el riesgo y permite demostrar valor de forma temprana ante los stakeholders.

Referencias

- [1] Ministerio Secretaría General de la Presidencia, “Ley N°21.180: De Transformación Digital del Estado,” Diario Oficial de la República de Chile, 2019.
- [2] Ministerio del Interior y Seguridad Pública, “Ley N°21.663: Ley Marco de Ciberseguridad,” Diario Oficial de la República de Chile, 2024.
- [3] Ministerio Secretaría General de la Presidencia, “Ley N°21.719: Regula la protección y el tratamiento de los datos personales y crea la Agencia de Protección de Datos Personales,” Diario Oficial de la República de Chile, dic. 2024.
- [4] NIST Big Data Public Working Group, “NIST Big Data Interoperability Framework: Volume 6, Reference Architecture,” NIST, Special Publication 1500-6r2, 2019.
- [5] DAMA International, DAMA-DMBOK: Data Management Body of Knowledge, 2nd ed. Technics Publications, 2017.
- [6] A. Cavoukian, “Privacy by Design: The 7 Foundational Principles,” Information and Privacy Commissioner of Ontario, Canada, 2009.
- [7] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design Science in Information Systems Research,” MIS Quarterly, vol. 28, no. 1, pp. 75–105, 2004.
- [8] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A Design Science Research Methodology for Information Systems Research,” Journal of Management Information Systems, vol. 24, no. 3, pp. 45–77, 2007.
- [9] Estonian Information System Authority (RIA), “X-tee | The Data Exchange Layer of Estonian e-Governance,” 2026.
- [10] Barcelona City Council, “Smart City Barcelona — CityOS,” 2026.
- [11] Government Digital Service (GDS), “The Data Ethics Framework,” Gov.uk, 2020.
- [12] Instituto de Previsión Social (IPS), “Requerimientos de Negocio — Proceso Bonificación de Salud (BRSALUD),” documento interno del proyecto, 2026.

- [13] Andes Digital, “Declaración de Trabajo (SoW) para un Proyecto de Big Data — IPS,” documento interno del proyecto, 2026.
- [14] The Open Group, TOGAF Standard, 10th ed., 2022; y ArchiMate 3.2 Specification, 2023.
- [15] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, “Zero Trust Architecture,” NIST, Special Publication 800-207, 2020.
- [16] Amazon Web Services, “AWS Well-Architected Framework,” 2023.
- [17] Amazon Web Services, “Región AWS Sudamérica (Chile),” aws.amazon.com/es/local/chile, 2025–2026.
- [18] Microsoft, “Microsoft inaugura su primera Región de Datacenters en Chile (Chile Central),” News Center Latinoamérica, jun. 2025.
- [19] Microsoft, “Azure Well-Architected Framework,” 2024.

Anexo A. Validación por juicio experto del Framework de Arquitectura de Referencia (FAR)

Este anexo resume los resultados de la tercera fuente de evidencia de la estrategia de triangulación descrita en la Sección 4.3: el juicio experto estructurado. Reporta de forma trazable el método de agregación, los índices de validez de contenido y el criterio de decisión aplicado a cada hipótesis, y sistematiza las observaciones cualitativas convergentes del panel.

A.1 Propósito y alcance del anexo

El anexo cumple una función doble: reportar los resultados de la aplicación del instrumento al panel de especialistas, y traducir las observaciones cualitativas en modificaciones concretas al cuerpo de la tesina, de modo que la evaluación externa retroalimente efectivamente el artefacto. Es autocontenido y sus resultados se incorporan por referencia en las Secciones 3.4, 4.2 (Actividad 5), 4.3, 5.7 y 7.

A.2 Composición y perfil del panel

El panel se constituyó con tres especialistas, conforme al rango de 3 a 4 previsto en la Sección 4.3, mediante muestreo intencional por criterio. Los perfiles cubren de forma complementaria la dirección de tecnología en una institución pública (E1), la arquitectura de plataformas transversales del Estado (E2) y la arquitectura empresarial e interoperabilidad entre organismos (E3). Los criterios de inclusión aplicados fueron: (i) más de diez años de experiencia acreditada en tecnologías de información; (ii) experiencia directa en el sector público chileno; y (iii) responsabilidad demostrable en decisiones de arquitectura, integración o gobierno de datos, satisfechos simultáneamente por los tres evaluadores.

Cód.	Evaluador	Cargo y organización	Perspectiva aportada al panel	Exp. TI pública
E1	Paul Alexander Cook Moyano	Jefe TI, Subsecretaría de Telecomunicaciones SUBTEL	Dirección de TI institucional; gobierno digital e interoperabilidad; conocimiento directo del dominio del caso de validación.	> 10 años
E2	Angel Merino	Jefe de la Oficina de Arquitectura y Gestión de Procesos, Dirección de Presupuestos (DIPRES).	Arquitectura y operación de plataformas transversales del Estado; gestión por procesos; sistemas de gestión financiera pública.	25 años
E3	Gonzalo Luengo Castro	Arquitecto Empresarial TOGAF 9 Certified; cofundador de Andes Digital. Consultor y arquitecto de soluciones en proyectos de integración de información del sector público.	Arquitectura empresarial e integración e interoperabilidad entre organismos públicos.	> 10 años

Tabla A.1. Composición del panel de expertos y perspectiva aportada por cada evaluador.

La heterogeneidad de perfiles constituye un control metodológico deliberado: el acuerdo entre evaluadores con marcos de referencia distintos es evidencia más robusta que el acuerdo entre pares del mismo perfil.

A.3 Método de agregación y criterio de decisión

El instrumento emplea dos escalas Likert de cinco puntos: valoración técnica para los 49 ítems de evaluación y escala de acuerdo para los 10 ítems de contraste de hipótesis. Las definiciones operacionales, fijadas antes de la aplicación, fueron: la concordancia por ítem (I-CVI), como proporción de evaluadores que responden 4 o 5 sobre el total de respuestas válidas; el índice de validez de contenido de la escala (S-CVI/Ave), como promedio de los I-CVI de cada agrupación; y el reporte de media, mediana y rango por ítem, conforme a la práctica recomendada para escalas ordinales con paneles reducidos.

Criterio de decisión. Tratándose de un panel reducido ($n = 3$), se aplica el criterio más exigente de Lynn (1986), que para paneles de hasta cinco expertos requiere un I-CVI = 1,00 por ítem, es decir,

acuerdo unánime. El umbral referencial de $\geq 0,80$ se reserva para paneles de seis o más expertos y se reporta únicamente como referencia comparativa, en coherencia con lo establecido en la Sección 3.4.

A.4 Resultados agregados de la evaluación del FAR

Los tres evaluadores respondieron la totalidad de los 49 ítems, sin respuestas omitidas ni «N/A», lo que arroja 147 respuestas válidas: 116 de valor 5 (78,9%) y 31 de valor 4 (21,1%), sin ninguna respuesta de valor 1, 2 o 3. En consecuencia, el I-CVI alcanzó 1,00 en los 49 ítems y el S-CVI/Ave es igual a 1,00. La media global fue de 4,79 sobre 5 (desviación estándar 0,41; mediana 5).

Dim.	Dimensión evaluada	Ítems	E1	E2	E3	Media	Mna.	I-CVI
D1	Pertinencia y relevancia del problema abordado	5	4.60	4.80	4.80	4.73	5,0	1,00
D2	Diseño y coherencia de la arquitectura (5 capas)	8	4.75	4.88	4.88	4.83	5,0	1,00
D3	Mapeo normativo (Compliance by Design)	6	4.67	4.67	4.83	4.72	5,0	1,00
D4	Modelo de seguridad Zero Trust	5	4.60	4.60	4.80	4.67	5,0	1,00
D5	Privacidad por diseño (Privacy by Design)	5	5.00	4.80	4.80	4.87	5,0	1,00
D6	Blueprint y replicabilidad	7	4.71	4.86	4.86	4.81	5,0	1,00
D7	Metodología de validación (DSR)	7	4.71	4.71	4.86	4.76	5,0	1,00
D8	Contribución académica e impacto práctico	6	4.83	4.83	5.00	4.89	5,0	1,00
—	Total Sección 2	49	4.73	4.78	4.86	4.79	5,0	1,00

Tabla A.2. Resultados agregados por dimensión (escala 1–5). Mna. = mediana del conjunto de respuestas de la dimensión. I-CVI = valor mínimo observado entre los ítems de la dimensión.

Las dimensiones mejor valoradas fueron la contribución académica e impacto práctico (4,89) y la privacidad por diseño (4,87), seguidas del diseño y coherencia de la arquitectura (4,83) y del blueprint y replicabilidad (4,81). Las de menor valoración relativa fueron el modelo de seguridad Zero Trust (4,67) y el mapeo normativo (4,72), ambas dentro del tramo «Bueno-Excelente» y con observaciones de complemento y no de corrección. Las medias individuales fueron 4,74 (E1), 4,78 (E2) y 4,86 (E3), con un rango de apenas 0,12 puntos, lo que evidencia alta consistencia; ningún ítem registró divergencia superior a un punto de la escala, descartándose la necesidad de una segunda ronda.

A.5 Resultados del contraste de hipótesis

A.5.1 Hipótesis principal (H1)

Los cinco ítems alcanzan un I-CVI de 1,00, con S-CVI/Ave de 1,00 (concordancia del 100%) y una media de 4,73. El ítem de síntesis H1.5 obtuvo la calificación máxima de forma unánime, al igual que H1.1 y H1.3, confirmando que el panel respalda tanto el mecanismo —el cumplimiento por diseño produce una conformidad sustantivamente superior a la de una arquitectura legada— como la razonabilidad de los umbrales fijados. El ítem de menor valoración fue H1.2 (validez del ICN y su rúbrica), con media 4,00 y acuerdo unánime en el valor 4: los tres evaluadores consideran el instrumento válido y suficiente, pero identifican de forma coincidente una oportunidad de robustecimiento consistente en ponderar los requisitos por criticidad y documentar la evidencia de cada valoración. H1.4 (ausencia de degradación del rendimiento) obtuvo 4,67, con la única reserva de E1 por permanecer pendiente la comparación definitiva contra el proceso heredado en condiciones productivas.

Ítem	Afirmación evaluada	E1	E2	E3	Media	Mna.	Rango	I-CVI
H1.1	Mayor cumplimiento que una arquitectura tradicional (legacy)	5	5	5	5.00	5	5–5	1.00
H1.2	Validez y suficiencia del ICN y de su rúbrica de valoración	4	4	4	4.00	4	4–4	1.00

H1.3	Meta de +30 p.p. e ICN $\geq$ 90% alcanzable y razonable	5	5	5	5.00	5	5-5	1.00
H1.4	Ausencia de degradación del rendimiento operativo	4	5	5	4.67	5	4-5	1.00
H1.5	La evidencia respalda la aceptación de H1	5	5	5	5.00	5	5-5	1.00
—	Promedio H1	—	—	—	4.73	5	4-5	1,00

Tabla A.3. Resultados por ítem de la hipótesis principal (H1). Escala de acuerdo 1–5.

Conclusión sobre H1. Al satisfacerse el criterio de Lynn (I-CVI = 1,00 en todos los ítems), la hipótesis principal se considera validada por el panel. El resultado converge con la verificación de cumplimiento de la Actividad 5, donde el ICN se eleva del 18% al 92% (+74 p.p.), lo que refuerza la validez del constructo al provenir de dos fuentes independientes.

A.5.2 Hipótesis secundaria (H2)

Los cinco ítems alcanzan un I-CVI de 1,00, con S-CVI/Ave de 1,00 y una media de 4,80, superior a la de H1. Obtuvieron unanimidad en la calificación máxima los ítems H2.1 (replicabilidad en instituciones de distinto tamaño y dominio), H2.3 (suficiencia de la tabla de parámetros para replicar mediante configuración) y H2.5 (juicio integrado). Es metodológicamente relevante que la valoración más alta corresponda al mecanismo que sustenta la replicabilidad: la separación entre una estructura invariante y un conjunto acotado de parámetros configurables. Los ítems con dispersión fueron H2.2 y H2.4 (media 4,67): en H2.2, los tres evaluadores consideran alcanzable el umbral de $\leq 30\%$ y dos de ellos lo estiman probablemente conservador, dado que la variabilidad se concentra en la capa de negocio; en H2.4, E1 y E2 sostienen que la instanciación única acredita factibilidad pero no cuantifica el umbral, mientras que E3 estima satisfechas las condiciones de validación propias de una arquitectura de referencia.

Ítem	Afirmación evaluada	E1	E2	E3	Media	Mna.	Rango	I-CVI
H2.1	El FAR es replicable en instituciones de distinto tamaño y dominio	5	5	5	5.00	5	5-5	1.00
H2.2	La adaptación no superaría el 30% de los componentes originales	4	5	5	4.67	5	4-5	1.00
H2.3	La tabla de parámetros permite replicar sin rediseño arquitectónico	5	5	5	5.00	5	5-5	1.00
H2.4	El caso BRSALUD es evidencia de factibilidad suficiente	4	4	5	4.33	4	4-5	1.00
H2.5	La evidencia respalda la aceptación de H2	5	5	5	5.00	5	5-5	1.00
—	Promedio H2	—	—	—	4.80	5	4-5	1,00

Tabla A.4. Resultados por ítem de la hipótesis secundaria (H2). Escala de acuerdo 1–5.

Conclusión sobre H2. La hipótesis secundaria se considera validada por el panel en sus dimensiones de diseño y factibilidad. El panel es unánime, no obstante, en que la confirmación cuantitativa del umbral de adaptación requiere una segunda instanciación en un dominio distinto, coincidiendo con la limitación declarada en las Secciones 5.7 y 7.

A.6 Síntesis de la validación de hipótesis

Ambas hipótesis satisfacen el criterio de decisión aplicable a paneles reducidos. La totalidad de los 59 ítems cerrados del instrumento —49 de evaluación y 10 de contraste de hipótesis— obtuvo un I-CVI de 1,00, sobre un total de 177 respuestas válidas, de las cuales 139 (78,5%) corresponden al valor máximo de la escala. No se registró ninguna respuesta en el tramo desfavorable (valores 1 a 3), resultado de concordancia excepcionalmente alto que descarta la necesidad de una segunda ronda de consulta.

Hipótesis	Nº de ítems	Media	Concordancia media (%)	S-CVI/Ave	Criterio aplicado	¿Validada?
Hipótesis principal (H1)	5	4.73	100,0%	1,00	Lynn (1986): I-CVI = 1,00	Sí
Hipótesis secundaria (H2)	5	4.80	100,0%	1,00	Lynn (1986): I-CVI = 1,00	Sí

Tabla A.5. Resumen de la validación de hipótesis por el panel de expertos (n = 3). Corresponde a la tabla de resultados prevista en el Anexo A del instrumento de evaluación.

A.7 Síntesis cualitativa de las observaciones del panel

A.7.1 Fortalezas convergentes

Los tres evaluadores coinciden, de forma independiente, en tres fortalezas principales. Primera, la integración normativa como decisión arquitectónica: el mérito no reside en los componentes individuales —Data Lake, Zero Trust y Privacy by Design son marcos previamente documentados— sino en su composición coherente y en la trazabilidad explícita a las tres leyes mediante la matriz de mapeo. Segunda, la separación entre estructura invariante y parámetros configurables, identificada como la propiedad que determina la replicabilidad efectiva y que distingue una arquitectura de referencia de una arquitectura de solución. Tercera, la validación sobre un proceso real, crítico y con datos sensibles: el panel destaca expresamente que la transparencia con que se reportan los resultados adversos (cobertura de 28 de ~50 fuentes y reconciliación del 73,3% con su causa identificada) refuerza y no debilita la validez del estudio. Adicionalmente, los tres evaluadores destacan el módulo de ejercicio de derechos ARCO+ de la Capa 5, por convertir una obligación legal en un componente arquitectónico verificable.

A.7.2 Observaciones y áreas de mejora convergentes

Las observaciones del panel se concentran en seis áreas, ninguna de las cuales fue calificada como impedimento para la presentación del trabajo; todas se formulan como líneas de evolución del artefacto. Los ejes principales son el robustecimiento metodológico del Índice de Cumplimiento Normativo —mediante ponderación por criticidad y documentación de la evidencia—, la incorporación de la dimensión de gobierno de la arquitectura y de los prerequisites organizacionales de la adopción, y el anclaje del marco en los habilitadores de gobierno digital ya operativos en el Estado.

Nº	Observación convergente	Planteada por	Ámbito	Tratamiento propuesto
O1	Robustecer el ICN mediante ponderación de los requisitos por criticidad y documentación de la evidencia que sustenta cada valoración de la rúbrica, de modo que el índice sea reproducible por un evaluador independiente. Se sugiere además reportar el ICN desagregado por capa además de por ley.	E1, E2, E3	Metodológico	Incorporar a la Sección 3.4 como refinamiento del instrumento y a la Sección 7 como trabajo futuro.
O2	Confirmar la referenciación de artículos de la matriz normativa contra los textos oficiales (BCN) y precisar la articulación de la Ley N°21.180 con la Ley N°19.880.	E1, E2, E3	Precisión normativa	Ya declarado como nota metodológica en la Sección 3.2; se mantiene como cierre pendiente.
O3	Vincular explícitamente la dimensión de interoperabilidad con la plataforma PISEE y la normativa técnica de la Secretaría de Gobierno Digital, apoyándose en los habilitadores estatales vigentes (Clave Única, PISEE, FirmaGob) en lugar de proponer una vía paralela.	E1, E2, E3	Arquitectura / adopción	Incorporar a las Secciones 3.2 y 5.4; es la observación con mayor potencial de elevar el ICN de la Ley N°21.180.
O4	Incorporar los prerequisites organizacionales y el gobierno de la arquitectura: designación efectiva del DPO, instancia de revisión de arquitectura con criterios de conformidad, y custodia institucional y ciclo de vida del blueprint.	E1, E2, E3	Gobernanza	Incorporar como nueva subsección del Capítulo 5 y como recomendación de adopción en la Sección 7.

O5	Agregar un modelo de costos y de esfuerzo de adopción (TCO a tres años, costo a escala productiva y horas-persona por capa), necesario para sustentar el caso de negocio institucional.	E2, E3	Económico	Incorporar a la Sección 7 como trabajo futuro; complementa el comparativo de la Sección 6.
O6	Explicitar la continuidad operacional con objetivos de RTO y RPO, estrategia de sitio alternativo y pruebas de restauración, dada la obligación de entrega en fecha fija del proceso.	E1, E2	Operacional	Incorporar a la Sección 5.6 y al checklist de adopción de la Sección 5.6.

Tabla A.6. Observaciones convergentes del panel y tratamiento propuesto en el cuerpo de la tesina.

A.7.3 Aportes específicos por perfil

E1 (dirección de TI institucional) recomendó la recertificación periódica de accesos y la separación de funciones entre quien define y quien aplica la política de acceso —señalando que la degradación del mínimo privilegio ocurre por acumulación de permisos en el tiempo—, así como documentar las excepciones de retención legal aplicables al derecho de supresión en el ámbito previsional. E2 (plataformas transversales del Estado) advirtió sobre el riesgo de reidentificación asociado al uso del RUN como identificador de vinculación transversal, recomendando documentar la custodia de la tabla de correspondencia y evaluar el riesgo en los conjuntos expuestos en la zona Gold; propuso además una política de retención e inmutabilidad de los registros de auditoría segregada de las cuentas administradoras, y un mecanismo de adopción compartida para instituciones de menor tamaño. E3 (arquitectura empresarial e integración) planteó formalizar contratos de datos versionados por fuente —esquema, frecuencia, criterios de calidad, tratamiento de rechazos y responsable de origen—, publicar el modelo ArchiMate en formato abierto e importable, y definir un modelo de conformidad verificable que permita a una institución declarar y demostrar su nivel de adopción.

A.7.4 Juicio global e idoneidad académica

Evaluador	Calificación global del FAR	Juicio sobre idoneidad académica
E1 — Paul Cook	5 — Excelente	Requiere modificaciones menores (correcciones puntuales, precisiones y complementos).
E2 — Angel Merino	5 — Excelente	Adecuado para ser presentado sin modificaciones.
E3 — Gonzalo Luengo	5 — Excelente	Adecuado para ser presentado sin modificaciones.

Tabla A.7. Calificación global y juicio de idoneidad académica por evaluador (Sección 5 del instrumento).

Los tres evaluadores asignaron la calificación global máxima al Framework. En cuanto a la idoneidad académica, dos lo consideran adecuado para su presentación sin modificaciones y uno estima necesarias modificaciones menores, referidas a las precisiones ya identificadas y no a aspectos de fondo. El panel es unánime en recomendar la aprobación del trabajo y coincide en recomendar la difusión del blueprint y de la matriz de mapeo normativo más allá del ámbito académico, poniéndolos a disposición de la Secretaría de Gobierno Digital y de la red de arquitectos y jefaturas de TI del sector público.

A.8 Incorporación de los resultados al cuerpo de la tesina

Los resultados del juicio experto se tradujeron en modificaciones concretas al documento, de modo que la validación externa retroalimente efectivamente el trabajo y no quede confinada al anexo. Las seis observaciones convergentes del panel se encuentran incorporadas en las siguientes ubicaciones: la ponderación del Índice de Cumplimiento Normativo por criticidad y la documentación de la evidencia por requisito (O1), en la Sección 3.4 bajo el epígrafe «Refinamiento del instrumento» y entre las limitaciones de la Sección 7; la vinculación de la interoperabilidad con los habilitadores de gobierno digital del Estado —PISEE, Clave Única y FirmaGob— (O3), en la Sección 3.2 y como parámetro de adopción en la Sección 5.4; los prerrequisitos organizacionales y el gobierno de la arquitectura (O4), en la nueva Sección 5.8; y la continuidad operacional con objetivos de tiempo y

punto de recuperación (O6), en el checklist de adopción de la Sección 5.6. Las observaciones restantes se recogen en los apartados correspondientes de las Secciones 5 y 7.

Asimismo, los resultados agregados del panel se incorporan por referencia en la Sección 4.2 (Actividad 5 del DSRM), en la Sección 4.3 relativa a la metodología de trabajo e instrumentos, en la Sección 5.7 al contrastar la validación del blueprint con el respaldo del panel a la hipótesis secundaria, y en las conclusiones y limitaciones de la Sección 7.

A.9 Limitaciones del juicio experto y consideraciones de validez

En coherencia con el estándar de transparencia aplicado al resto del trabajo, se declaran las siguientes limitaciones del proceso de validación por juicio experto. En cuanto al tamaño del panel, este se compone de tres especialistas, en el límite inferior del rango de 3 a 4 previsto en la Sección 4.3; un panel reducido presenta menor estabilidad estadística, razón por la cual se adoptó el criterio más exigente de Lynn (1986) —unanimidad por ítem— en lugar del umbral referencial de $\geq 0,80$, criterio que la totalidad de los ítems satisfizo.

Respecto de la composición, la Sección 4.3 contemplaba la participación de especialistas provenientes de organismos reguladores y de la academia, perfiles que el panel efectivamente constituido no incorpora, al privilegiar la perspectiva de arquitectura e implementación en el sector público. La incorporación de un evaluador con perfil regulador —por ejemplo, de la Agencia de Protección de Datos Personales o de la Agencia Nacional de Ciberseguridad— fortalecería particularmente la evaluación del mapeo normativo y de la privacidad por diseño, y se propone como línea de continuidad.

En relación con la deseabilidad social y el sesgo de benevolencia, toda evaluación no anónima solicitada por el autor está expuesta a un sesgo favorable. Este se mitigó mediante tres controles: la exigencia de fundamentar las calificaciones extremas, la inclusión de preguntas abiertas dirigidas explícitamente a identificar debilidades, y la selección de evaluadores con posición profesional consolidada e independiente. La presencia de calificaciones de valor 4 acompañadas de observaciones específicas, así como la formulación de seis áreas de mejora convergentes, indica que el panel ejerció un juicio discriminante y no meramente aprobatorio.

En cuanto a los vínculos declarados, dos evaluadores manifestaron circunstancias que se dejan a consideración del comité académico: E1 se desempeñó como Director de Tecnología del Instituto de Previsión Social, institución cuyo proceso BRSALUD constituye el caso de validación, lo que informa técnicamente su evaluación; y E3 es cofundador de Andes Digital, consultora citada como fuente documental del trabajo por su participación en la definición del alcance del proyecto asociado al caso de estudio. Ambos declararon emitir su juicio a título personal y sustentado en el contenido del documento evaluado; E2 no declaró vínculo alguno. Finalmente, el panel evaluó el artefacto según su documentación y las métricas reportadas, sin acceso directo al entorno de la prueba de concepto, por lo que la validez de las conclusiones sobre viabilidad técnica descansa en la evidencia documentada en las Secciones 4 y 5.

Ninguna de estas limitaciones altera el resultado de la validación, que satisface el criterio de decisión establecido a priori. Se consignan para permitir al lector ponderar el alcance de la evidencia y para orientar la continuidad del proceso de validación.

A.10 Conclusión del anexo

La aplicación del instrumento de juicio experto a un panel de tres especialistas del sector público chileno, con perfiles complementarios en dirección de tecnología institucional, arquitectura de plataformas transversales del Estado y arquitectura empresarial e interoperabilidad, arrojó una valoración uniformemente favorable del Framework de Arquitectura de Referencia. La media global de la evaluación fue de 4,79 sobre 5, sin registrarse ninguna calificación en el tramo desfavorable de la escala, y los tres evaluadores asignaron la calificación global máxima al artefacto.

Ambas hipótesis de la investigación alcanzaron un índice de validez de contenido S-CVI/Ave de 1,00, satisfaciendo el criterio de Lynn (1986) aplicable a paneles reducidos y quedando, en consecuencia, validadas por el panel. Este resultado converge con las dos restantes fuentes de evidencia de la estrategia de triangulación —las métricas de la prueba de concepto y la verificación de cumplimiento mediante la matriz normativa—, lo que fortalece la validez de constructo del estudio y mitiga el sesgo asociado a una fuente única de validación.

Las observaciones del panel, sistematizadas en la Sección A.7 y traducidas en modificaciones concretas al cuerpo del documento según se detalla en la Sección A.8, se orientan de forma convergente hacia tres ejes: el robustecimiento metodológico del Índice de Cumplimiento Normativo, la incorporación de la dimensión de gobierno y de los prerequisites organizacionales de la adopción, y el anclaje del marco en los habilitadores de gobierno digital ya operativos en el Estado. Ninguna de ellas cuestiona el diseño del artefacto ni las conclusiones del trabajo; todas constituyen líneas de evolución que, incorporadas, incrementan la probabilidad de adopción efectiva del FAR por parte de los organismos públicos.

Referencia metodológica: Lynn, M. R. (1986). Determination and quantification of content validity. Nursing Research, 35(6), 382–385.