

# MODERN DATA PLATFORM

A PRACTICAL EXAMPLE OF DATA MESH IN PRACTICE

CDC-ONC INDUSTRY DAY 2023

FEBRUARY 28TH, 2023

PRESENTED BY VITALIY BAKLIKOV, SKYWARD IT SOLUTIONS



# INTRO

- VP OF ENGINEERING AT SKYWARD
- PREVIOUSLY SVP, DATA ANALYTICS AT DBS BANK
- BEEN BUILDING DATA PLATFORMS AND ML SYSTEMS SINCE 2003



# AGENDA



What is Data Mesh



Four Pillars of Data Mesh

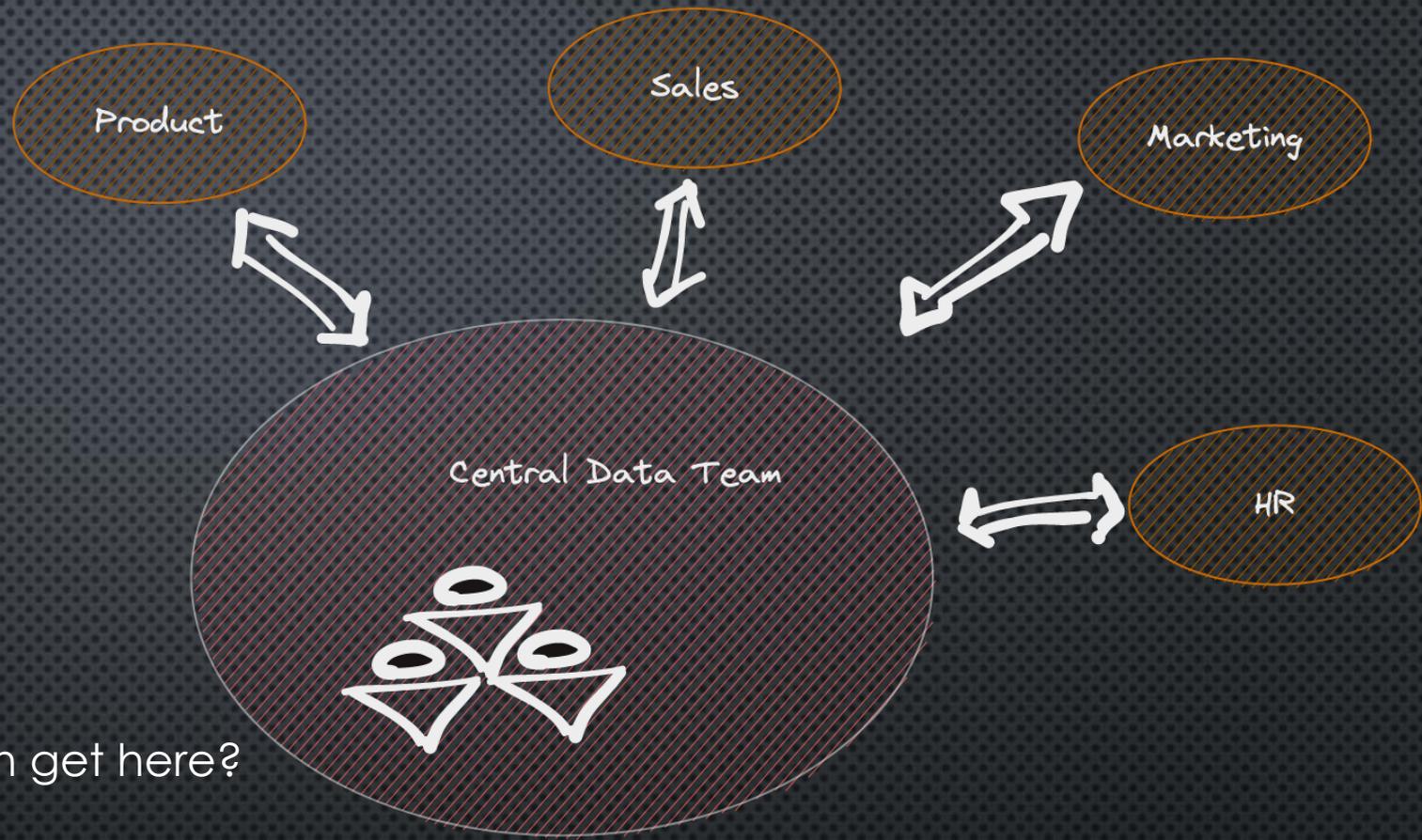


Data Mesh in Practice



Conclusion

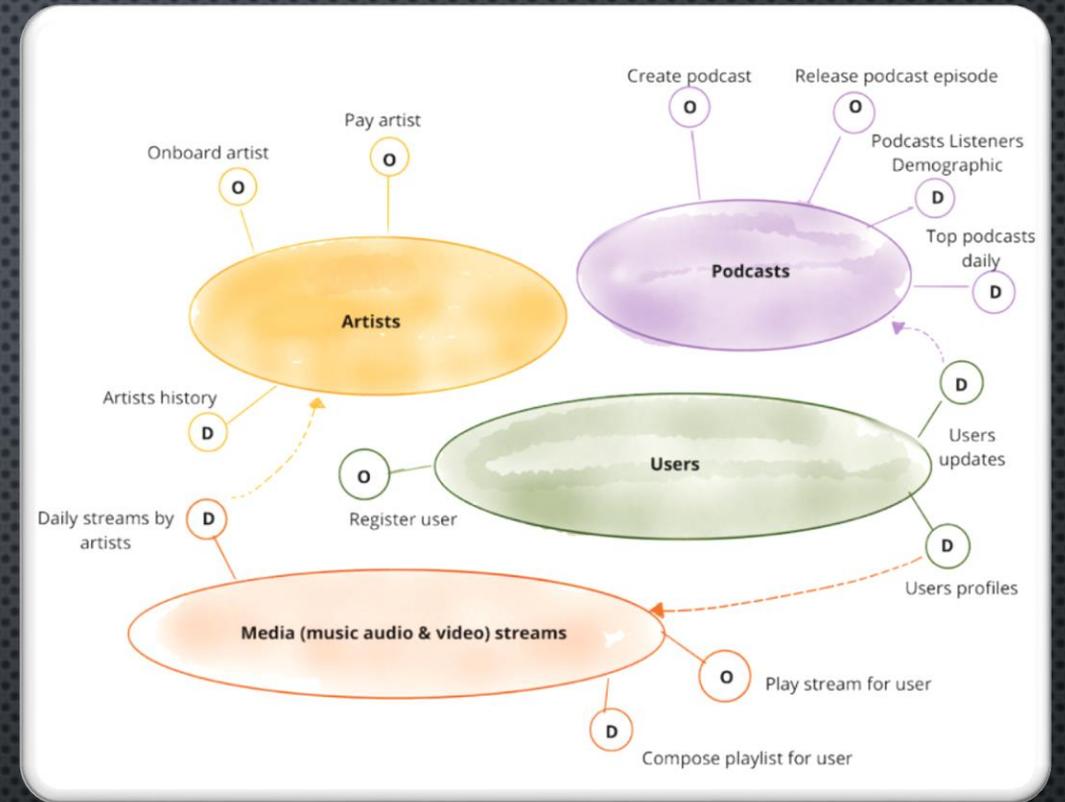
# WHAT IS DATA MESH



But before we define it, How did we even get here?

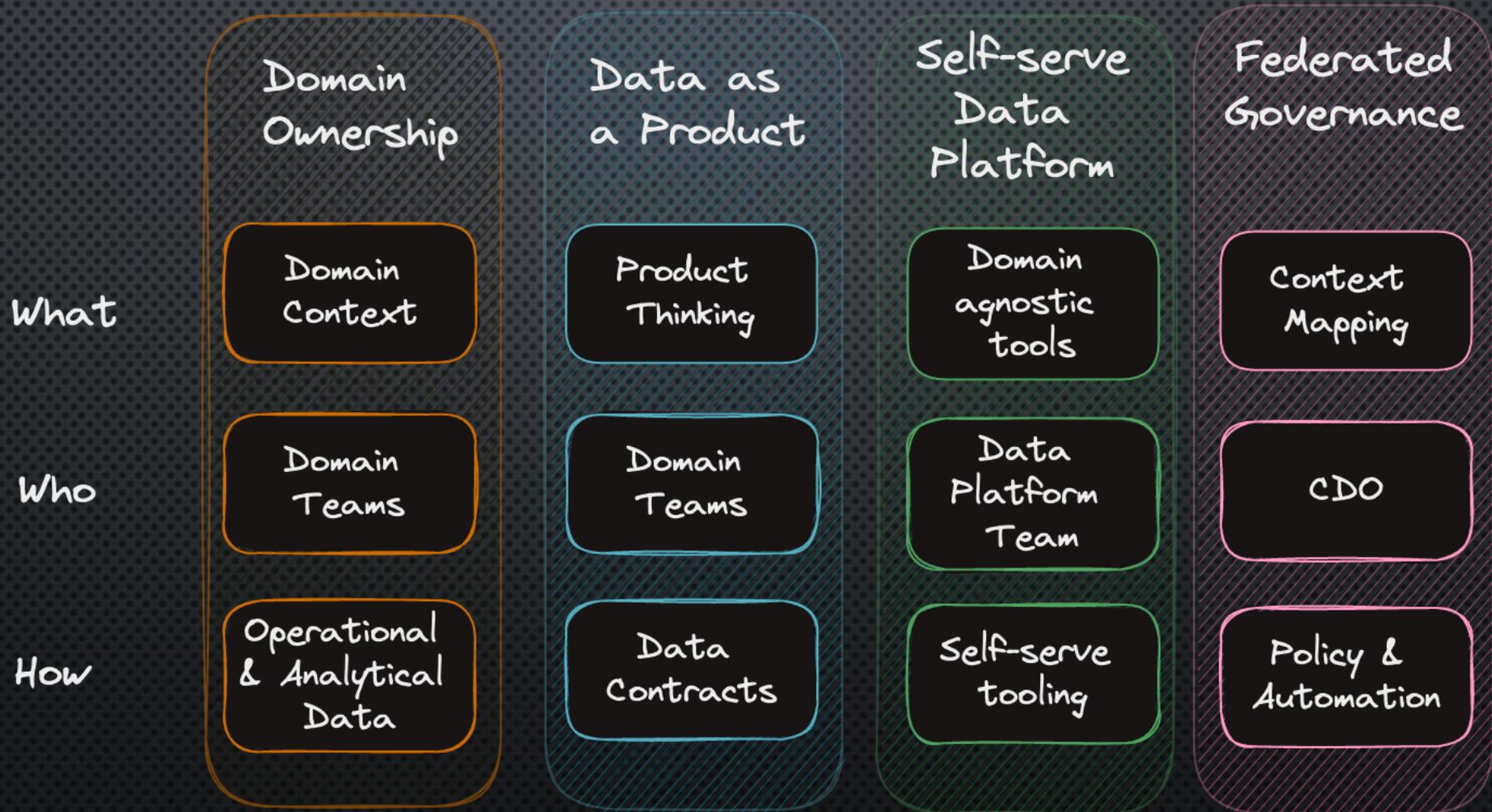
- Central Data Team (DBAs or IT) is rarely SME on Data
- Orgs are growing, Data is growing, Central Teams DON'T KEEP UP

# MEET DATA MESH



An architectural paradigm for organizing Data Ecosystems in large, complex Organizations

# 4 PILLARS OF DATA MESH



What

Domain  
Ownership

Domain  
Context

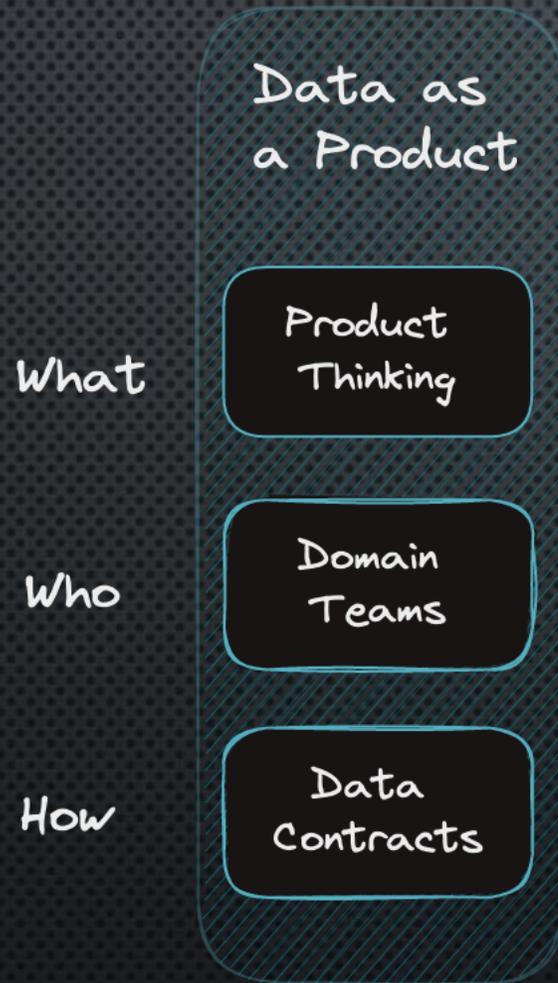
Who

Domain  
Teams

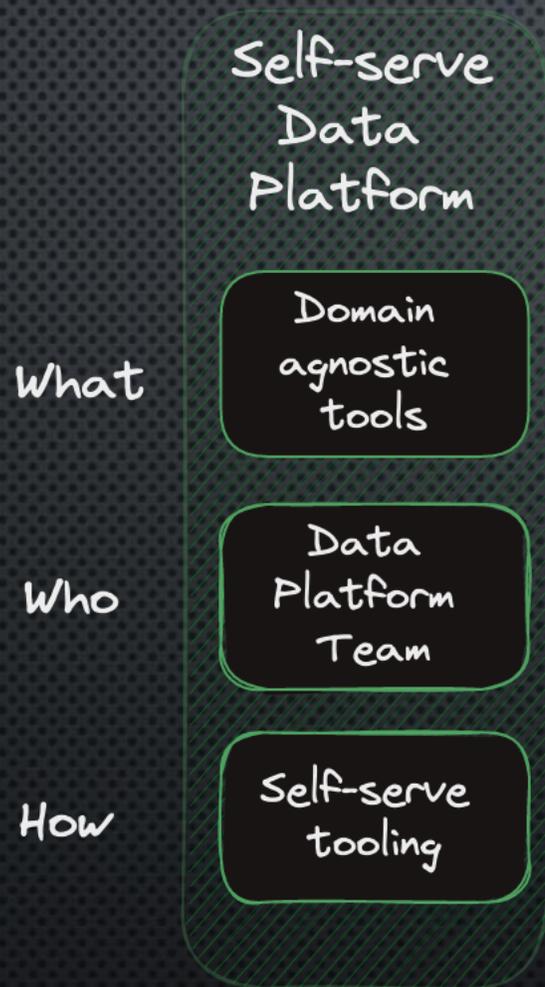
How

Operational  
& Analytical  
Data

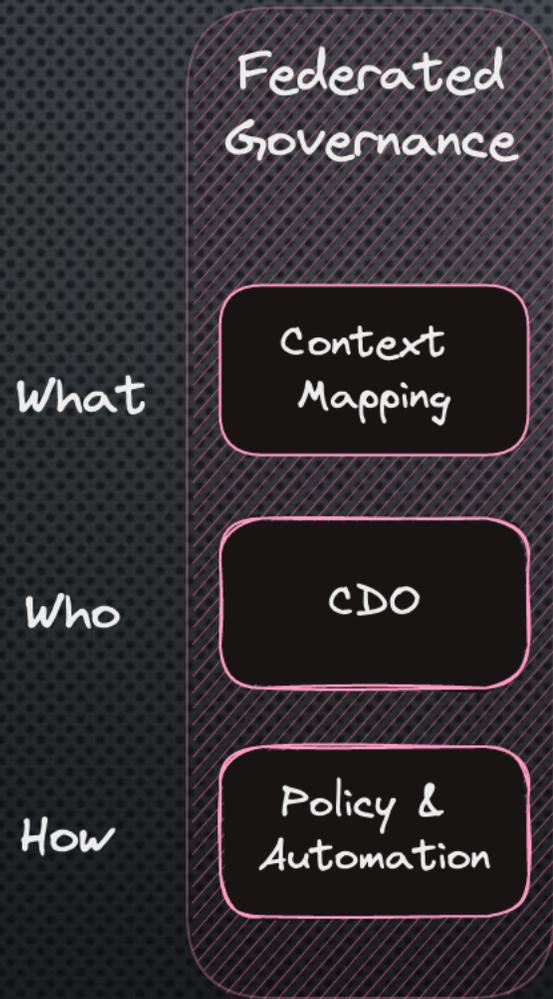
- DOMAIN TEAM SMEs OWN DATA
- SELF-SUFFICIENT TO BUILD AND RUN
- ANALYTICAL DATA COMPOSED AROUND DOMAINS



- PRODUCT THINKING PHILOSOPHY AROUND ANALYTICAL DATA
- DATA PRODUCTS ARE “OFFERED” BY DOMAIN TEAMS TO THE ORG CUSTOMERS
- DATA CONTRACTS ARE WELL DEFINED



- PRODUCT THINKING TO DATA INFRASTRUCTURE
- CUSTOMERS: DATA PRODUCERS, DATA CONSUMERS, DATA GOVERNORS
- THE STORE IN DATA MARKETPLACE ANALOGY

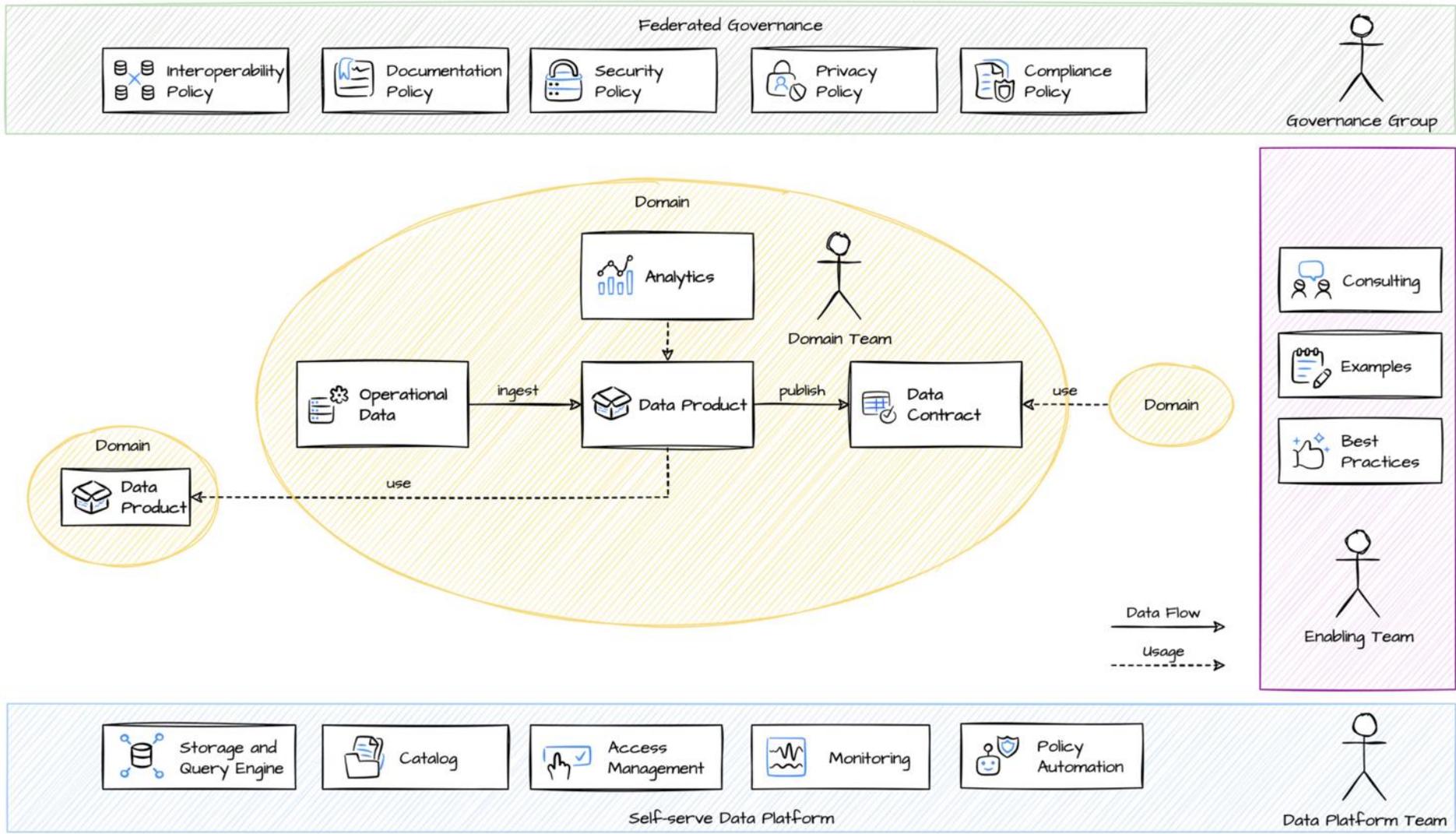


- THE MESH IN DATA MESH
- DATA POLICY & INDUSTRY REGULATION
- PREFERABLY AUTOMATED AND ADHERED TO

# DATA MESH IN PRACTICE.... NOT THAT SIMPLE! :)

- STARTED ON DATA MESH JOURNEY IN LATE 2017 (BEFORE THE TERM WAS EVEN DEFINED)
- DATA IN SILOS – MAINFRAME, HADOOP, EXADATA, TERADATA, SAS, ETC.
- LEGACY TECH STACK AND PROCESSES – CONTROLLED BY CENTRAL DATA TEAM
- CLOUD ASPIRATIONS





- STARTED ON-PREM THEN AWS
- STORAGE – IBM CLEVERSAFE OBJECT STORE, AWS S3
- QUERY ENGINE – PRESTO, ATHENA
- ETL – APACHE SPARK, EMR
- **STANDARDIZE DATA INGESTION!**

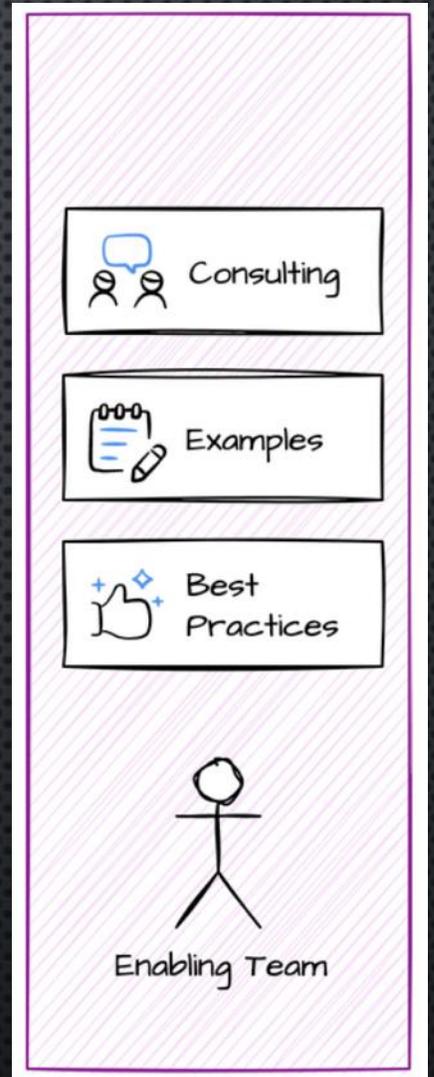
- DATA CATALOG – COLLIBRA
- METADATA DRIVEN INGESTION

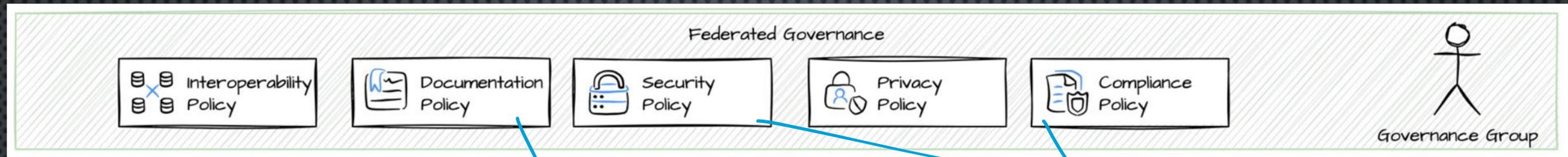
- ACLS AND RBAC ENGINE
- TOKENIZATION AND ENCRYPTION
- METADATA MATURITY LEVELS



# TIP 1: INVEST IN ENABLING TEAM!

THIS IS THE TEAM THAT WILL DRIVE YOUR FIRST DATA PRODUCT





**Policy 1: Data Format**

Format: CSV  
 Standard: RFC 4180  
 Delimiter: COMMA  
 Header-Record: Required

**Policy 2: Location**

FileServer  
 S3 Bucket owned by Domain Team

**Policy 3: Data Discovery**

A page on Enterprise Confluence that lists:  
 Domain Team  
 Data Product Name  
 Product Owner  
 Location  
 Data Freshness  
 Tags  
 Data Model

**Policy 4: Access Control**

Role based ACL

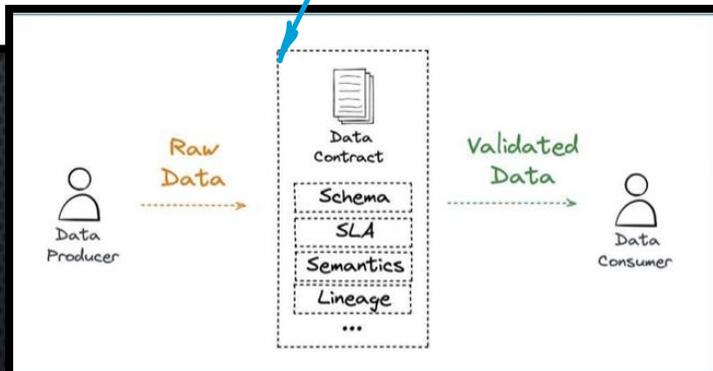
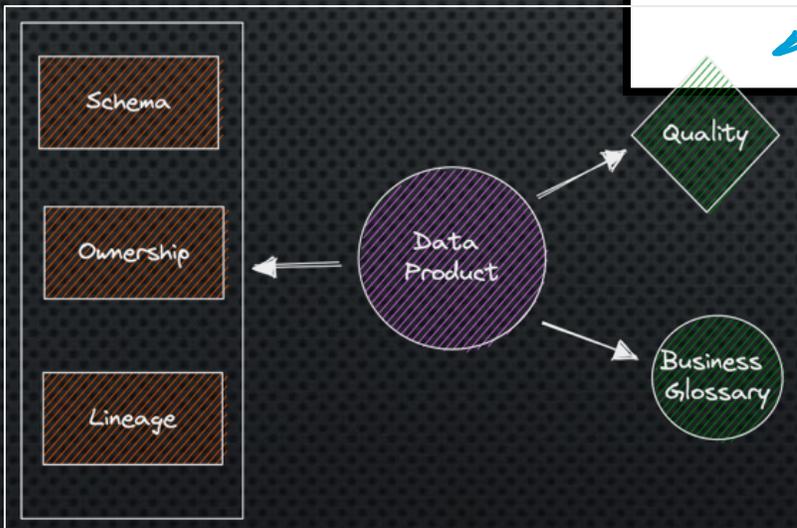
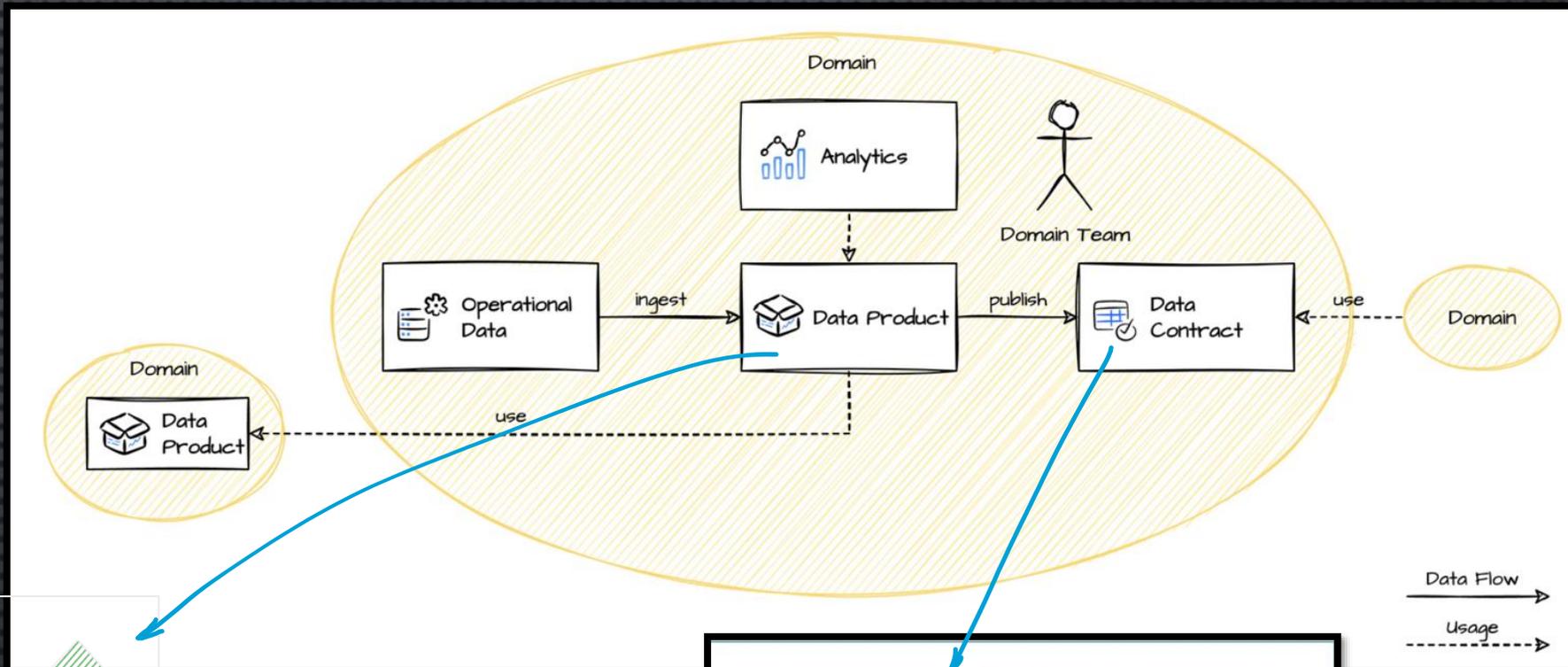
**Policy 5: Data Classification**

All PII data is Masked with \*\*\*\*\*

How can you automate Governance compliance?

TIP 2: IF YOU'RE BUILDING YOUR 1<sup>ST</sup> DATA PRODUCT, GO EASY





**Tip 3:** Start small and Re-Use artefacts. Data Modeling - take an existing definition OMOP CDM for example. HL7 FHIR, SNOMED CT Vocab, Athena OHDSI

# CONCLUSION



Data Mesh is akin to Data Marketplace



Product Thinking at every step of the way



It only works if your teams are supportive of each other and the initiative



Be Prepared that ONE team will be carrying the rest of the ORG



**Skyward**  
IT SOLUTIONS

THANK YOU!

CDC-ONC INDUSTRY DAY 2023  
FEBRUARY 28TH, 2023

PRESENTED BY VITALIY BAKLIKOV, SKYWARD IT SOLUTIONS

