

A complex network graph with numerous nodes represented by small circles of varying sizes and colors (white, light orange, yellow, pink, purple) connected by a dense web of thin white lines. The background has a warm, orange-to-red gradient.

# Designing Data Governance for Data Mesh Architecture

Lauren Maffeo

Data Science DC Meetup

February 13, 2023



# Agenda

1. The Big Data Dilemma(s)
2. Why Data Mesh?
3. Intro to Data Governance
4. Apply Design Thinking to Data





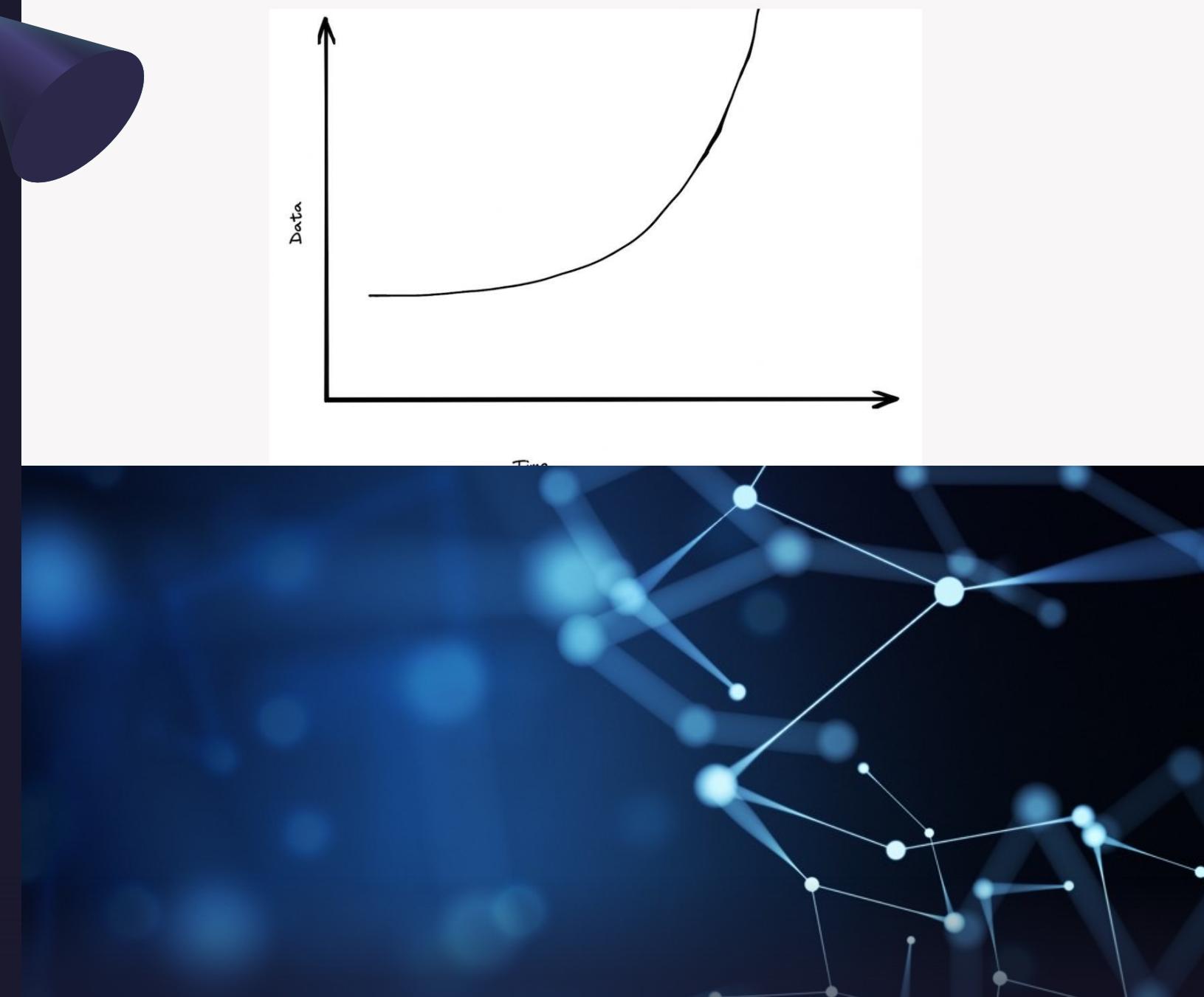
# Hi! I'm Lauren

- Civic tech designer of data/digital products (websites, databases, strategies, services)
- Advising an ACDO at a U.S. federal agency on data strategy
- Author of **Designing Data Governance from the Ground Up**
- 10+ years in tech, 5 contributing to open source (Drupal, OS.com, WWCode)

By year  $\text{CURRENT\_YEAR}() + \epsilon$  there will be {unfathomably large amount} of data generated

# Obligatory Intro Slide

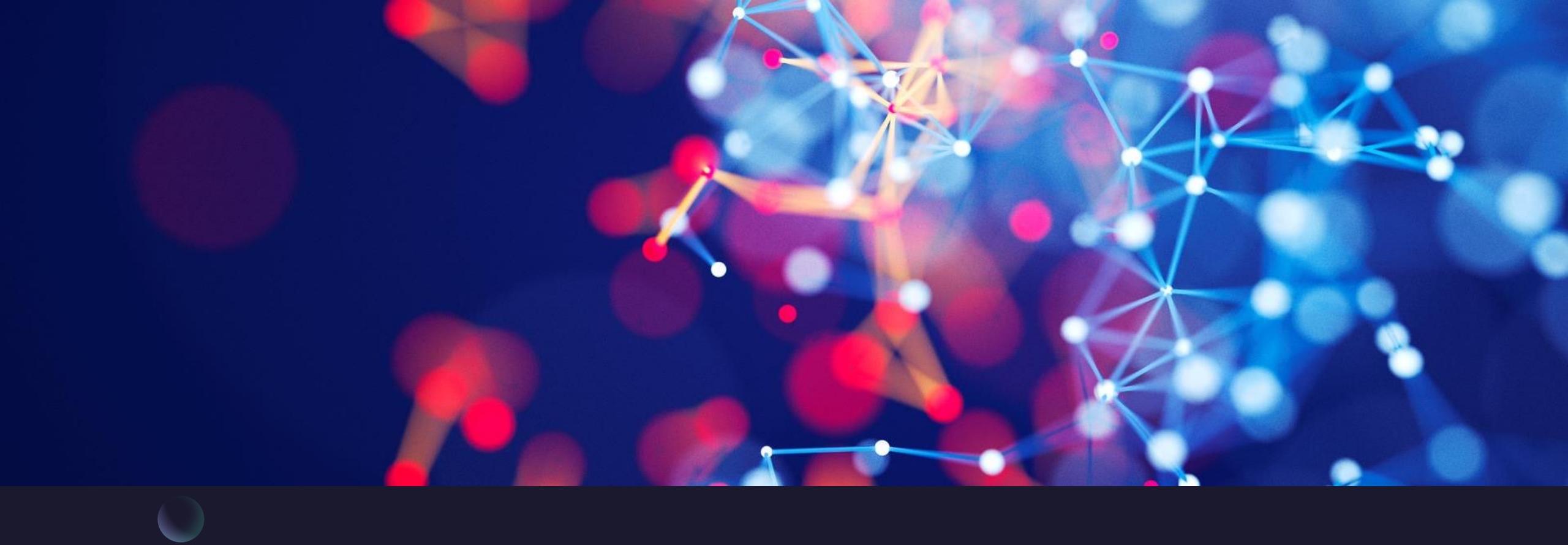
Credit: [Jordan Tigani of MotherDuck.com](#)



A complex network graph composed of numerous small, glowing red and blue spheres connected by thin lines, set against a dark blue background.

# Big Data is Dead, Long Live Big Data

- Data volumes grow by 63% per month on average
- Mean # of data sources per org is 400
- 99% will migrate to the cloud over 2 years
- 90% of execs say it's hard to prepare data for analytics

A complex network graph composed of numerous small, glowing nodes (red, blue, white) connected by thin lines, set against a dark blue background.

# Change is Hard

- 1 in 4 org leaders say they're data-driven
- Cultural hardships > Technical blockers
- App. 13% of ML models make it to production

“Everybody  
wants to change  
the world, but  
one thing’s clear:  
No one ever  
wants to change  
themselves.”

Nothing More, “Do You  
Really Want It?”





Federated data governance can help solve these problems.

It automates the standards needed to decentralize data.

# Data Governance + Data Mesh = Federated Data Governance

## GOVERNANCE

- A **strategy** to manage the people, processes, and tools related to how your org stores data/metadata
- Standards which share how your org collects, stores, processes, and destroys data
- Scaled to improve service delivery

## MESH

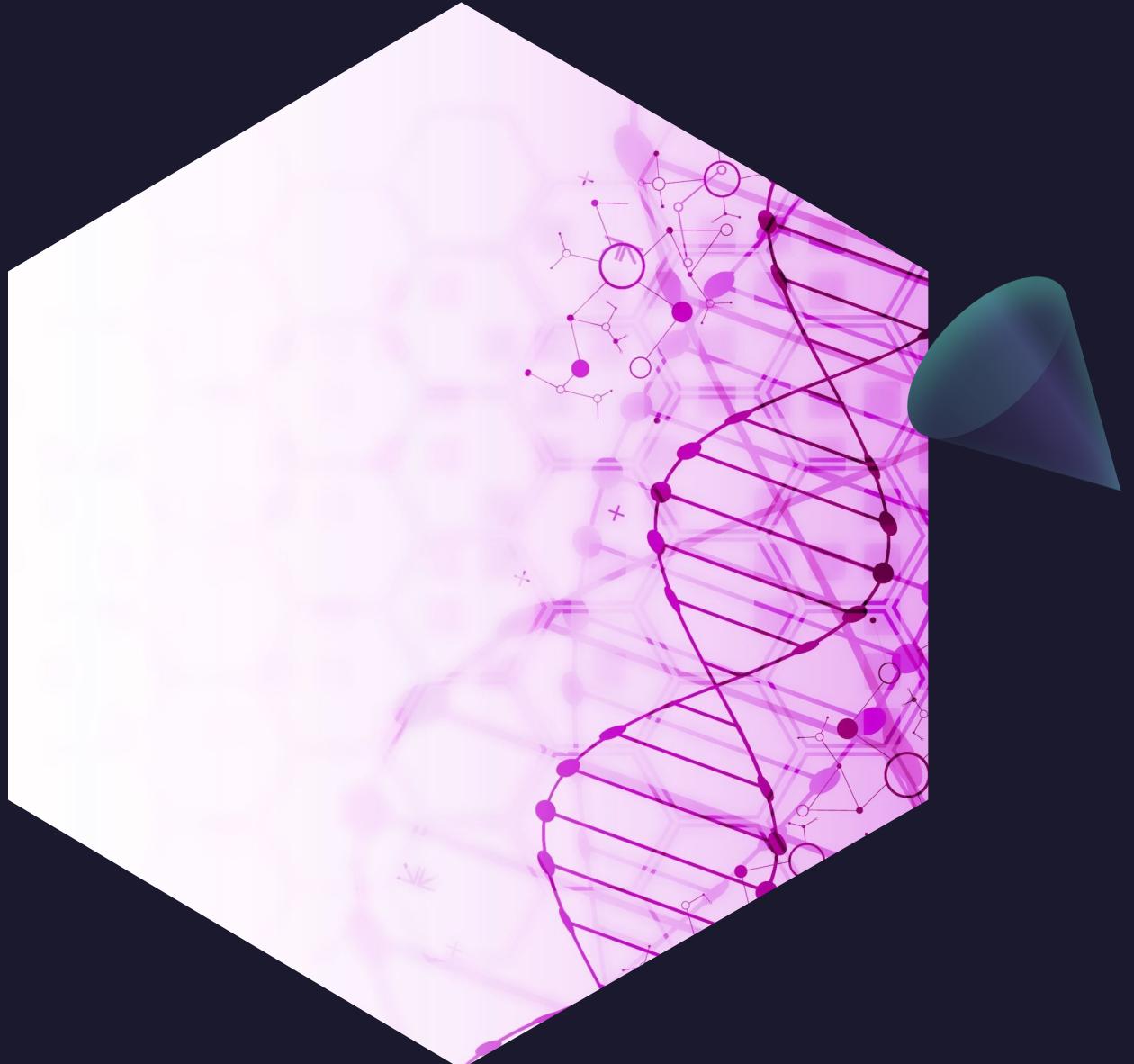
- Introduced by Zhamak Deghani
- Promotes a **data-as-products** mindset
- Manages products according to domains
- Keeps data available in self-service architecture
- Ensures that data is formatted/stored/discovered against equal standards

## FEDERATED DG

- Automates your governance standards throughout your decentralized architecture
- Guides domain-specific data stewards through the process to build data products
- Defines parameters for how to build/define/access/use data products in the data mesh

# Warning:

- Data mesh was first mentioned in 2019
- It is thus a young, evolving process
- Data mesh is no singular tool, role, or standard
- You can't do data mesh without governance



# Why Data Mesh?

## CORE TENANTS

1  
Domain-driven design

2  
Autonomous domain teams

3  
Decentralized microservice architecture

## BIG BENEFITS

- Data domain owners (aka data stewards) know/own/make decisions about their data
- Data mesh architecture supports cross-domain data analysis
- It interconnects data (akin to APIs in a microservice architecture)

# 4 Phases of Federated Governance



## Interoperability

Defines the parameters that allow stewards to leverage data products across the data mesh.



## Documentation

Helps everyone learn how to find and use data products.



## Security

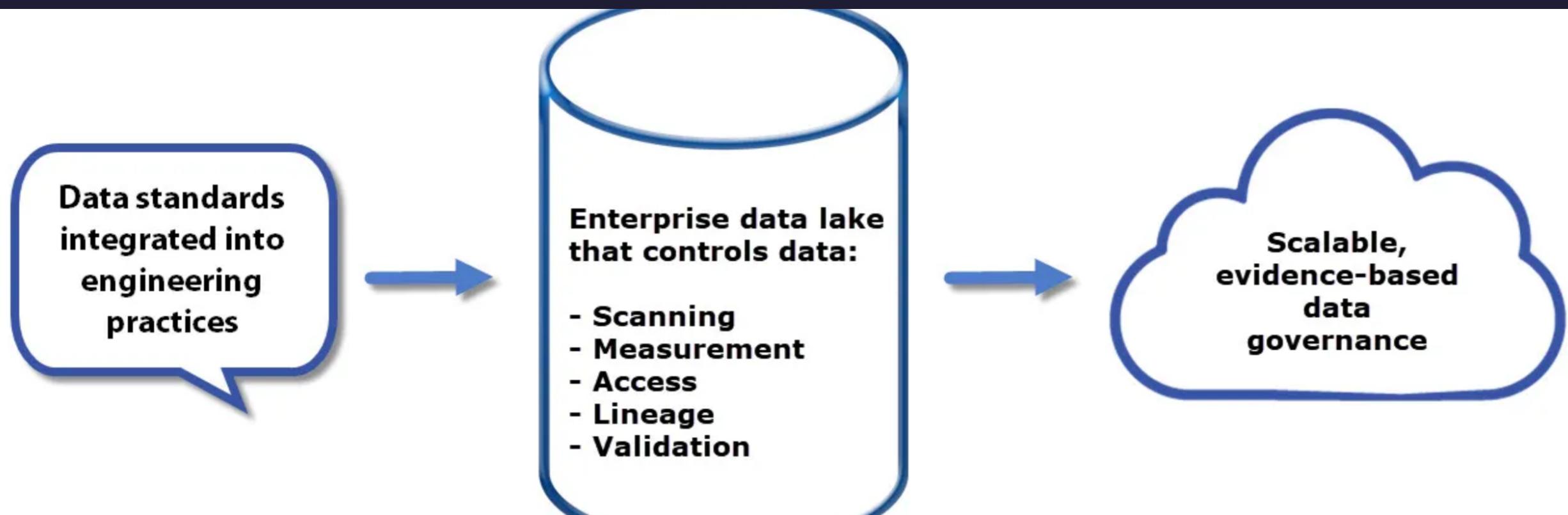
Sets role-based access by creating identity and access management policies within your cloud platform.

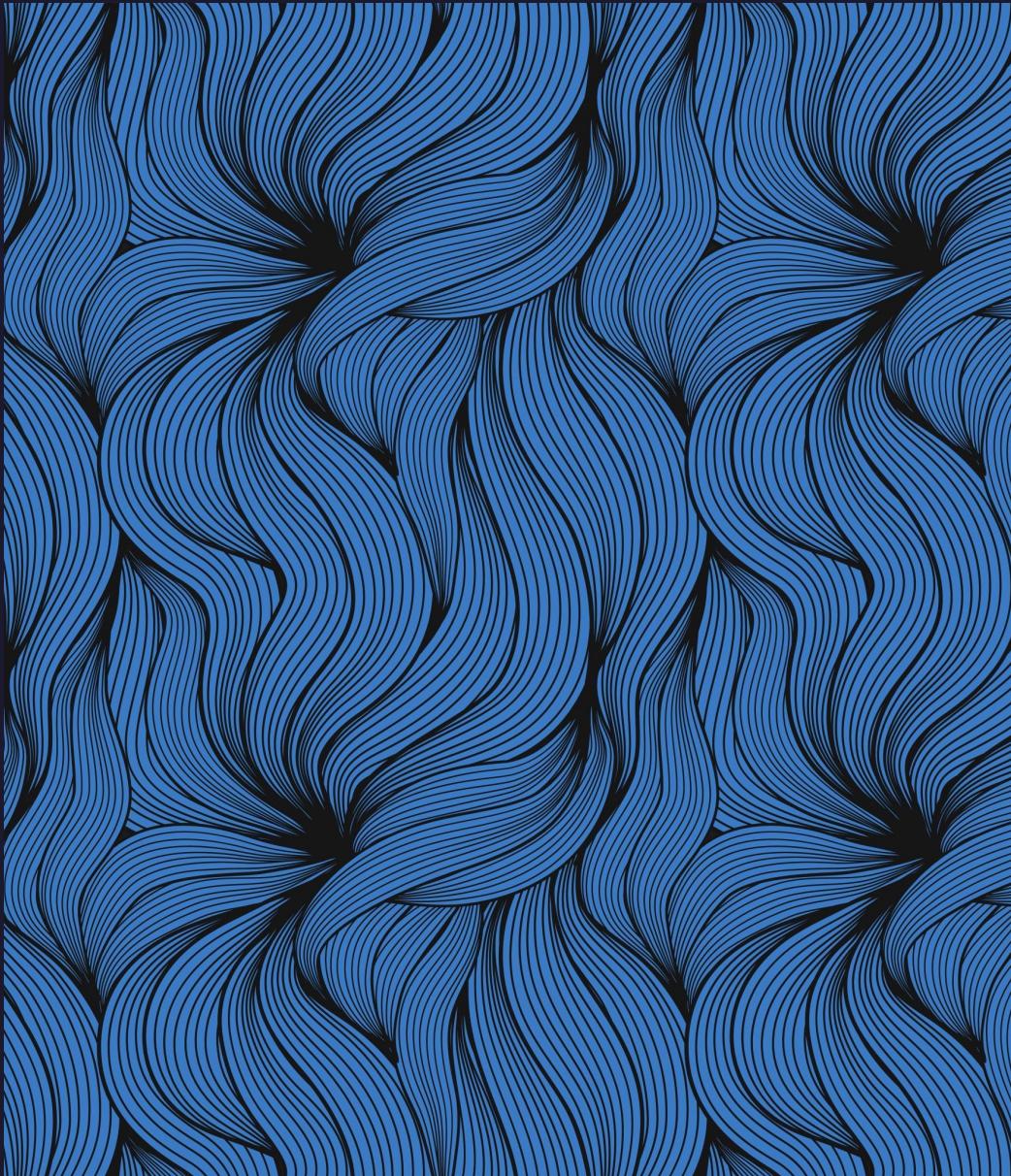


## Compliance

Automates requirements for legal rights, PII, etc. within your data mesh.

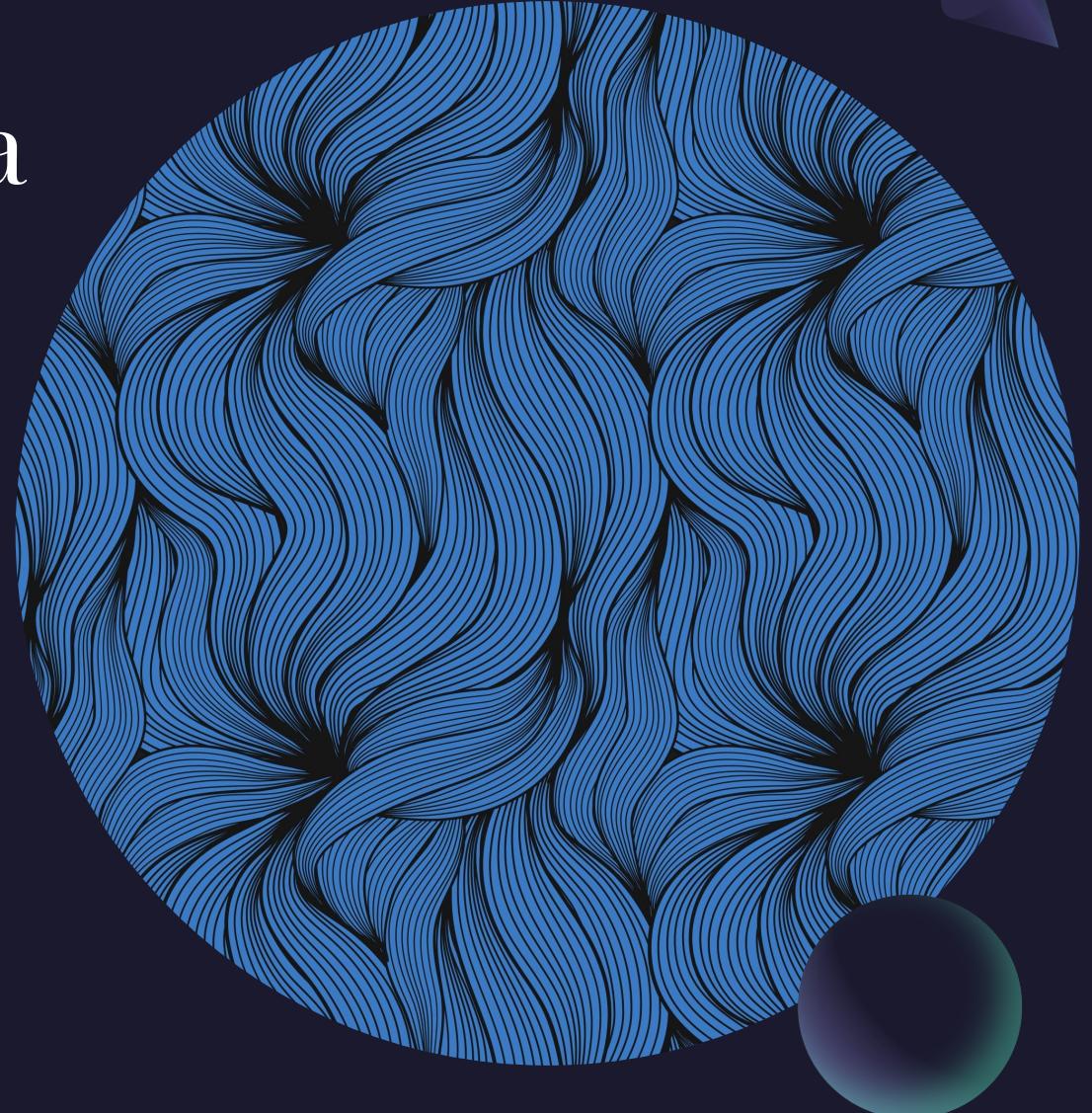
# Federated Data Governance in Action



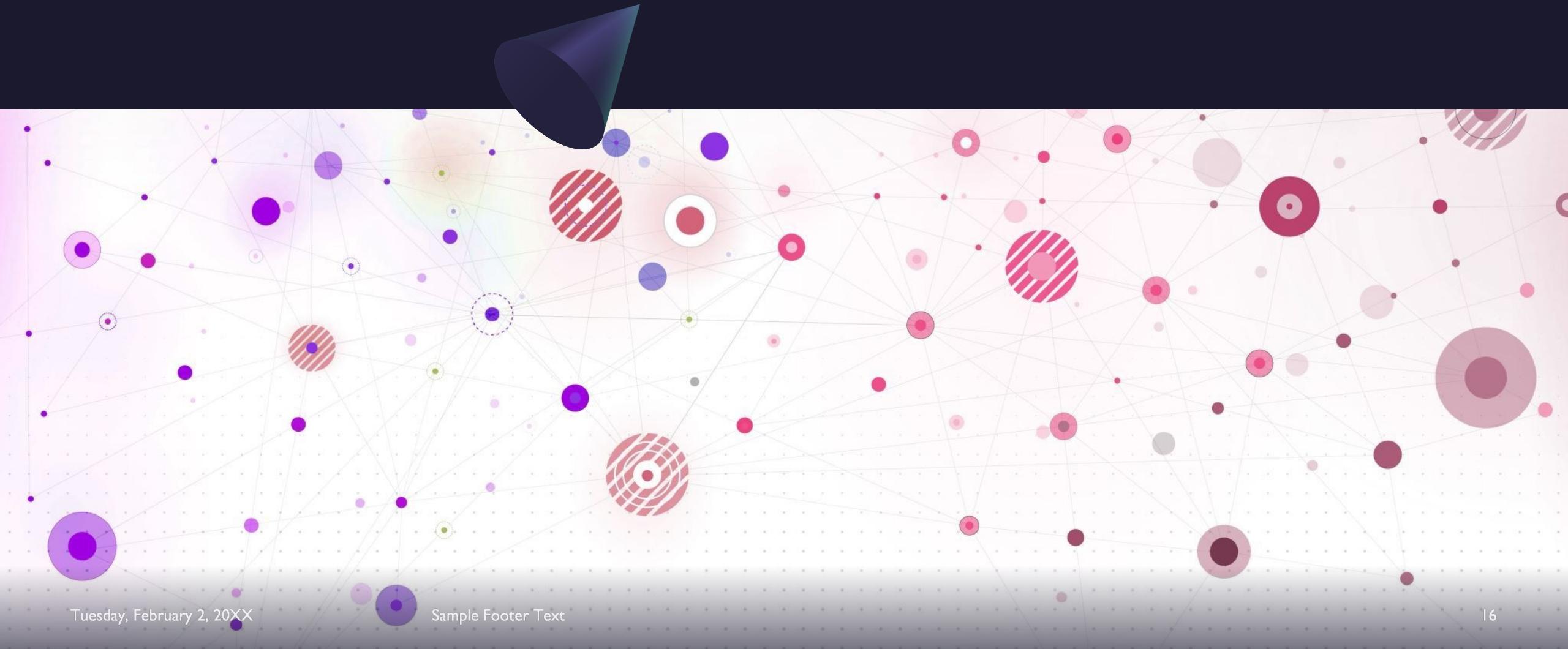


If you don't invest in  
building a data-driven  
culture, your data  
mesh efforts will fail.

To reap the rewards of data mesh, you must co-create governance with your cross-functional teams.



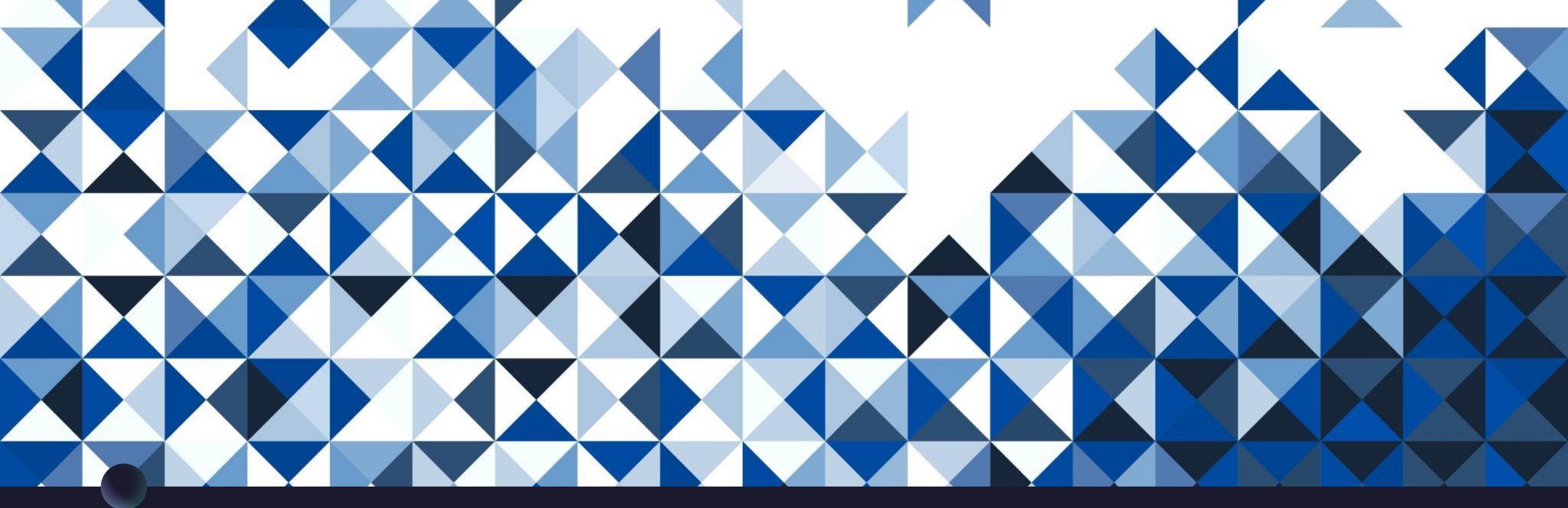
# Designing Data Governance from the Ground Up



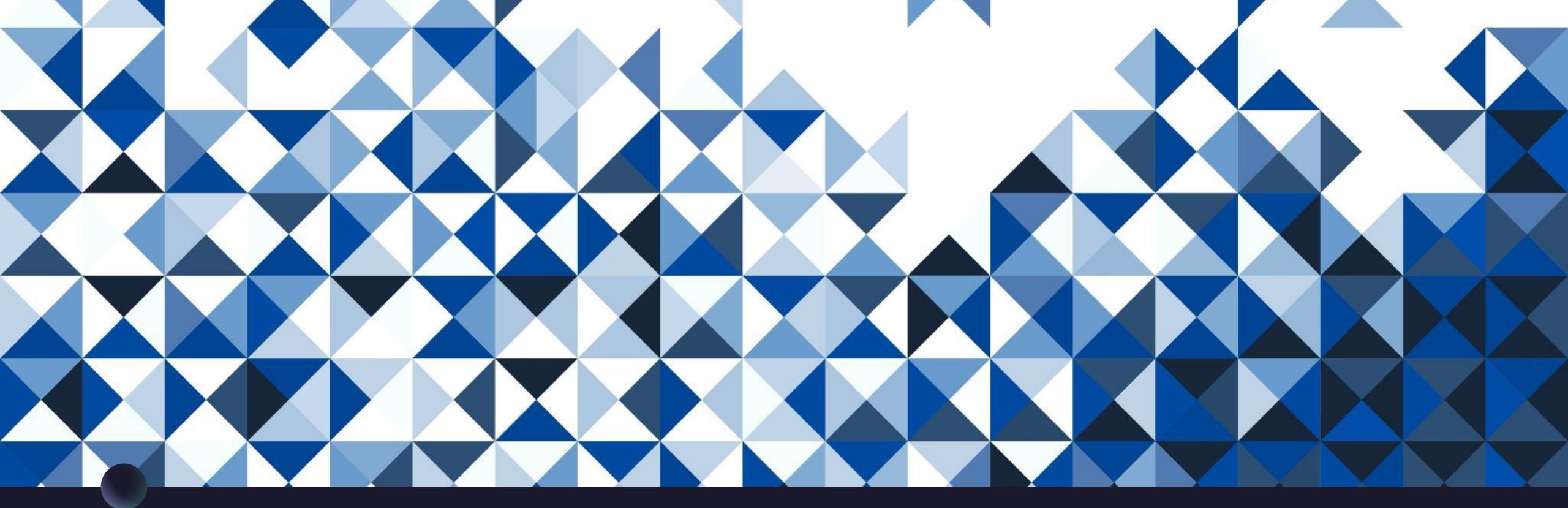
# Answer 2

## questions:

- *What is my company's mission statement?*
- *How does our data use fulfill that mission?*



# Define Data Domains



## To define data domains, answer:

1. What are your organization's key areas of interest?
2. Which problem spaces does your org address?
3. How is your org architected as-is?
4. Who leads each problem space and its associated data?

# Business Capability Map of Oceanic Airlines

## Customer services management

Online ticket management

Offline ticket management

Delay and resolution management

Customer management

Discount and loyalty management

Bookings & commissions

Advertising and marketing

Re and

## Staff and personnel management

Recruitment & employee management

Personnel management

Groundcrew personal management

Pilot management

Aerospace engineer management

Purser management

Baggage handling and lost items

Ra p

## Airflight management

Airplane maintenance

Fuel optimization

Flight optimization planning

Airport and lounges management

Engines and spares management

Flight plan and overview management

Aviation insurance management

La ma

## Supporting services management

Assets and financing

Cost management

IT services management

Car leasing and pick up services

Income and taxes management

Partnership and communication

Emission and fare trading management

Re pr ma

# Appoint domain-specific data stewards



**Will**

Customer



**Robin**

Product/Service



**Gabrielle**

Sales/Transactions



**Sergio**

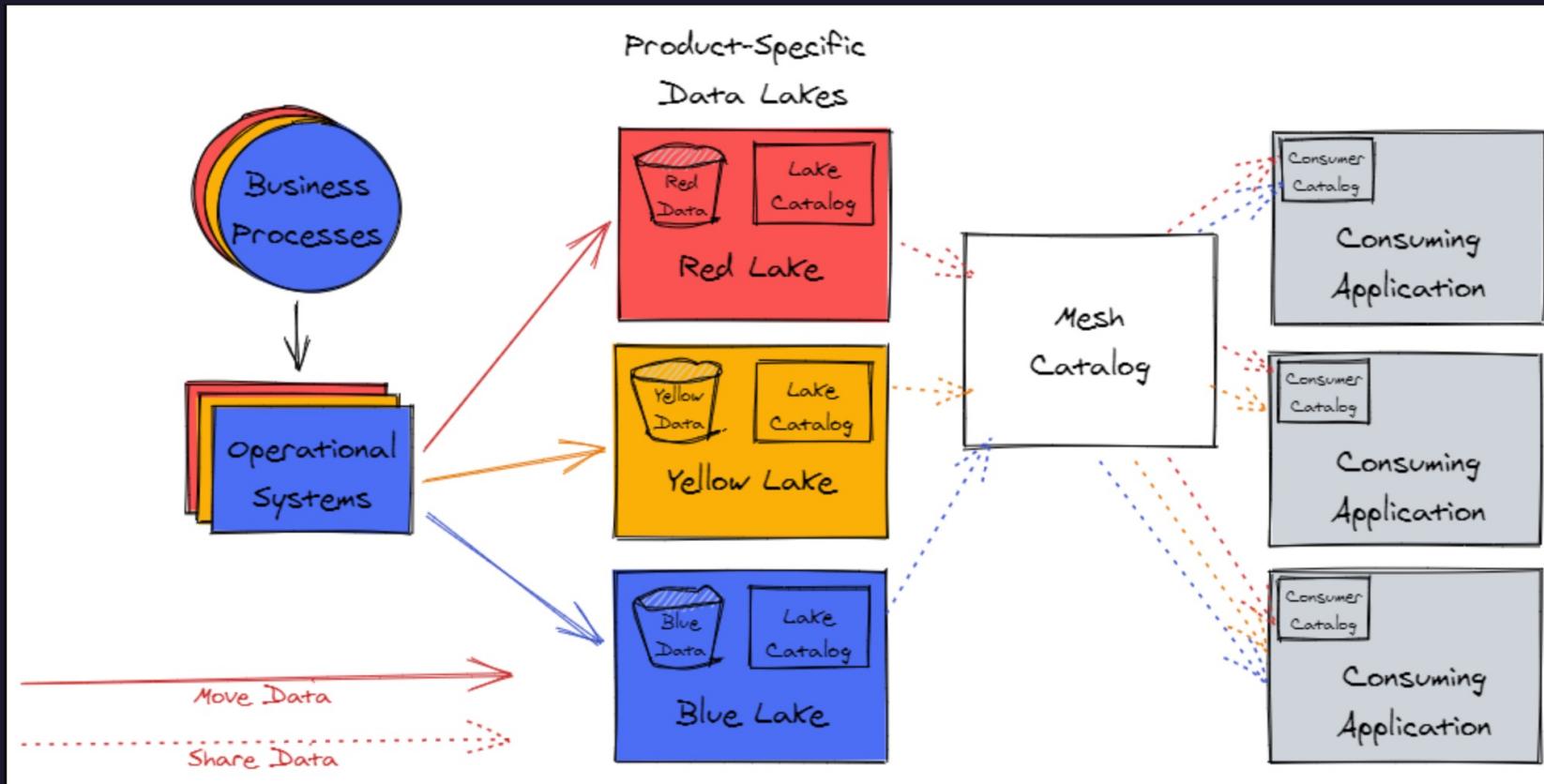
Legal

# Define Data Quality

Data Domain	Component Name	Goal	Success Conditions	Failed Conditions
Customer				
Product/Service				
Sales/Transactions				
Legal				

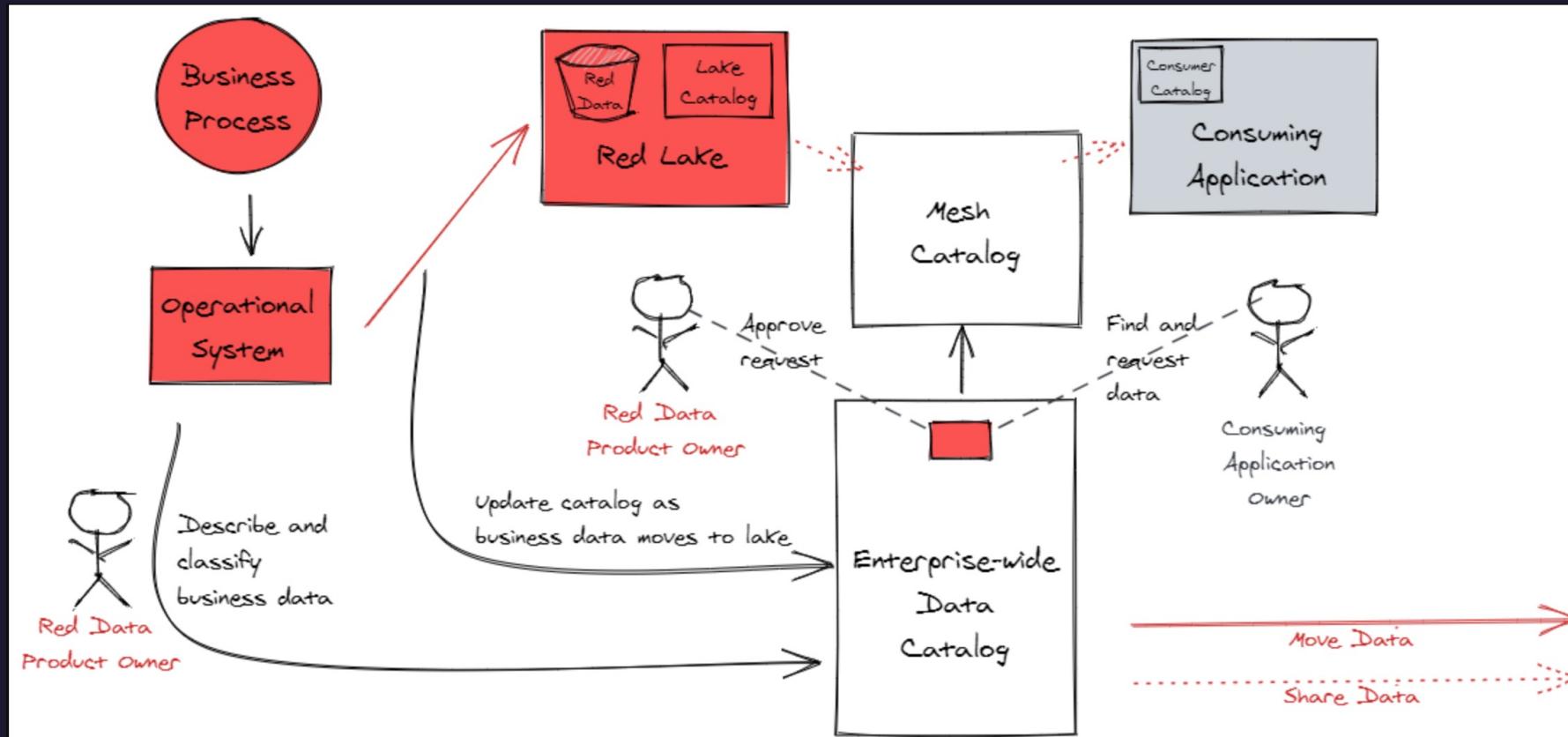


# Data Mesh at JP Morgan



Source: <https://aws.amazon.com/blogs/big-data/how-jpmorgan-chase-built-a-data-mesh-architecture-to-drive-significant-value-to-enhance-their-enterprise-data-platform/>

# Data Mesh at JP Morgan



Source: <https://aws.amazon.com/blogs/big-data/how-jpmorgan-chase-built-a-data-mesh-architecture-to-drive-significant-value-to-enhance-their-enterprise-data-platform/>

# Summary



Done well, data mesh lets teams access, develop, and manage data autonomously.

It also gives your data stewardship team an easier way to keep your data secure.

But there's a catch!

Data mesh works only if you've done the hard work to create a data-driven culture and automate that culture's standards.



# Thank You!

<https://pragprog.com/titles/lmmllops/designing-data-governance-from-the-ground-up/>

<https://www.linkedin.com/in/laurenmaffeo>

