Data Mesh

Introduction

INNOQ

JOCHEN CHRIST
@JOCHEN_CHRIST

Hi,
I am Jochen

Jochen Christ

Senior Consultant at INNOQ

- 💪 *Java*
- 👍 *Remote Mob Programming*
- 🤩 *Data-driven Development*

O'REILLY®

Deutsche Ausgabe

Data Mesh

Eine dezentrale Datenarchitektur entwerfen

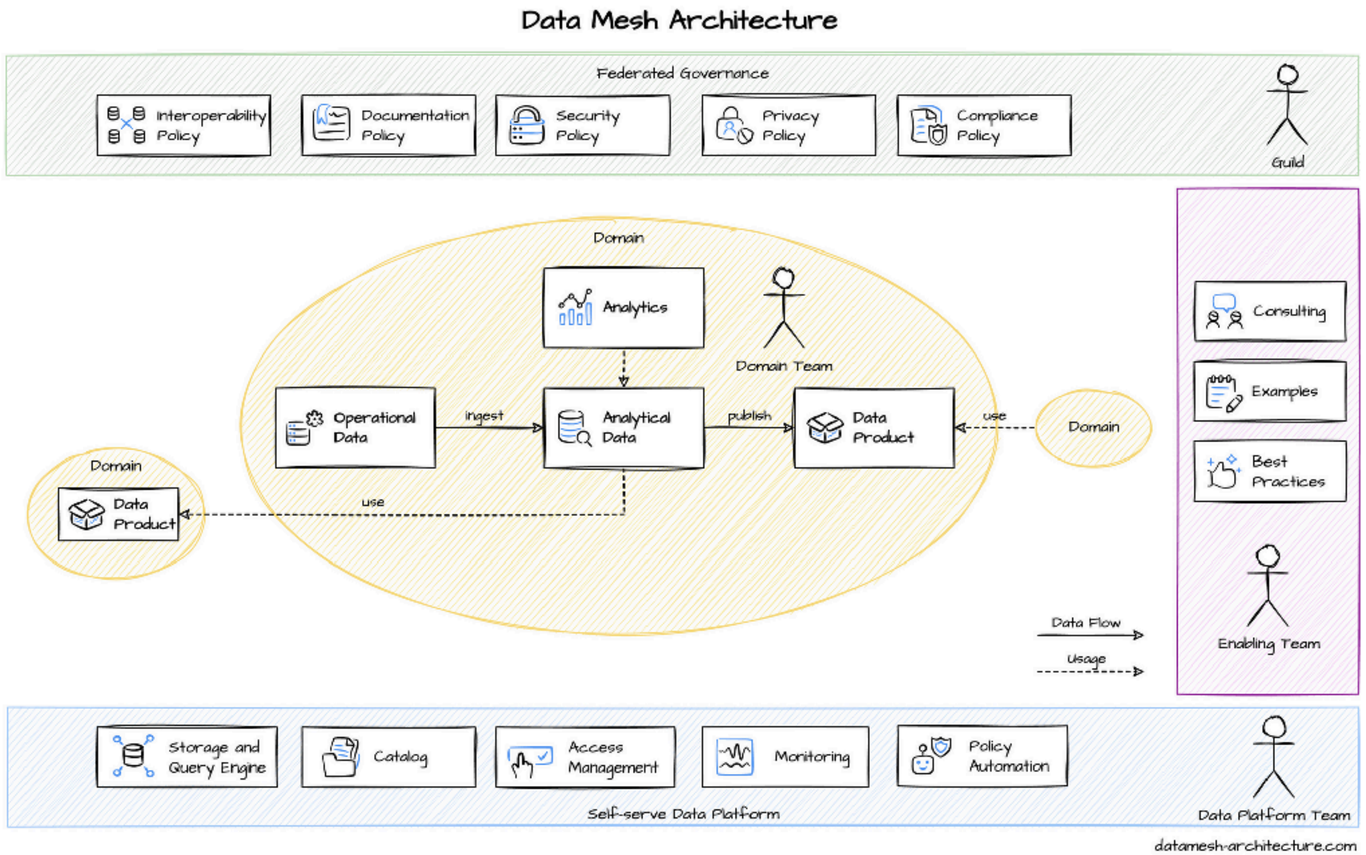
Zhamak Dehghani
Vorwort von Martin Fowler
Übersetzung von Jochen Christ und Simon Harrer

Data Mesh Architecture

https://www.datamesh-architecture.com

How To Design a Data Mesh?

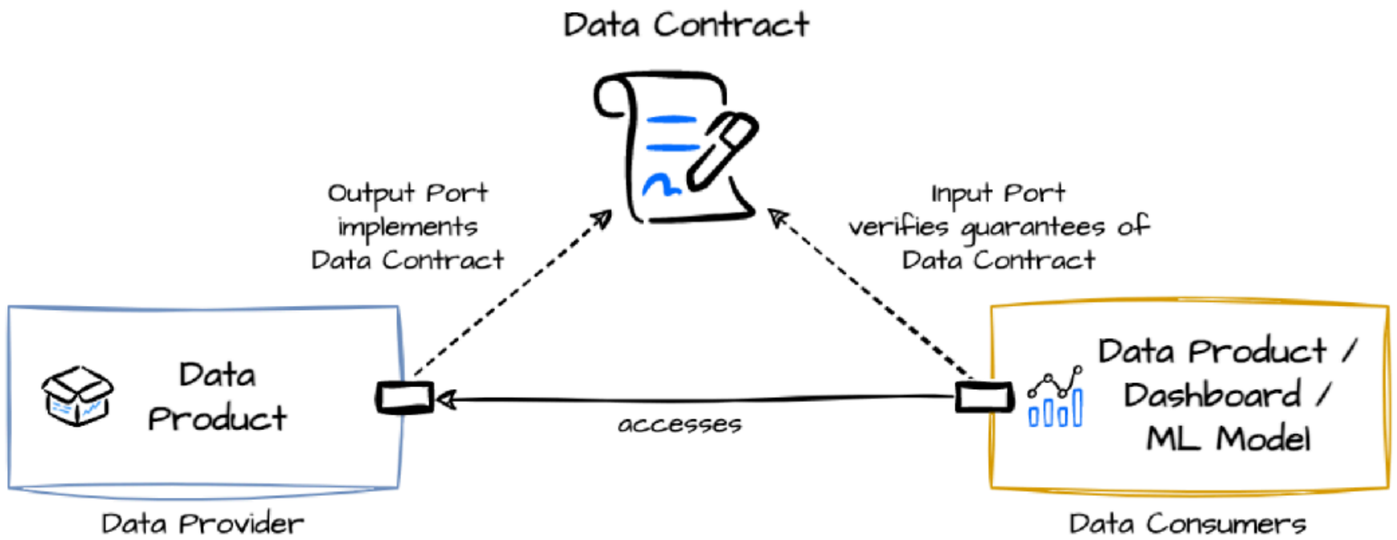
A data mesh architecture is a decentralized approach that enables domain teams to perform cross-domain data analysis on their own. At its core is the domain with its responsible team and its operational and analytical data. The domain team ingests operational data and builds analytical data models to perform their own analysis. It uses analytical data to build data products based on other domains' needs.



datamesh-architecture.com

datacontract.com

Data Contract Specification

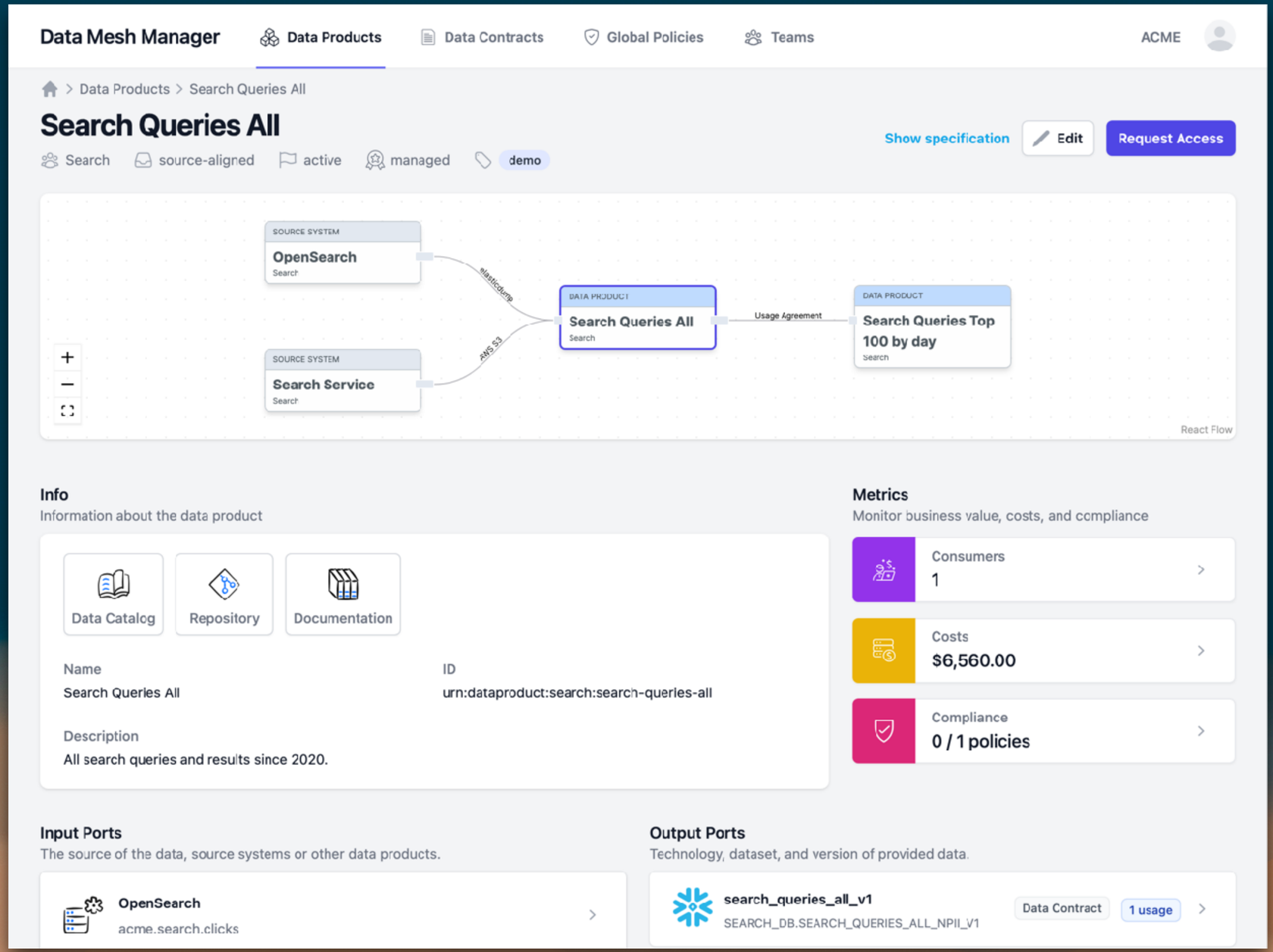


Data contracts bring data providers and data consumers together.

A *data contract* is a document that defines the structure, format, semantics, quality, and terms of use for exchanging data between a data provider and their consumers. A data contract is implemented by a data product's output port or other data technologies. Data contracts can also be used for the input port to specify the expectations of data dependencies and verify given guarantees.

The *data contract specification* defines a YAML format to describe attributes of provided data sets. It is data platform neutral, yet supports well-known formats to express schemas (e.g., dbt models, JSON Schema, Protobuf, SQL DDL) and quality tests (e.g., SodaCL, SQL queries) to avoid unnecessary abstractions. The data contract specification is an open initiative to define a common data contract format. Think of an OpenAPI specification, but for data sets.

datacontract.com



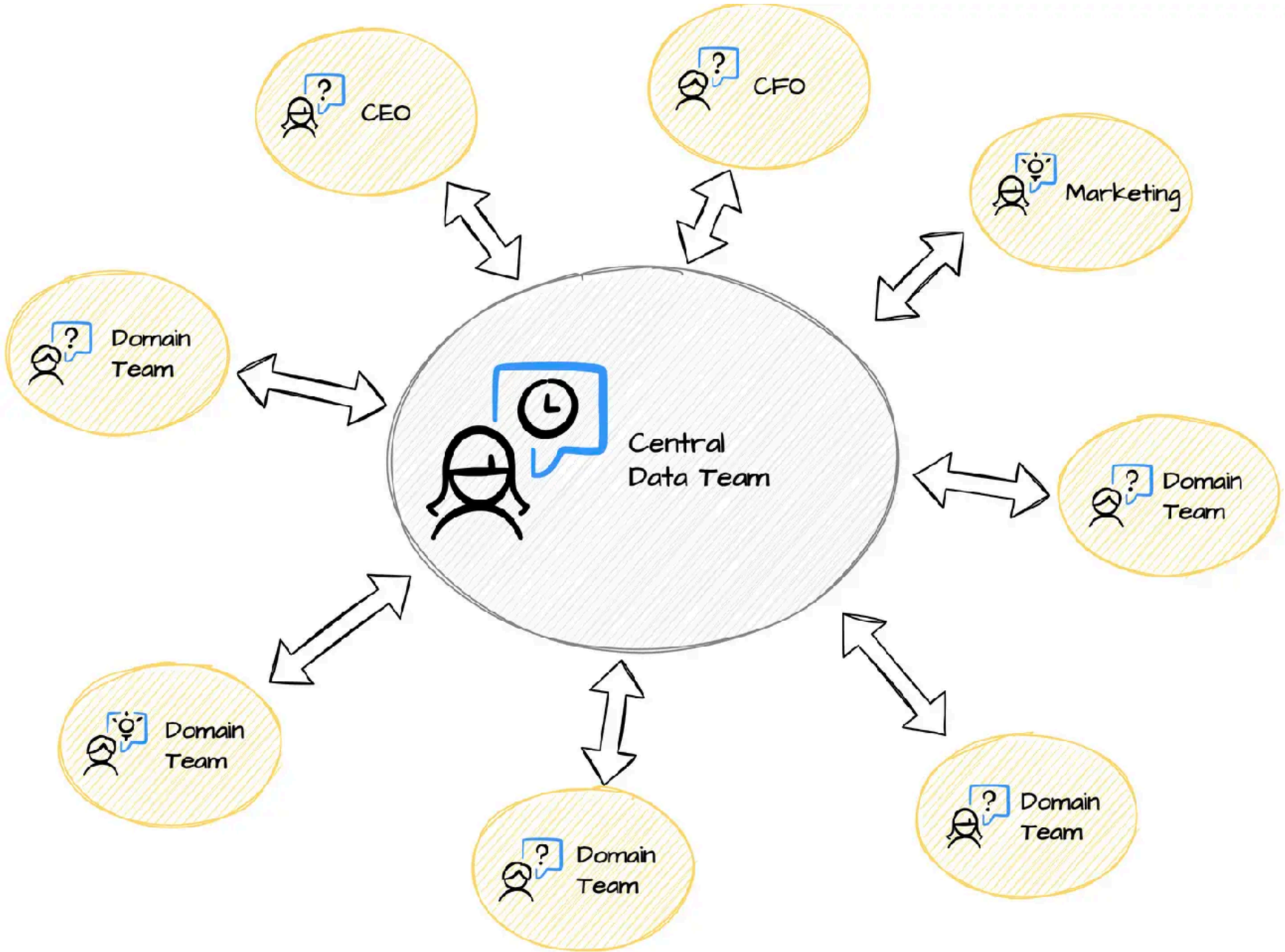
datamesh-manager.com

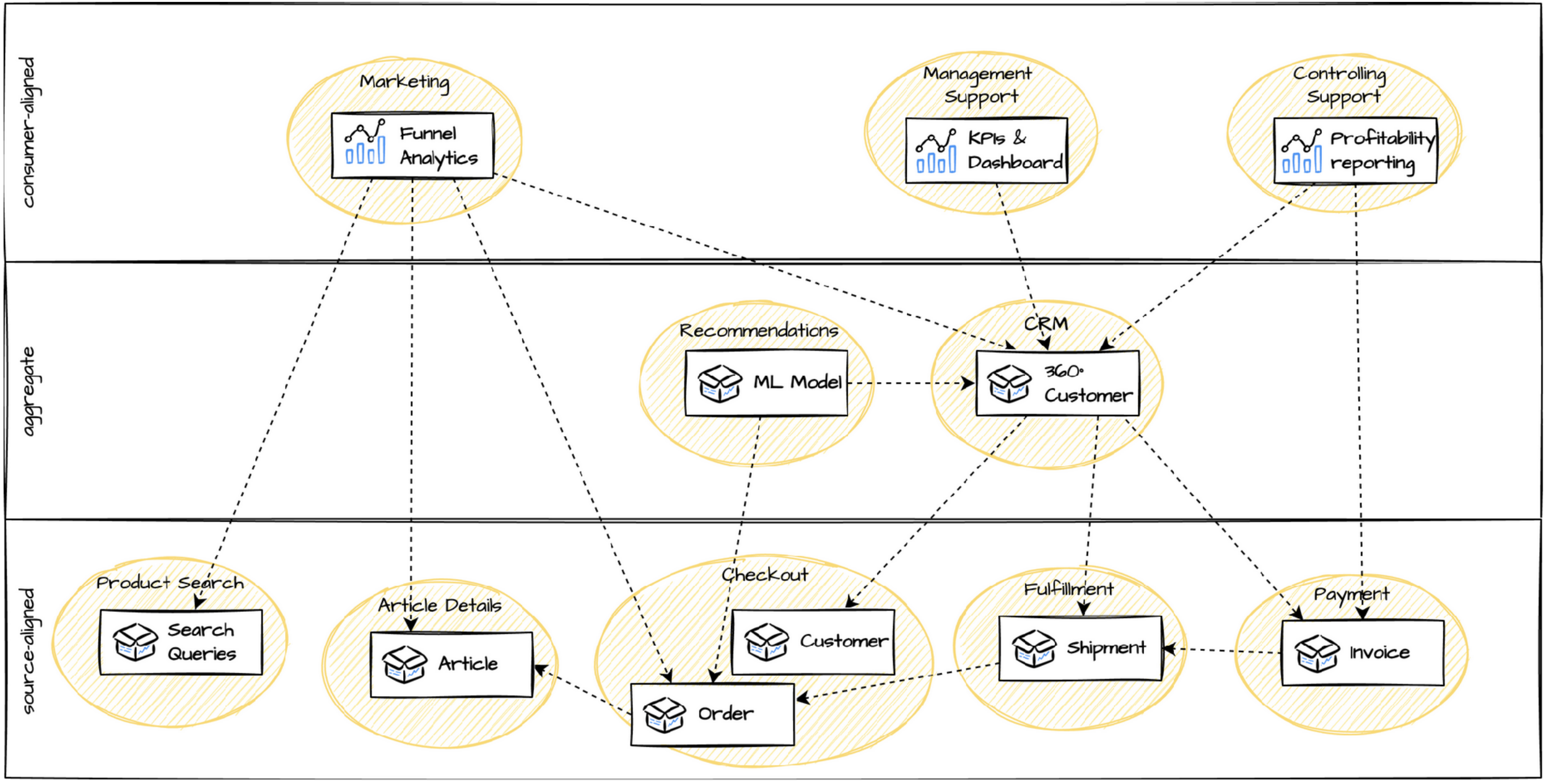
Decentralized Data Architecture

What?

A decentralized data architecture gives
ownership and competence
for (analytical) data
to the teams that
understand the business context.

-- Jochen


CEO
CFO
Marketing
Domain Team
Central Data Team
Domain Team
Domain Team
Domain Team
Domain Team
Domain Team


consumer-aligned
aggregate
source-aligned
Marketing
Funnel Analytics
Management Support
KPIs & Dashboard
Controlling Support
Profitability reporting
Recommendations
ML Model
CRM
360° Customer
Product Search
Search Queries
Article Details
Article
Checkout
Customer
Order
Fulfillment
Shipment
Payment
Invoice

Decentralized Data Architecture

Why?

Make qualified data-driven decisions
in your domain

Use data to better understand your users and system behavior. Derive features from insights, qualify value, and fast iterations. Also qualified rejection of unnecessary tasks.

Do the right things, purpose, motivation

Build innovative services
in your domain

Enhance your customer experience with data technologies, such as LLMs, visualizations, classifications, and ML models for predictions and recommendations.

Customer value through innovation

Provide data as business value
for other domains

Domain data is valuable for other business units as reference data and to aggregate. Needs managed, explained, high-quality and easy accessible data as products.

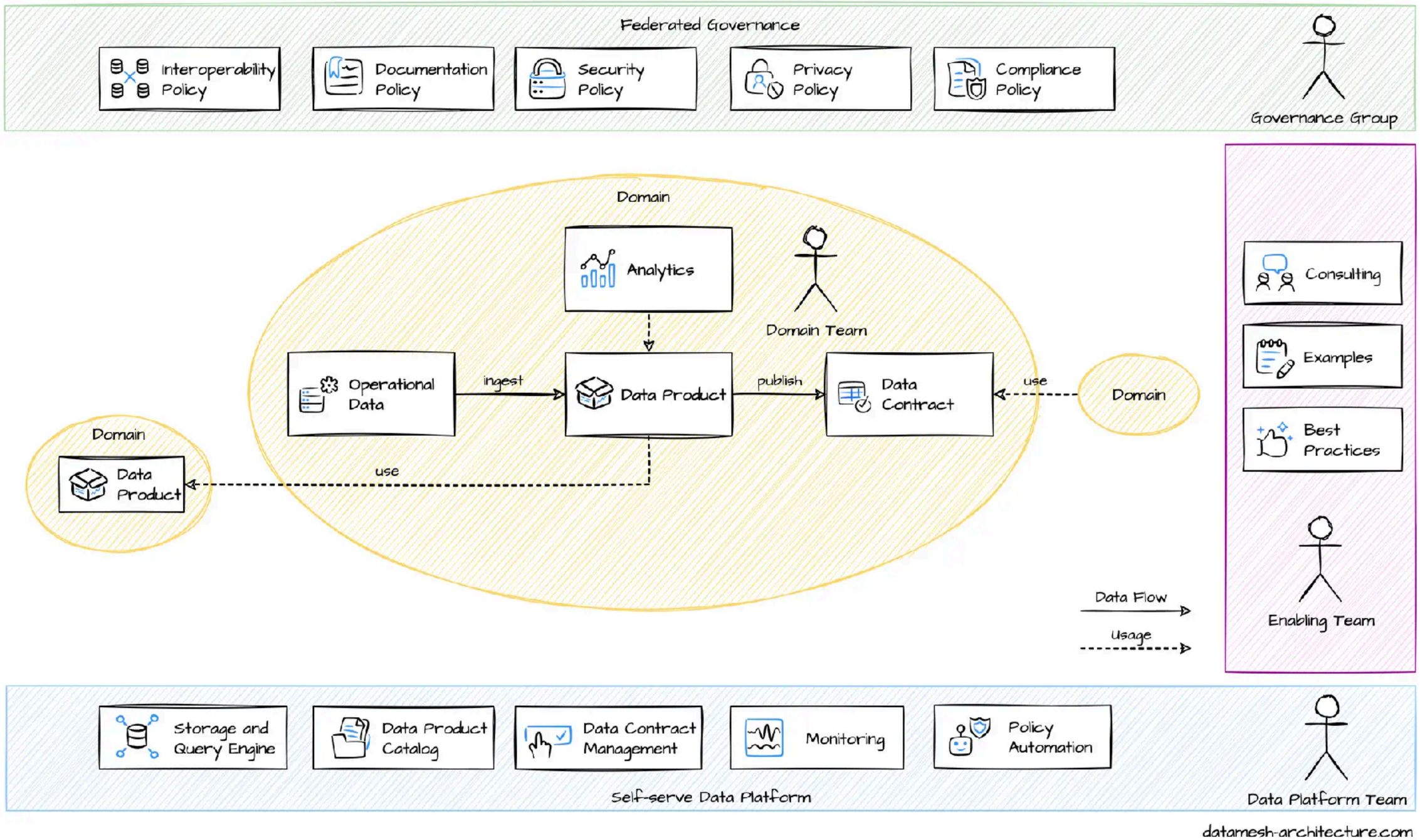
Company success

What Is Data Mesh?

	Domain Ownership	Data as a Product	Self-serve Data Platform	Federated Governance
Strategic Domain-driven Design	Domain Bounded Context	Product Thinking	Domain-agnostic	Context Mapping
Socio-technical Perspective	Domain Teams	Data Product by Domain Team	Data Platform Team	Guild
Technology	Operational & Analytical Data	Interoperability Interfaces	Self-serve Data Platform	Data Governance & Automation

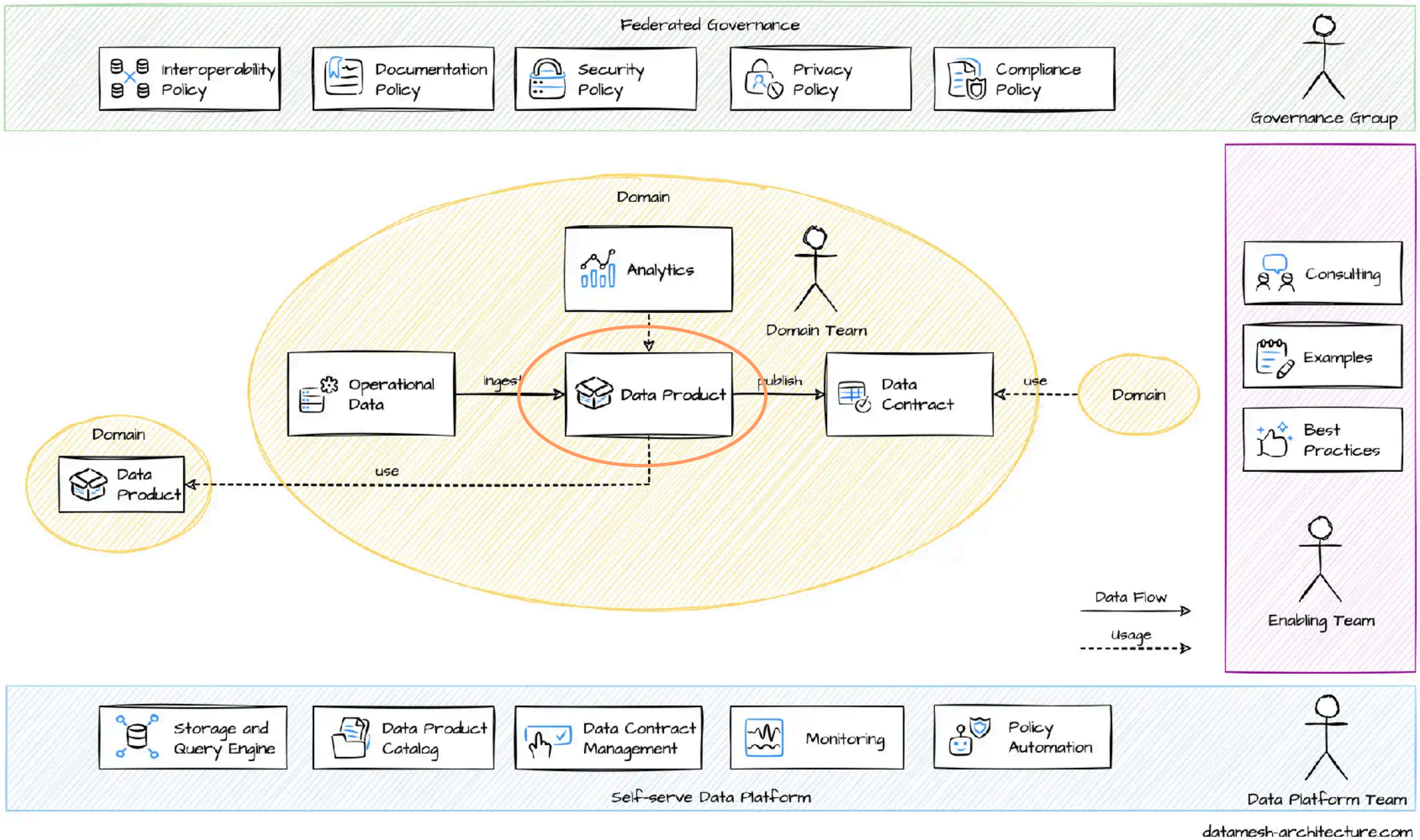
datamesh-architecture.com

Data Mesh Architecture



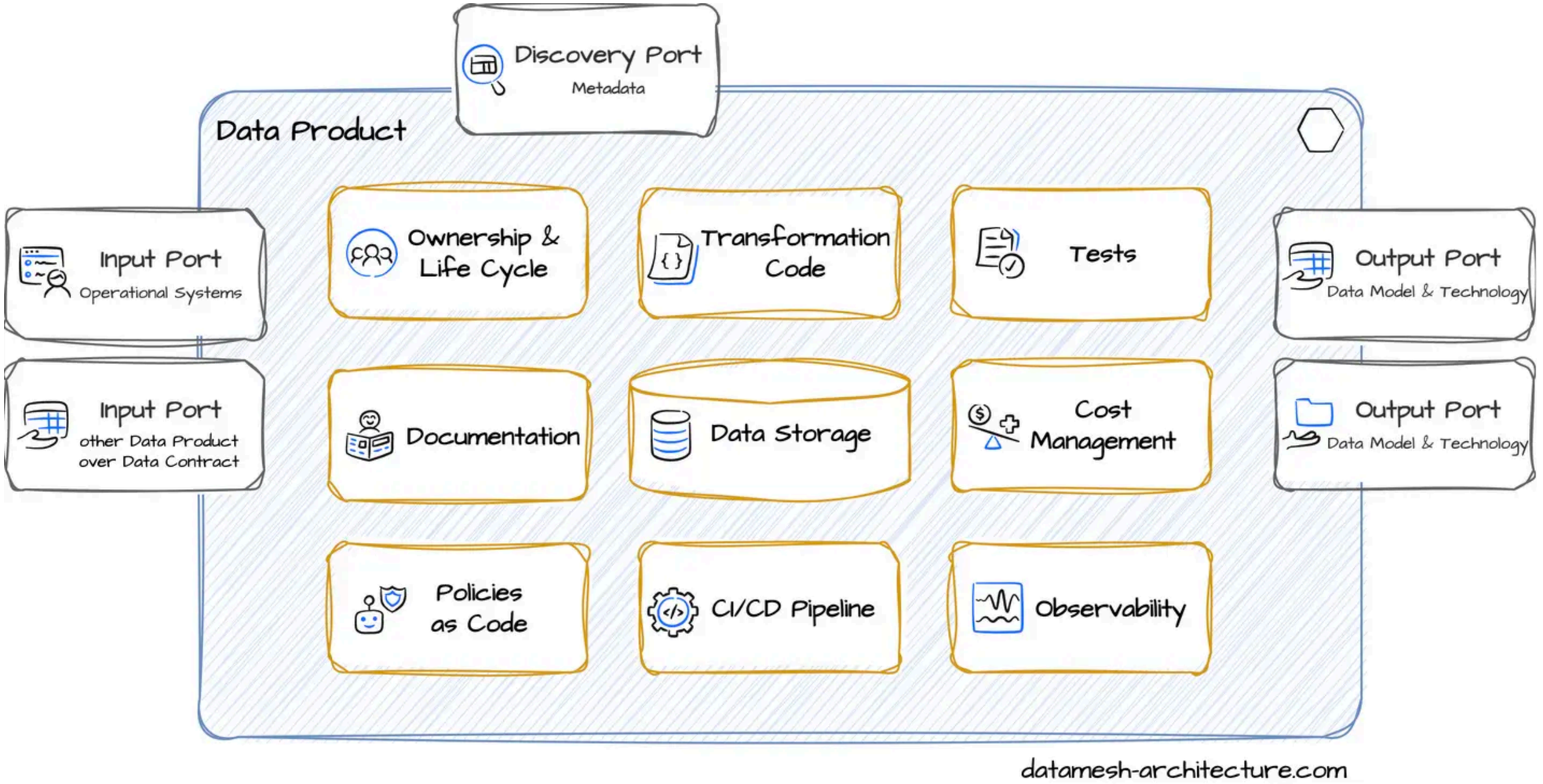
datamesh-architecture.com

Data Mesh Architecture



datamesh-architecture.com

Data Product are Modules



Output Port Example

Row	sku	location	available	updated_at
1	9520010951145	20	0	2021-02-28 12:29:21 UTC
2	9520010951145	20	1	2021-03-02 09:07:21 UTC
3	9520010951145	20	0	2021-03-03 16:36:21 UTC
4	9520010951145	20	1	2021-03-04 13:03:21 UTC
5	9520010951145	20	2	2021-03-05 17:26:21 UTC
6	9520010951145	20	3	2021-03-06 03:35:21 UTC
7	9520010951145	20	2	2021-03-06 17:25:21 UTC
8	9520010951145	20	1	2021-03-07 18:10:21 UTC

- Technical Endpoint
- Hides implementation details
- Large data set
- Read-only
- Technology
  - Tables
  - Files in Bucket
  - Topic
- Data Model
  - With PII
  - Without PII
- Version

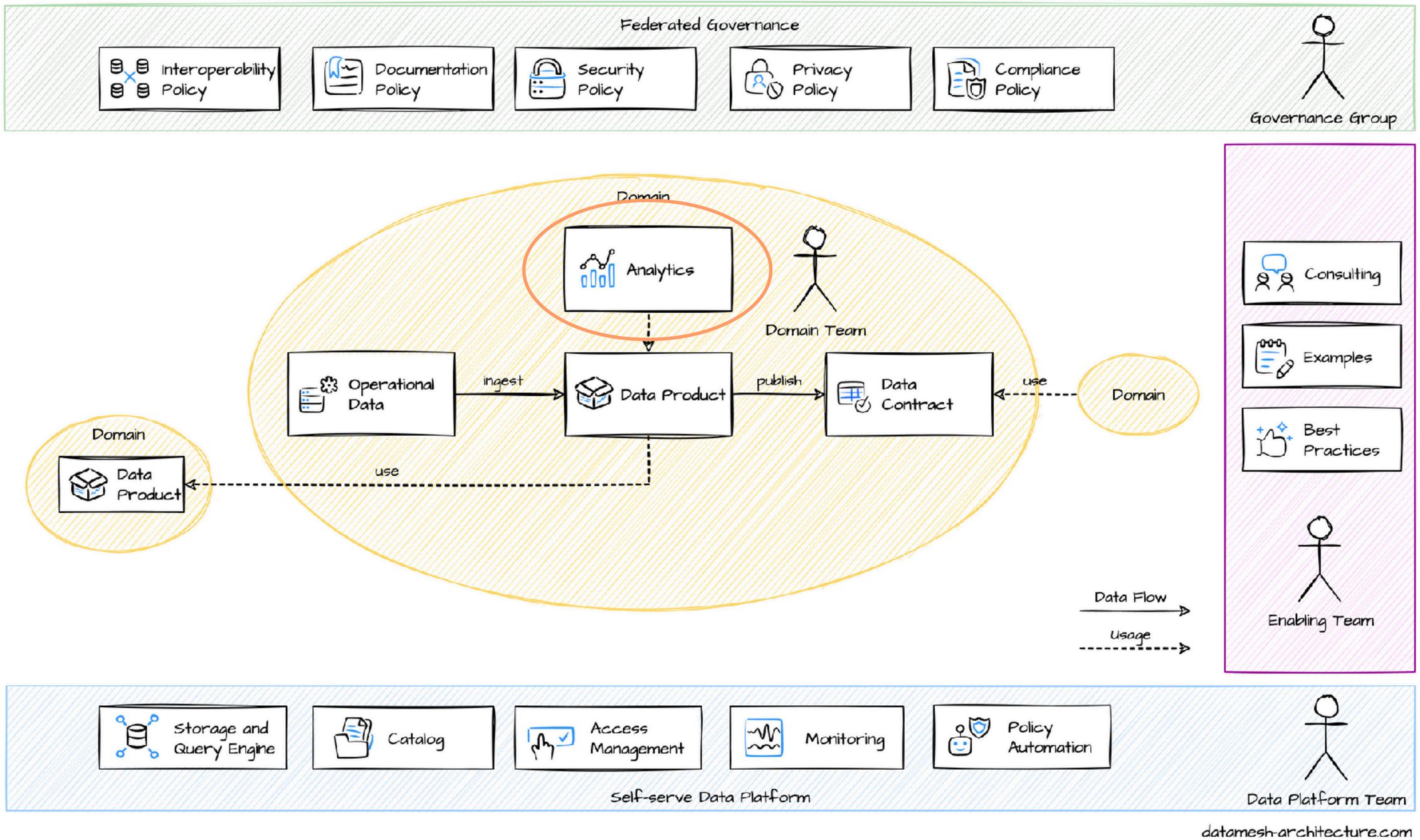
Transformation: SQL

entities__inventory_history.sql

Raw

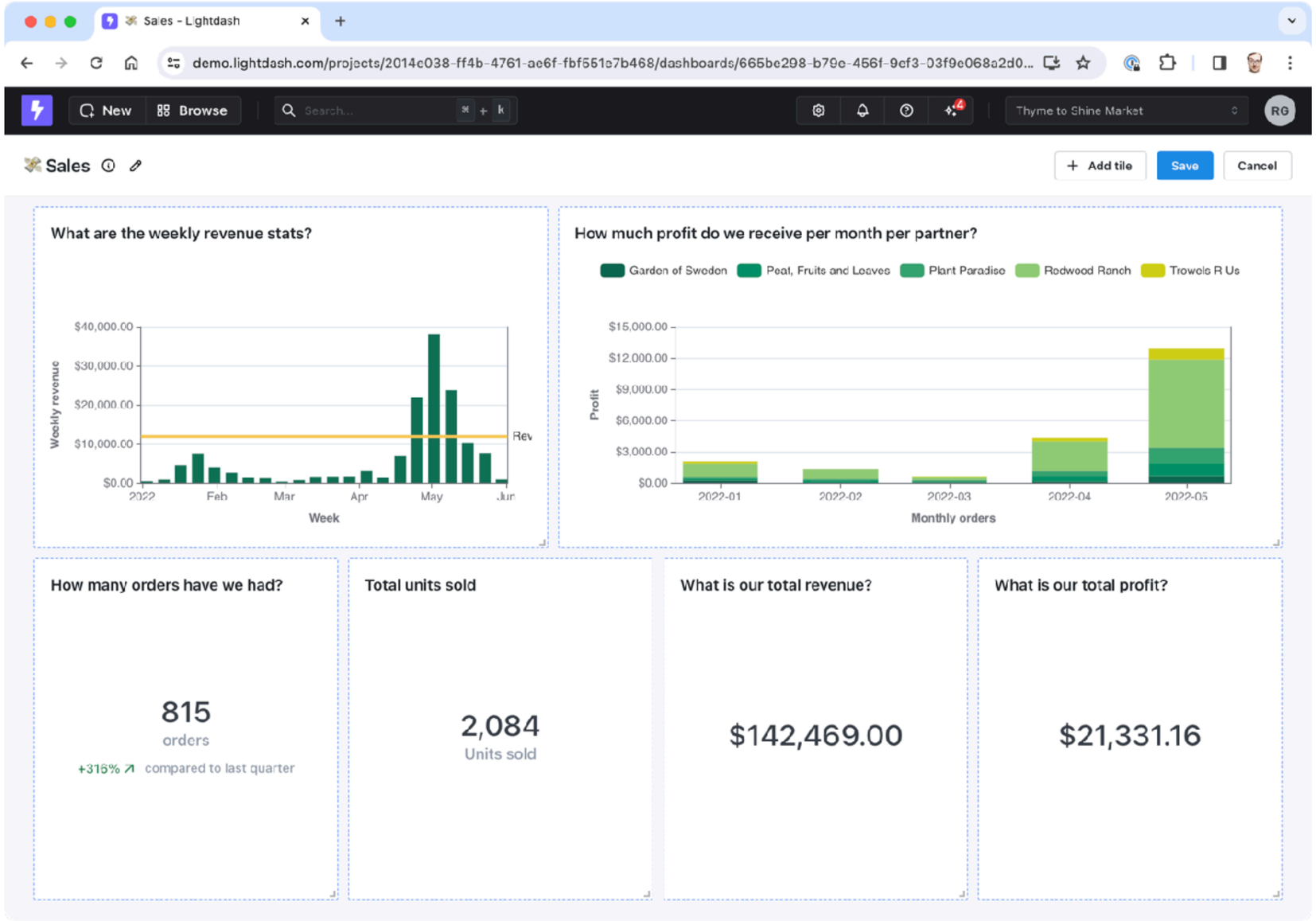
```sql
-- Step 1: Deduplicate
WITH inventory_deduplicated AS (
    SELECT *
    EXCEPT (row_number)
    FROM (
       SELECT *,
              ROW_NUMBER() OVER (PARTITION BY id ORDER BY time DESC) row_number
       FROM `datameshexample-fulfillment.raw.inventory`)
    WHERE row_number = 1
),
-- Step 2: Parse JSON to columns
inventory_parsed AS (
    SELECT
        json_value(data, "$.sku")                          AS sku,
        json_value(data, "$.location")                     AS location,
        CAST(json_value(data, "$.available") AS int64)     AS available,
        CAST(json_value(data, "$.updated_at") AS timestamp) AS updated_at,
    FROM inventory_deduplicated
)
-- Step 3: Actual Query
SELECT sku, location, available, updated_at
FROM inventory_parsed
ORDER BY sku, location, updated_at
```

Data Mesh Architecture

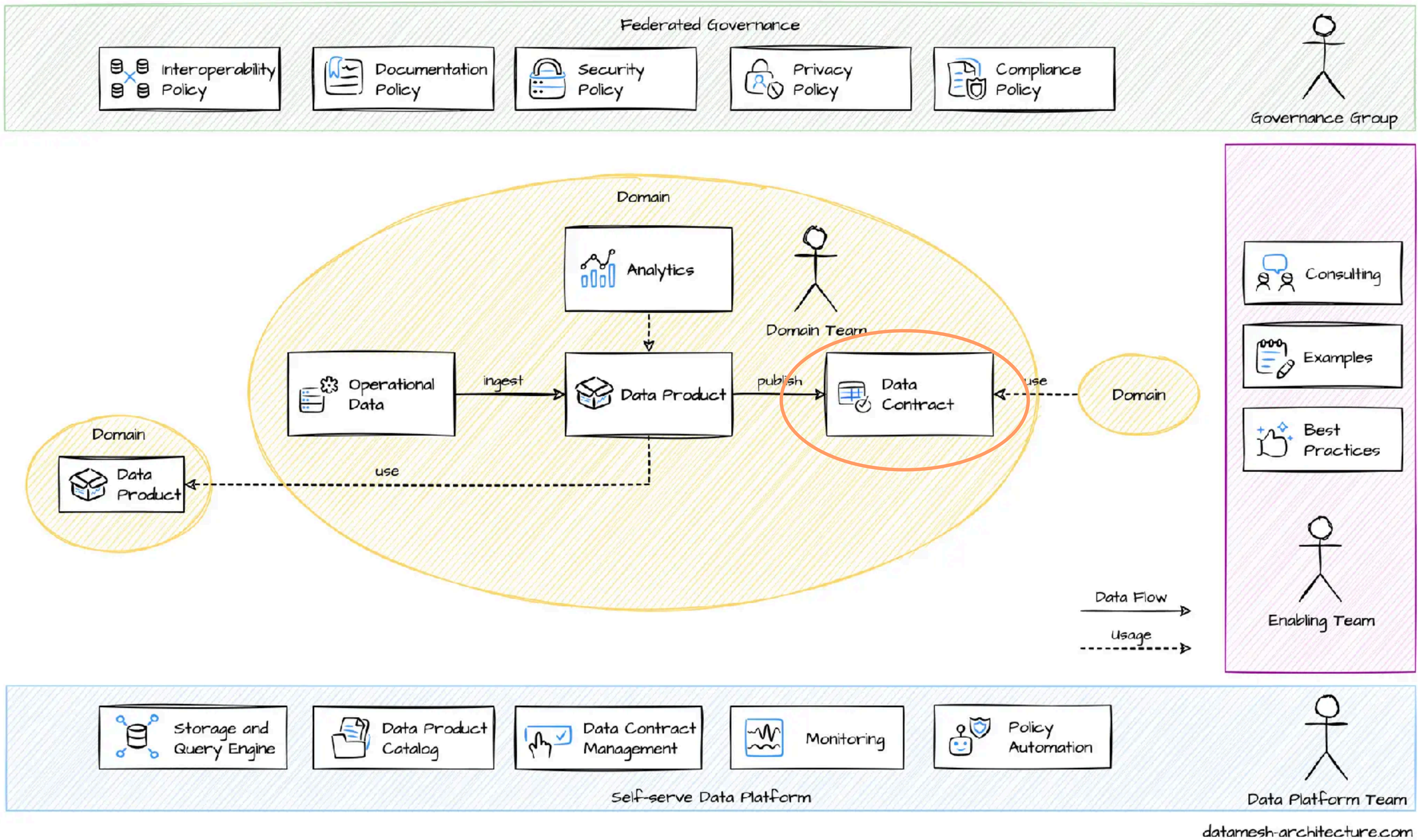


datamesh-architecture.com

Analytics: Enable Data Culture

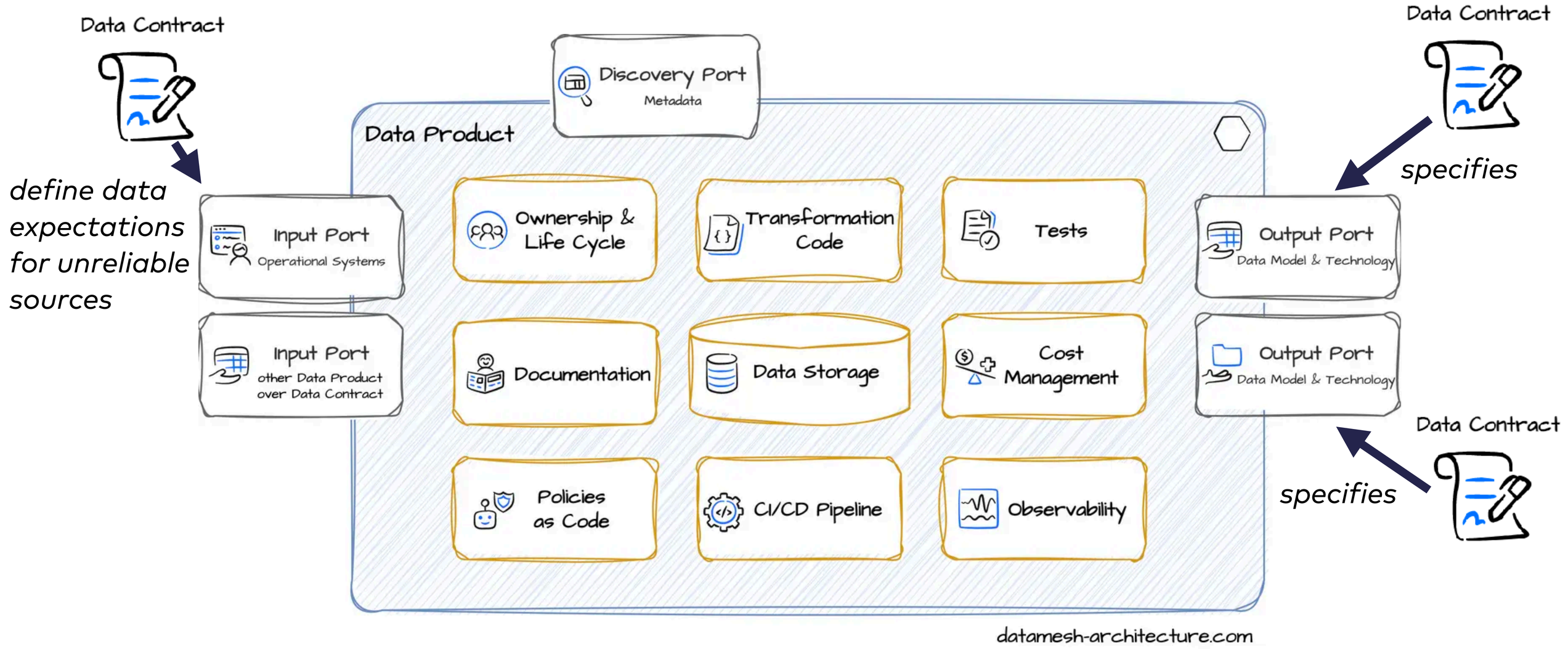


Data Mesh Architecture



datamesh-architecture.com

Data Product and Data Contracts



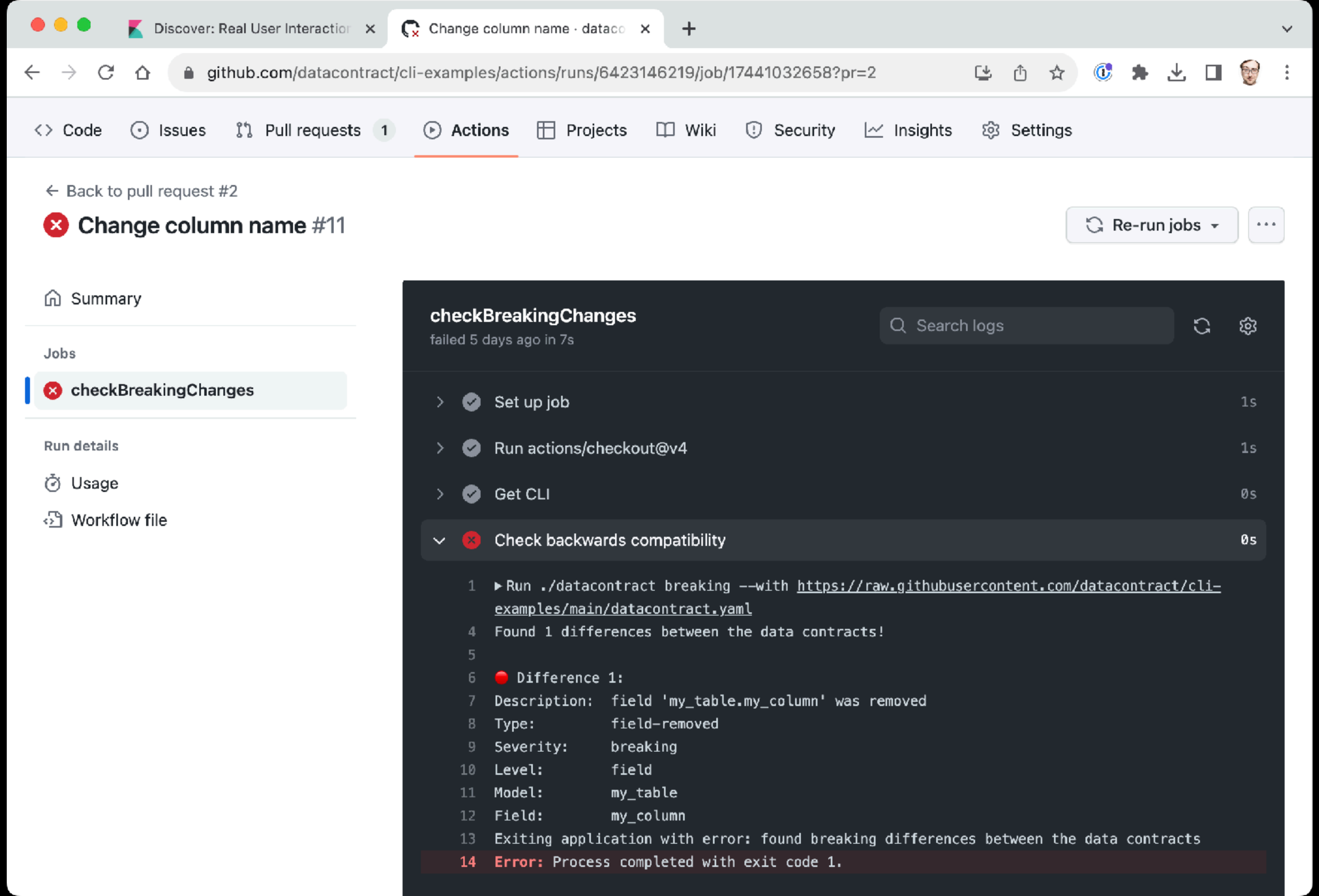
Data Contract

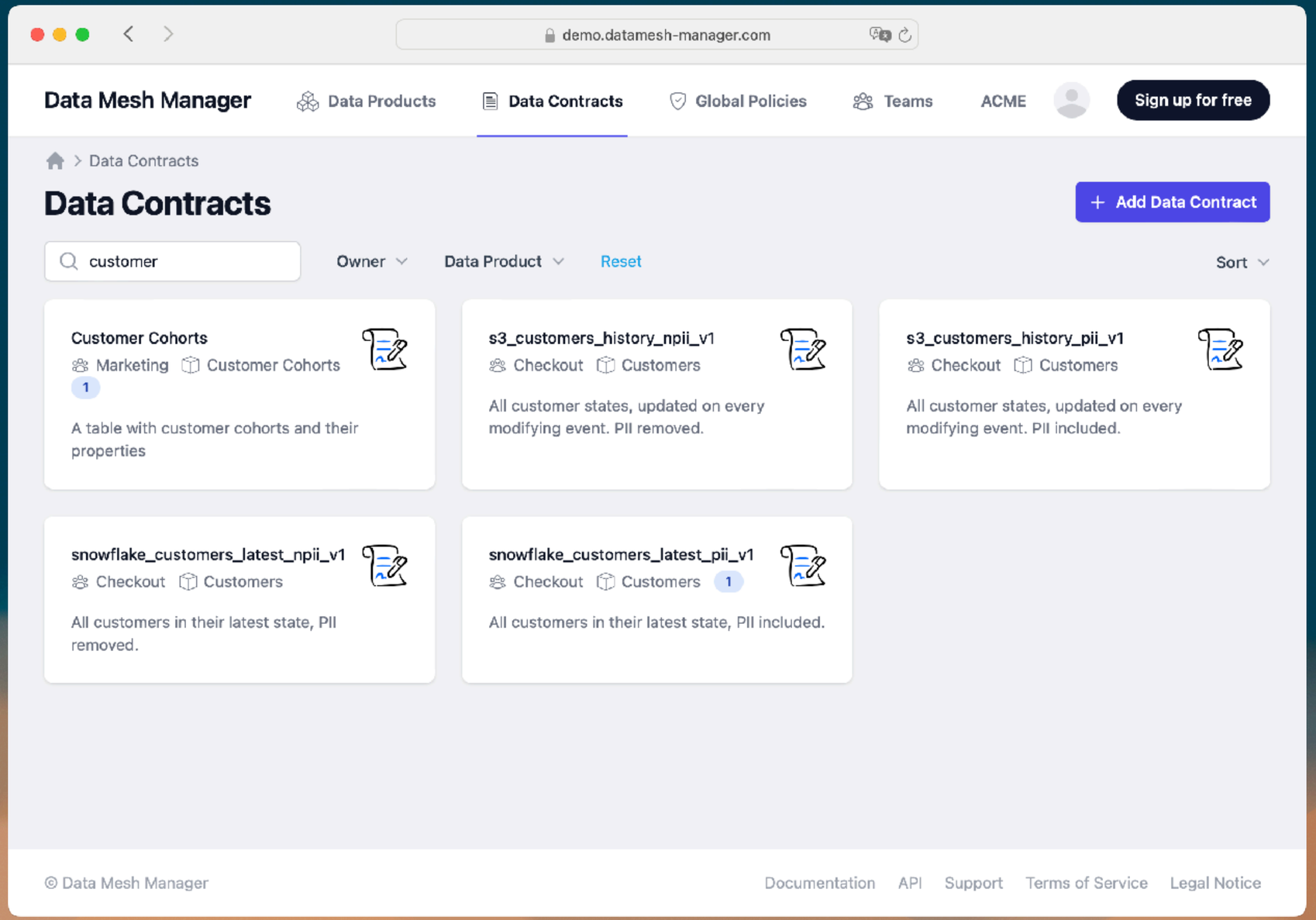
```yaml
dataContractSpecification: 0.9.1
id: web-orders-with-consent-v1
info:
  title: Web Orders With Consent V1
  version: 1.0.0
  description: "All orders made through the web channel.\r\nFiltered for orders where customers have expressed consent for analytical use."
  owner: checkout
  contact:
    url: https://teams.example.com/datacontracts/web-orders-with-consent-v1
terms:
  usage: "The data can be used for analytical and data science use cases, as the customer has expressed their consent."
  limitations: "As the dataset is filtered, these data set cannot be used to aggregate financial KPIs.\r\nNot suited for real-time use cases."
  billing: $1000 per month
  noticePeriod: P3M
models:
  orders:
    type: table
    description: A successful sale in the web shop
    fields:
      order_id:
        type: string
        description: Primary key of the order
      billing_customer_id:
        type: string
        description: Customer ID of the billing customer
      shipment_customer_id:
        type: string
        description: Customer ID of customer to ship the order to
      sold_timestamp:
        type: timestamp_tz
        description: The timestamp of the final confirmation step in the web form.
      total_amount:
        type: bigint
        description: The total order amount in the smallest unit of the currency (such as Eurocents)
```

datacontract.com
cli.datacontract.com

- Interface Specification (like OpenAPI, but for data)
- YAML
- Define Requirements
- Make expectations explicit
- Make domain knowledge explicit
- Common language for data providers and consumers
- Owned by a team
- Contract-first
- Enforce Contract in CI/CD

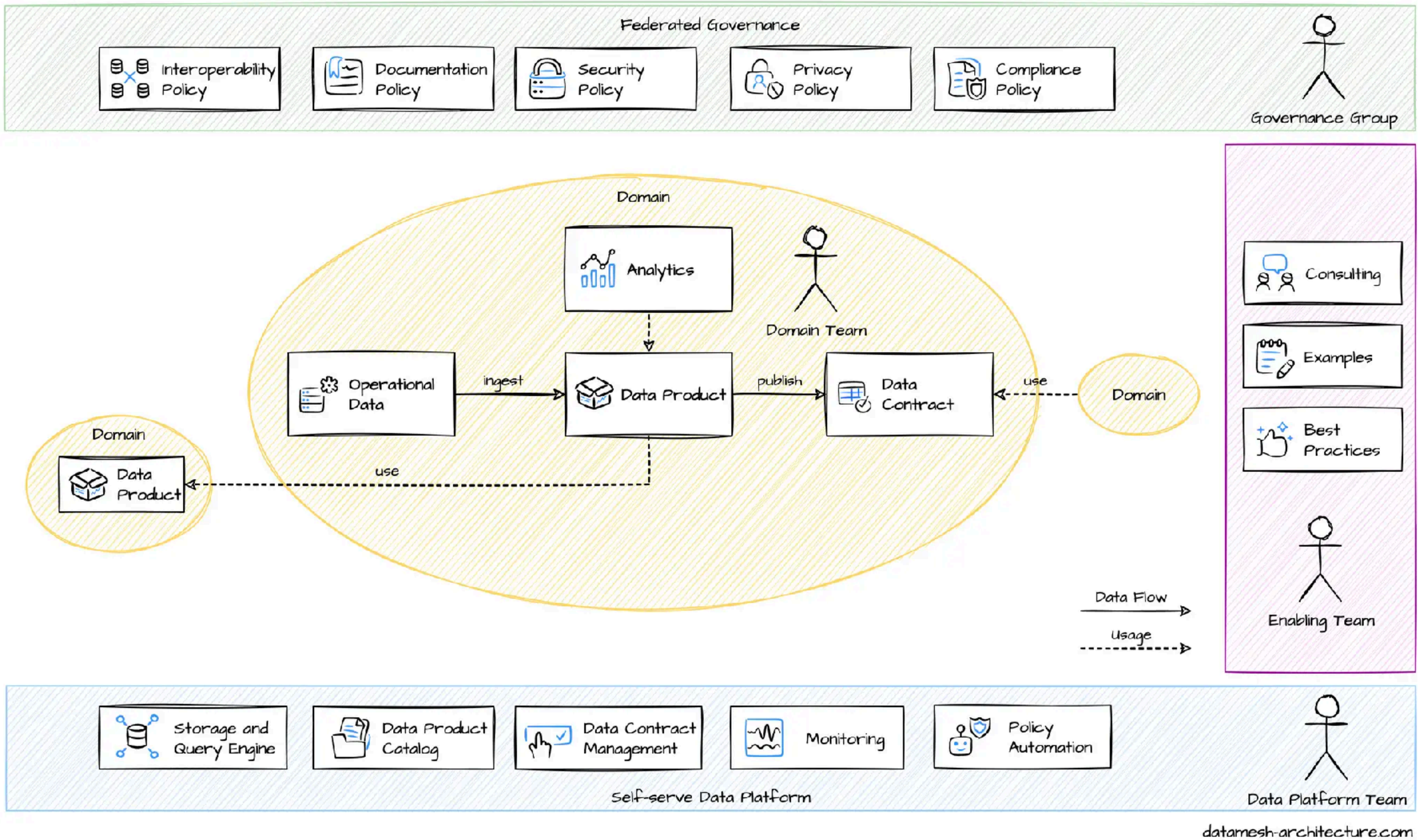
```yaml
dataContractSpecification: 0.9.1
id: web-orders-with-consent-v1
info:
  title: Web Orders With Consent V1
  version: 1.0.0
  description: "All orders made through the web channel.\r\nFiltered for orders where customers have expressed consent for analytical use."
  owner: checkout
  contact:
    url: https://teams.example.com/datacontracts/web-orders-with-consent-v1
terms:
  usage: "The data can be used for analytical and data science use cases, as the customer has expressed their consent."
  limitations: "As the dataset is filtered, these data set cannot be used to aggregate financial KPIs.\r\nNot suited for real-time use cases."
  billing: $1000 per month
  noticePeriod: P3M
models:
  orders:
    type: table
    description: A successful sale in the web shop
    fields:
      order_id:
        type: string
        description: Primary key of the order
      billing_customer_id:
        type: string
        description: Customer ID of the billing customer
      shipment_customer_id:
        type: string
        description: Customer ID of customer to ship the order to
      sold_timestamp:
        type: timestamp_tz
        description: The timestamp of the final confirmation step in the web form.
      total_amount:
```


Discover: Real User Interactior
Change column name · datacc
github.com/datacontract/cli-examples/actions/runs/6423146219/job/17441032658?pr=2
Code
Issues
Pull requests 1
Actions
Projects
Wiki
Security
Insights
Settings
Back to pull request #2
Change column name #11
Re-run jobs
Summary
Jobs
checkBreakingChanges
Run details
Usage
Workflow file
checkBreakingChanges
failed 5 days ago in 7s
Search logs
Set up job 1s
Run actions/checkout@v4 1s
Get CLI 0s
Check backwards compatibility 0s
1 ▶ Run ./datacontract breaking --with https://raw.githubusercontent.com/datacontract/cli-examples/main/datacontract.yaml
4 Found 1 differences between the data contracts!
5
6 🔴 Difference 1:
7 Description: field 'my_table.my_column' was removed
8 Type: field-removed
9 Severity: breaking
10 Level: field
11 Model: my_table
12 Field: my_column
13 Exiting application with error: found breaking differences between the data contracts
14 Error: Process completed with exit code 1.



datamesh-manager.com

Data Mesh Architecture



datamesh-architecture.com

DALL·E 2: elephant in a room with a software engineer, pixel art

The Elephant in the Room

(The Problem with Data Mesh)

Dev

Fullstack Dev

Fullstack DevOps

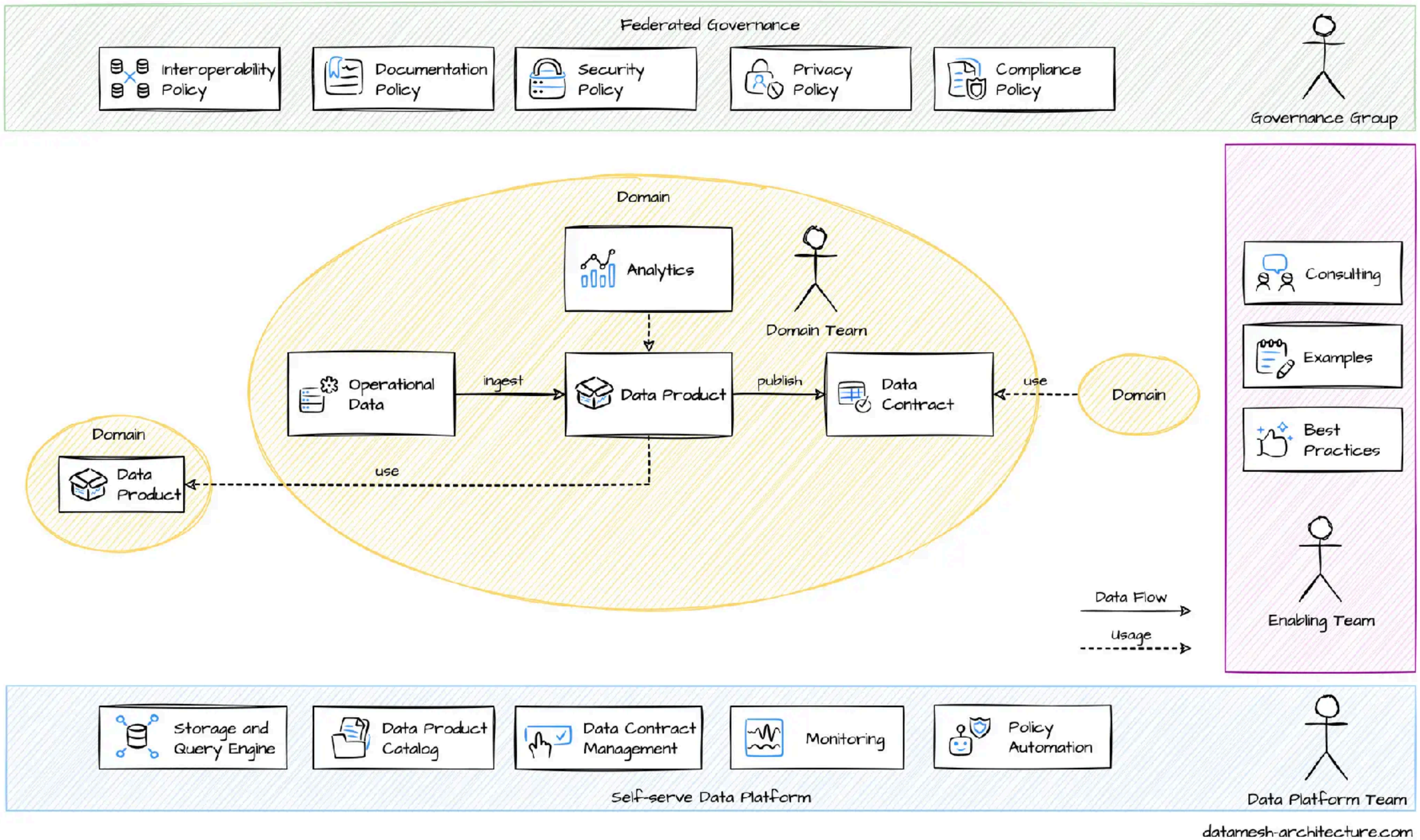
Fullstack DevSecOps

Fullstack BizDevSecOps

Fullstack
BizDevSecDataOps

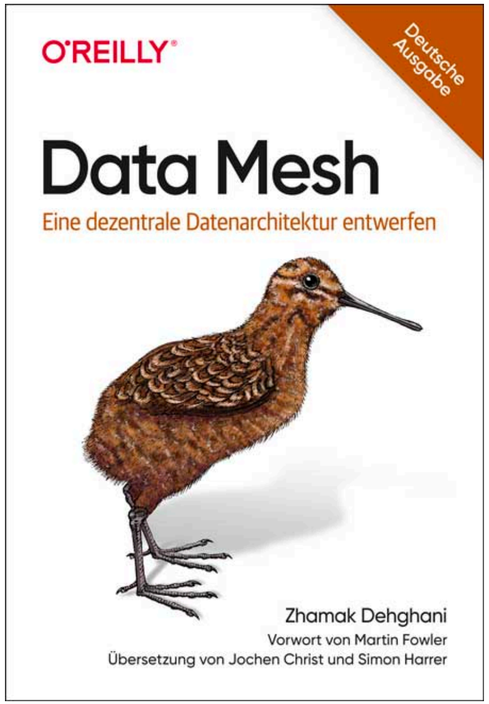
DALL·E 2: A person sits in front of a computer and is exhausted, pixel art

Data Mesh Architecture

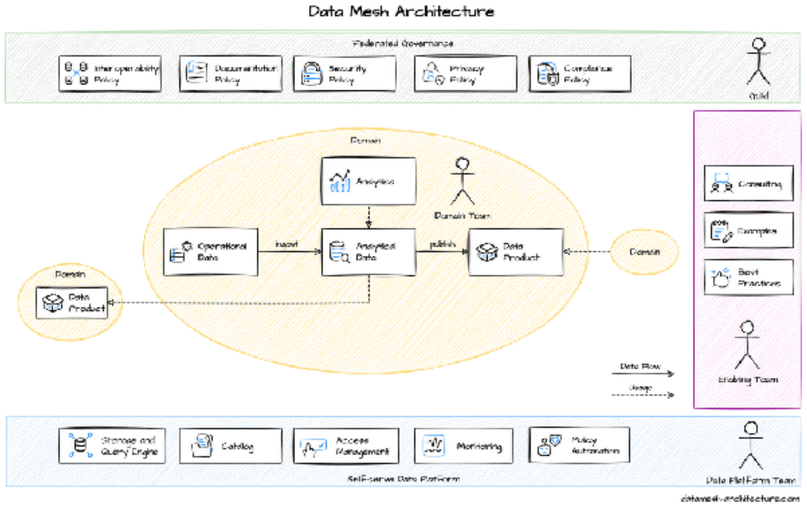


datamesh-architecture.com

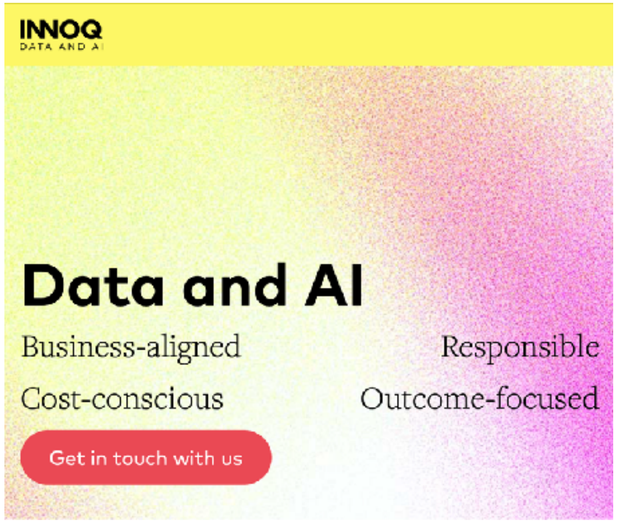
Learn more



oreilly.de/produkt/data-mesh



datamesh-architecture.com



INNOQ.ai
Data Mesh Consulting, Trainings,
Data Product Engineering

Data Mesh

Introduction

INNOQ

JOCHEN CHRIST
@JOCHEN_CHRIST