

# Data Mesh: Delivering Data-Driven Value at Scale

Resumen del capítulo 1

Zhamak Dehghani

November 22, 2025

## Data Mesh: Delivering Data-Driven Value at Scale

### Prefacio

El Data Mesh es el empujón que nos coloca en una nueva trayectoria sobre cómo abordamos los datos: cómo los imaginamos, capturamos, compartimos y creamos valor a partir de ellos a escala. Esta nueva trayectoria nos aleja de la **centralización** de los datos y su propiedad, moviéndonos hacia un modelo **descentralizado**.

Este enfoque abraza la complejidad de las organizaciones modernas y busca obtener valor de los datos a escala, a pesar del desorden y la complejidad organizacional. Históricamente, la industria ha tenido estos cambios (como el nacimiento de Unix, la arquitectura distribuida o los microservicios). Data Mesh busca establecer las condiciones para abordar la complejidad en el corazón de los datos, específicamente en analítica e IA.

La tesis se formuló en 2018 tras observar patrones de fallo comunes en empresas que, a pesar de grandes inversiones, luchaban por escalar sus soluciones de gestión de datos.

---

### Prólogo: Imagina el Data Mesh

La autora utiliza una empresa ficticia, **Daff, Inc.** (una compañía global de streaming de música similar a Spotify), para ilustrar qué es el Data Mesh en la práctica.

#### Data Mesh en Acción (El escenario futuro - 2022)

Daff ha pivotado hacia el Data Mesh. La empresa se organiza en unidades de negocio llamadas **dominios** (ej. *player*, *partnership*, *playlist*, *artist*).

- **Cultura:** Existe una cultura ubicua de curiosidad y experimentación de datos.

- **Organización:** Cada dominio combina desarrollo de software y capacidades de negocio, y es responsable de los componentes de software que soportan ese dominio.
- **Colaboración:** Los equipos experimentan continuamente. Por ejemplo, el equipo de *playlist* colabora directamente con el dominio de *partnership* para integrar datos de plataformas de ejercicio y crear listas de reproducción para correr o andar en bicicleta.

#### **El proceso de creación de valor:**

El equipo de *partnership* crea un nuevo “producto de datos” (*partner playlists*) en cuestión de horas. Utilizan herramientas de gestión del ciclo de vida del producto de datos de la plataforma.

1. Transforman los datos.
2. Armonizan IDs (ej. Track ID).
3. Despliegan el producto de datos a la malla (*mesh*).
4. Lo hacen disponible para que el equipo de *playlist* lo consuma.

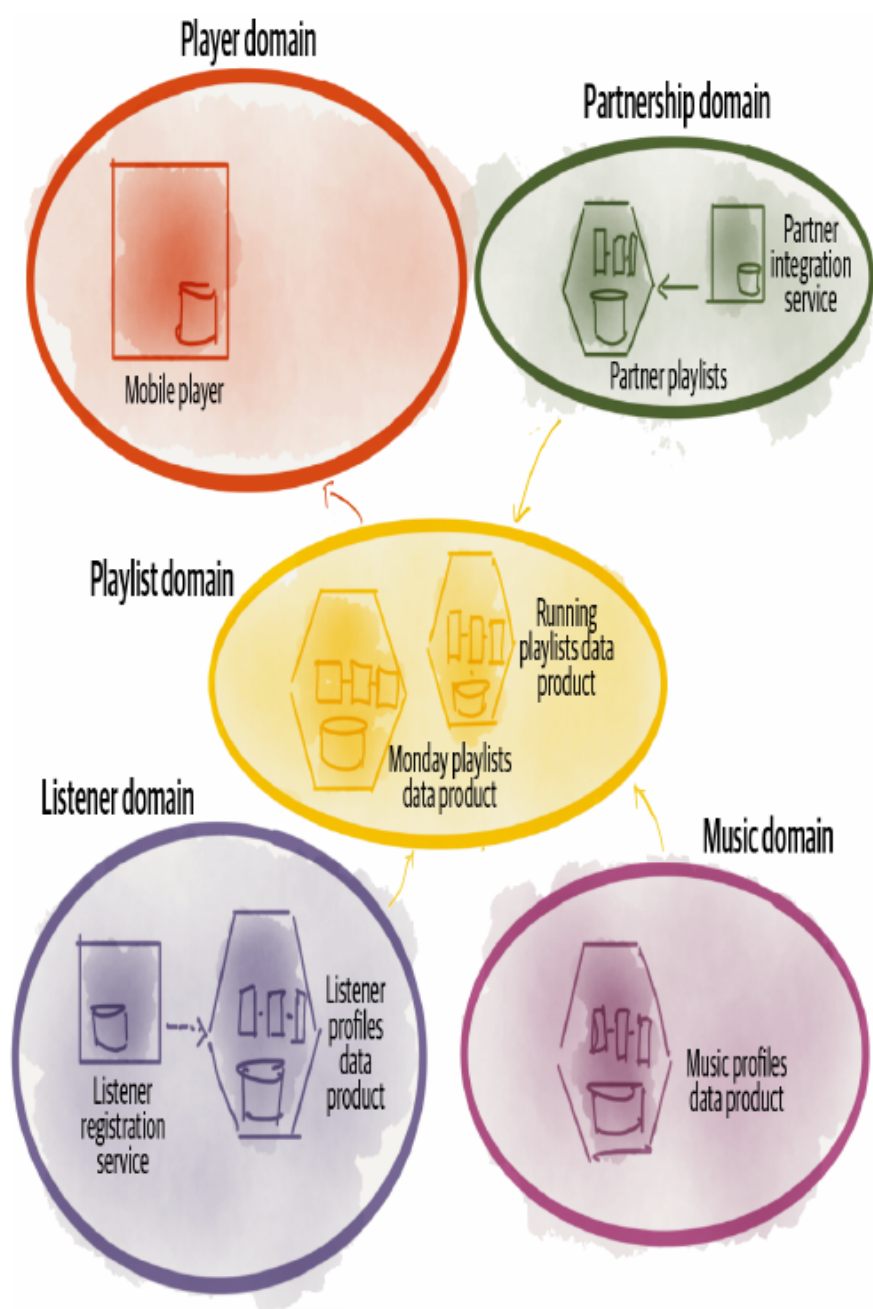


Figura P-1: Escenario de creación de listas de reproducción inteligentes con data mesh

### Principios en juego en este escenario:

1. **Propiedad de dominio descentralizada:** El equipo de *partnership* es dueño de sus datos, eliminando la brecha entre quienes crean el dato y quienes lo usan.
2. **Datos como producto:** Los equipos tienen la responsabilidad de proveer datos que sean descubribles, comprensibles y usables.
3. **Plataforma de datos de autoservicio:** Una plataforma “invisible” que automatiza la infraestructura, permitiendo a los equipos compartir datos sin fricción (automatizando acceso, encriptación, registro).
4. **Gobernanza computacional federada:** Políticas globales (como quién es dueño de qué) definidas por un grupo federado y automatizadas por la plataforma.

### El trabajo con datos antes del Data Mesh (El escenario pasado - 2019)

Tres años antes, la situación era la norma de la industria pero ineficiente:

- **Cuellos de botella:** Un equipo centralizado de datos e IA tenía que priorizar solicitudes de toda la organización.
- **Fricción y Lentitud:** Para crear una playlist de deportes, los científicos de datos tenían que pedir acceso a un equipo de gobernanza central, esperar días, y luego pedir a ingenieros de datos centrales que construyeran pipelines ETL.
- **Desconexión:** Los ingenieros centrales no entendían el dominio de *partnership*, lo que llevaba a errores, datos obsoletos y falta de confianza.
- **Arquitectura Monolítica:** Un Data Lake o Warehouse central que se convertía en un cuello de botella ante la proliferación de fuentes de datos.

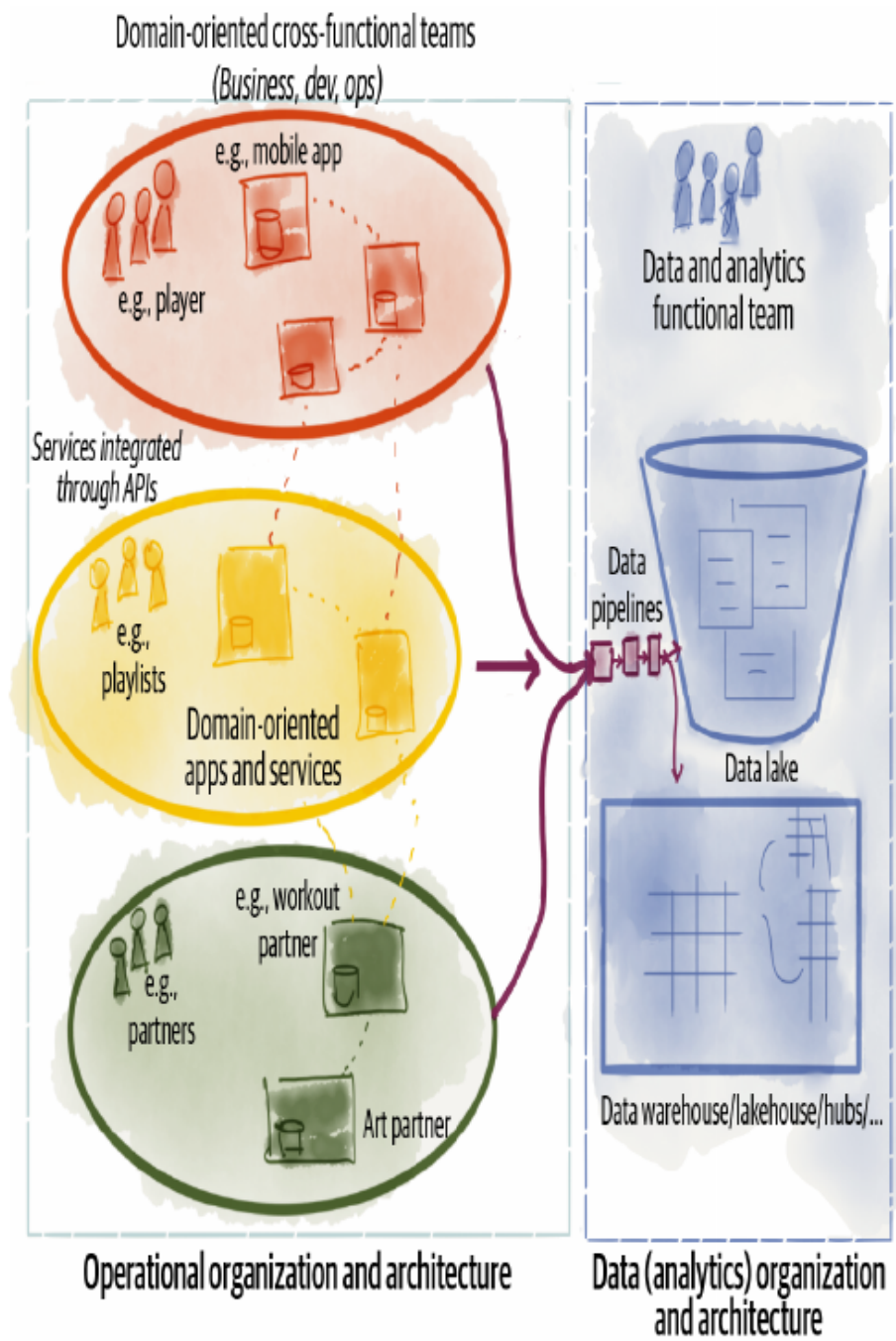


Figura P-2: Organización y arquitectura de Daff antes de data mesh

## El Efecto de Red Positivo

El éxito de Daff radica en el efecto de red creado por la conectividad par-a-par (*peer-to-peer*) de los dominios intercambiando productos de datos. Cuantos más productos de datos se conectan, más inteligencia e *insights* de alto orden se generan.

---

## Parte I. ¿Qué es Data Mesh?

*“La única simplicidad en la que se puede confiar es la simplicidad que se encuentra al otro lado de la complejidad.” — Alfred North Whitehead*

Para entender Data Mesh, debemos analizarlo desde sus **primeros principios**. Data Mesh se clasifica como un **paradigma sociotécnico**: un enfoque que reconoce las interacciones entre las personas y la arquitectura técnica en organizaciones complejas.

---

## Capítulo 1. Data Mesh en Resumen (Data Mesh in a Nutshell)

**Definición:** Data Mesh es un enfoque sociotécnico descentralizado para compartir, acceder y gestionar datos analíticos en entornos complejos y de gran escala (dentro o entre organizaciones).

### Los Resultados (The Outcomes)

Data Mesh busca lograr tres objetivos principales para obtener valor a escala:

1. **Responder con gracia al cambio:** Manejar la volatilidad, incertidumbre y complejidad esencial del negocio.
2. **Sostener la agilidad ante el crecimiento:** Evitar que la organización se ralentice a medida que añade más fuentes y consumidores de datos.
3. **Aumentar la relación de valor de los datos respecto a la inversión.**

### Los Cambios (The Shifts)

Data Mesh introduce cambios multidimensionales respecto a los enfoques anteriores de gestión de datos (como Data Warehouse y Data Lake).

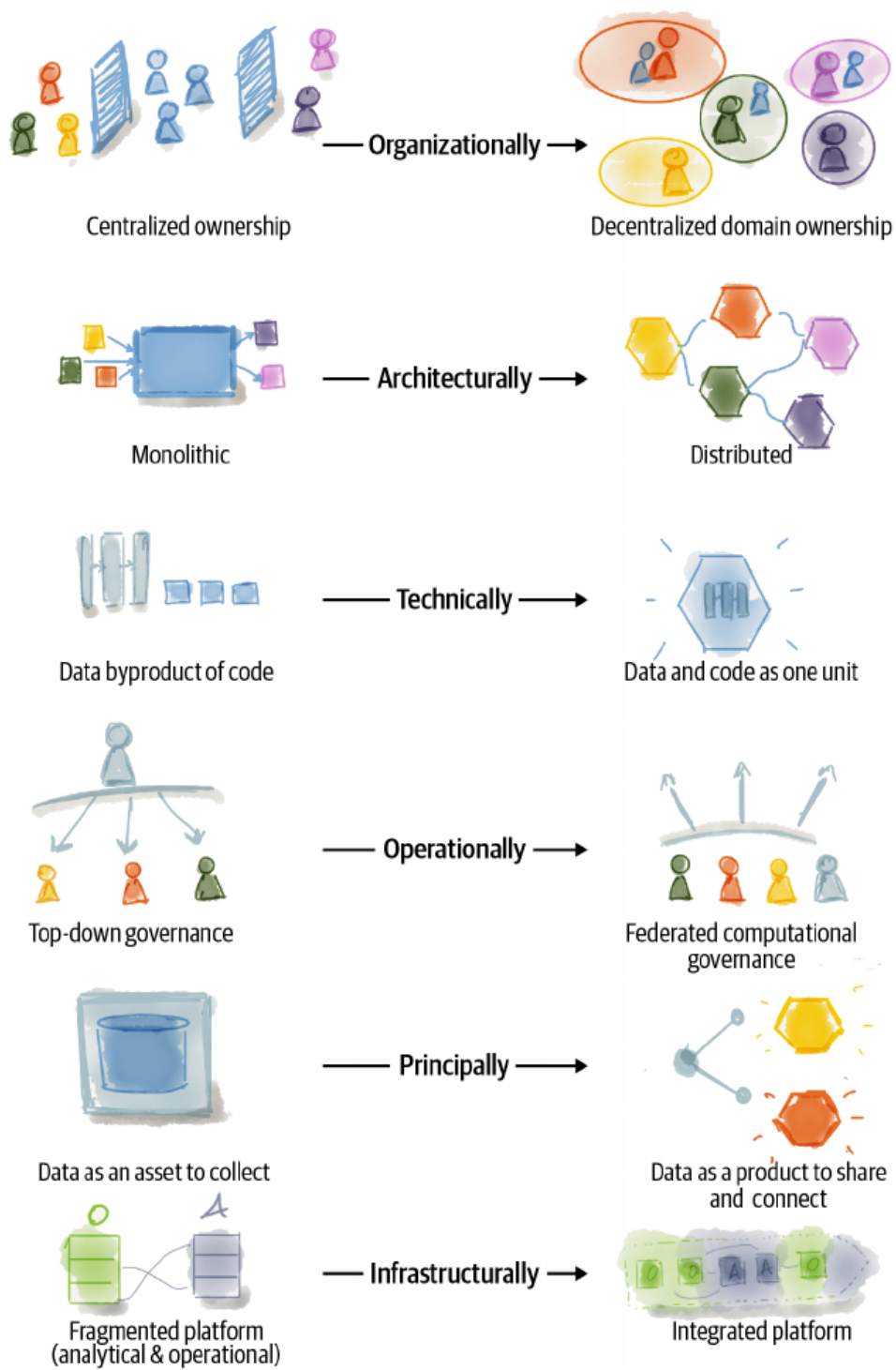


Figura 1-1: Dimensiones de cambio del data mesh

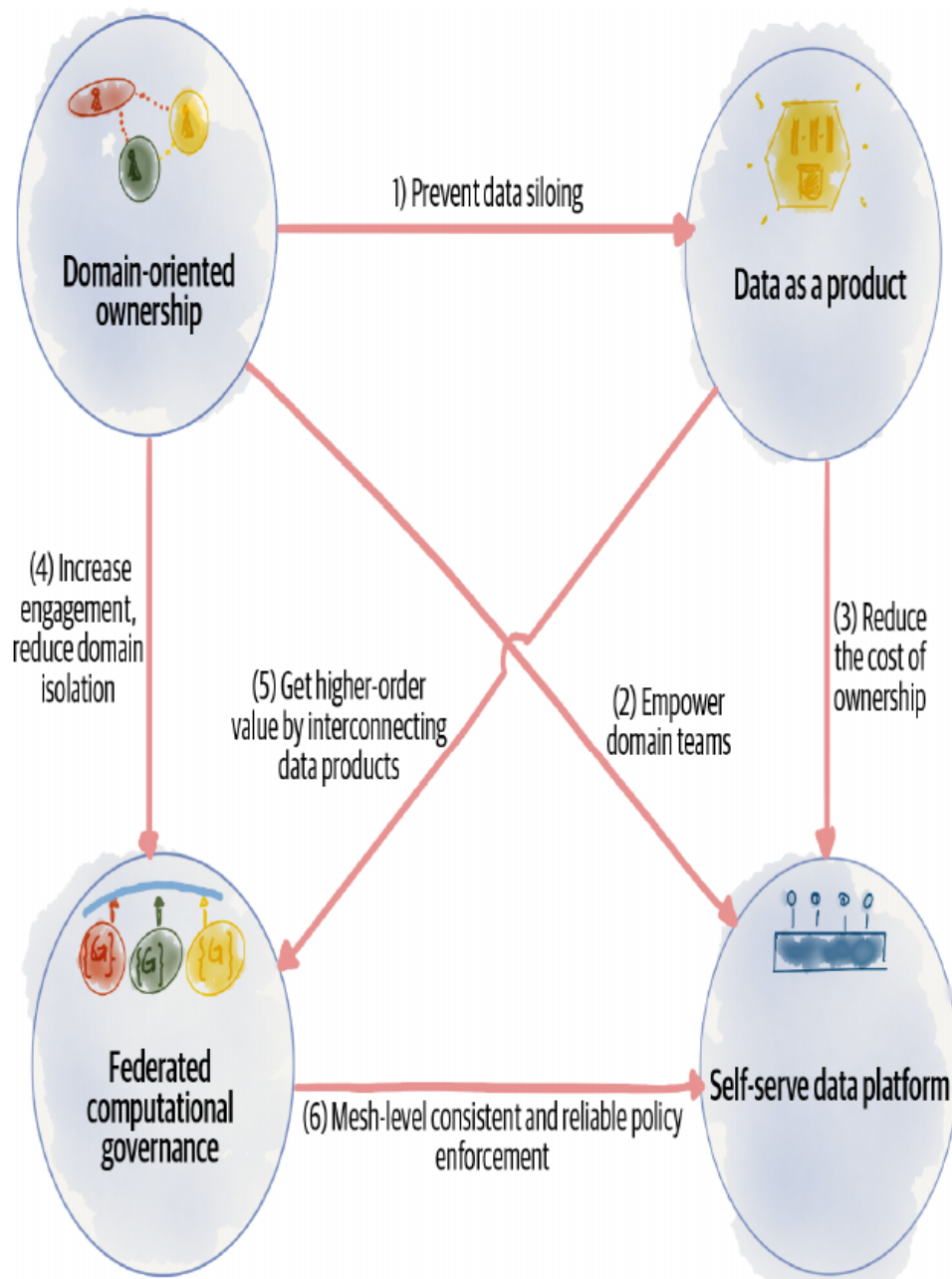
La siguiente tabla resume estos cambios fundamentales explicados en la Figura 1-1:

Dimensión	Enfoque Anterior (Centralizado)	Enfoque Data Mesh (Descentralizado)
<b>Organizacional</b>	Propiedad centralizada por especialistas de la plataforma de datos.	Propiedad descentralizada del dominio; la responsabilidad vuelve a los dominios de negocio.
<b>Arquitectural</b>	Monolítico (Warehouses/Lakes). Recolección de datos.	Distribuido. Conexión de datos a través de una malla distribuida de productos de datos.
<b>Tecnológico</b>	Datos como subproducto del código de pipeline.	Datos y código que los mantiene como una unidad autónoma y viva ( <i>Data Quantum</i> ).
<b>Operacional</b>	Gobernanza “top-down” (de arriba abajo) con intervención humana.	Gobernanza computacional federada con políticas embebidas en los nodos de la malla.
<b>De Principios</b>	Datos como un “activo” para recolectar.	Datos como un <b>producto</b> para servir y deleitar a los usuarios.
<b>Infraestructural</b>	Servicios de infraestructura fragmentados y aislados.	Conjunto bien integrado de infraestructura para sistemas operacionales y de datos.

### Los Principios (The Principles)

Hay cuatro principios simples que sustentan la arquitectura lógica y el modelo operativo del Data Mesh. Se complementan entre sí.





*\* Direction of the arrow shows the dependency from one principle to another; implementing the from principle creates the challenge that the to principle addresses.*

Figura 1-2: Cuatro principios de data mesh y su interacción

**1. Principio de Propiedad del Dominio (Domain Ownership)** Descentralizar la propiedad de los datos analíticos a los dominios de negocio que están más cerca de los datos (ya sea la fuente o los consumidores principales).

- **Objetivo:** Escalar la compartición de datos alineada con el crecimiento organizacional, optimizar para el cambio continuo y aumentar la veracidad de los datos.

**2. Principio de Datos como Producto (Data as a Product)** Para evitar que la descentralización cree silos de datos, los datos de dominio deben compartirse como un producto.

- **Características de usabilidad:** Los datos deben ser descubribles, direccionables, comprensibles, confiables, accesibles nativamente, interoperables, valiosos por sí mismos y seguros.
- **Data Quantum:** Introduce una nueva unidad de arquitectura lógica que encapsula datos, metadatos, código y políticas de forma autónoma.

**3. Principio de la Plataforma de Datos de Autoservicio (Self-Serve Data Platform)** Para evitar la duplicación de esfuerzos y la alta carga cognitiva en los equipos de dominio, se necesita una nueva generación de plataformas de datos.

- **Objetivo:** Abstraer la complejidad de la gestión de datos, reducir el costo total de propiedad y movilizar a una población más grande de desarrolladores generalistas para construir productos de datos.

**4. Principio de Gobernanza Computacional Federada (Federated Computational Governance)** Para asegurar la interoperabilidad y seguridad en un modelo descentralizado.

- **Modelo:** Un equipo federado (representantes de dominio + plataforma + expertos) decide las políticas.
- **Ejecución:** Las políticas se codifican y automatizan (computacionalmente) en cada producto de datos a través de la plataforma, en lugar de depender de intervenciones manuales.

### **Modelo de Data Mesh de un Vistazo**

Operacionalmente, los dominios con equipos multifuncionales logran los objetivos de negocio mediante aplicaciones y productos de datos. La plataforma de autoservicio automatiza la infraestructura y la aplicación de políticas globales.

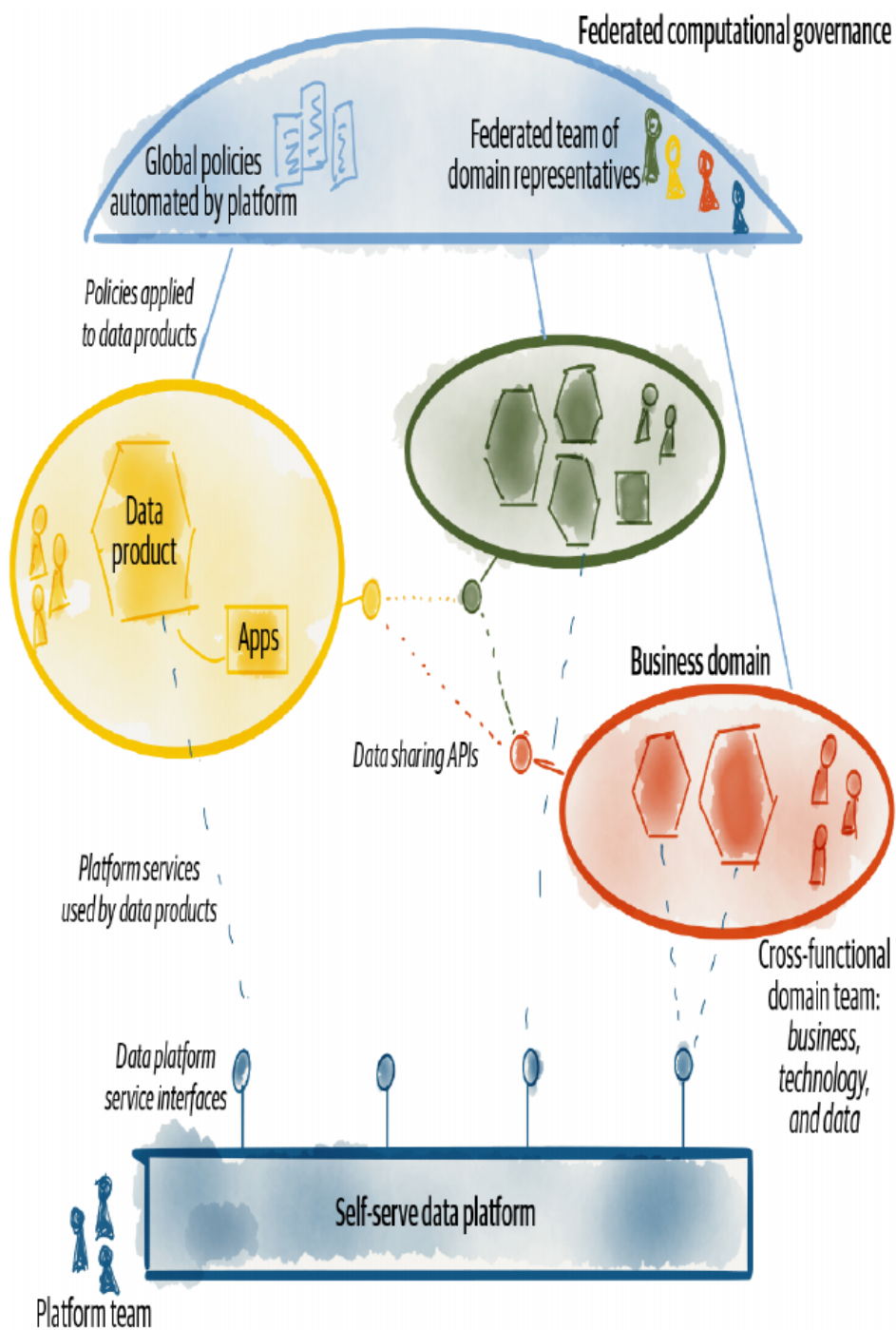


Figura 1-3: Modelo operativo de los principios de data mesh

## Los Datos (The Data)

Data Mesh se enfoca en **datos analíticos**, pero reconoce la necesidad de integrar los dos modos de datos:

### 1. Datos Operacionales (“Data on the inside”):

- Soportan la ejecución del negocio en tiempo real.
- Transaccionales (OLTP).
- Mantienen el estado actual.
- Optimizados para la lógica de la aplicación (microservicios).

### 2. Datos Analíticos (“Data on the outside”):

- Visión histórica, integrada y agregada de los datos.
- Mantenidos en sistemas OLAP (Warehouses, Lakes).
- Optimizados para lógica analítica (entrenar modelos ML, reportes).
- Usados para **optimizar** el negocio y la experiencia del usuario.
- Son inmutables y variantes en el tiempo.

## El Origen

La autora hace referencia a **Thomas Kuhn** (*La Estructura de las Revoluciones Científicas*) para explicar que Data Mesh surge de una crisis de paradigma. El modelo anterior (Warehouses/Lakes centralizados) presentaba “anomalías” y fallos al intentar escalar en empresas modernas digitalizadas. Data Mesh no es una invención aislada, sino una adaptación de prácticas exitosas de la ingeniería de software (Microservicios, Diseño Guiado por el Dominio, DevOps) aplicadas al problema de los datos analíticos.