Data Mesh: a lean perspective

Ned Letcher & David Colls
LAST Conference
Melbourne December 2021

/thoughtworks

Ned Letcher

Lead Data Science Engineer

@nletcher

@ned2

ned.sh

/thoughtworks

David Colls

Director, Data & AI

@davidcolls

@safetydave

safetydave.net

/thoughtworks

Today

Data mesh from a LAST perspective

First

Data production and its wastes

Overview of the modern analytical data factory, and common issues as waste

Next

What is data mesh?

Brief introduction to data mesh principles and building blocks

LAST

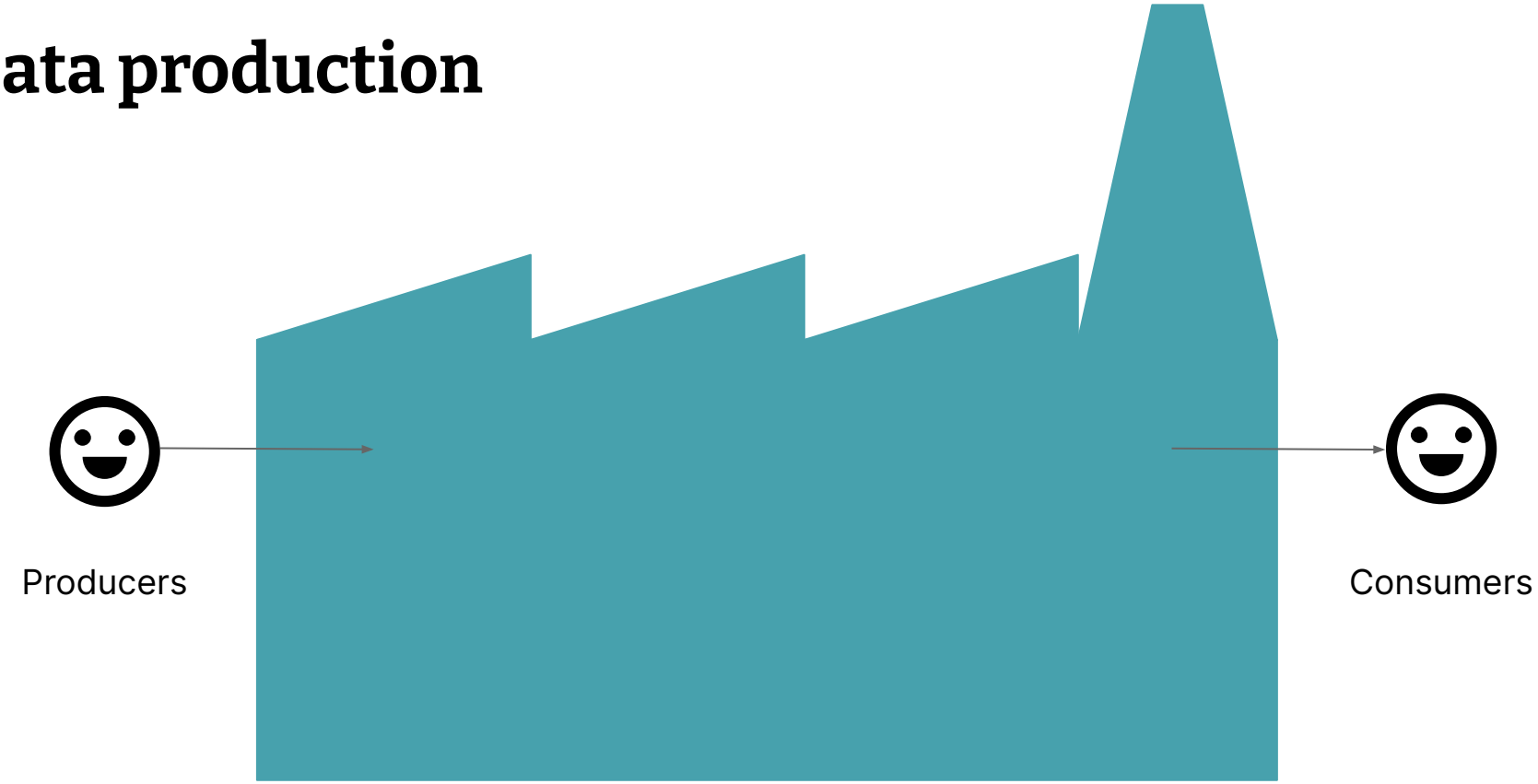
Lean thinking and data mesh

Discussion of a lean approach to analytical data manufacturing

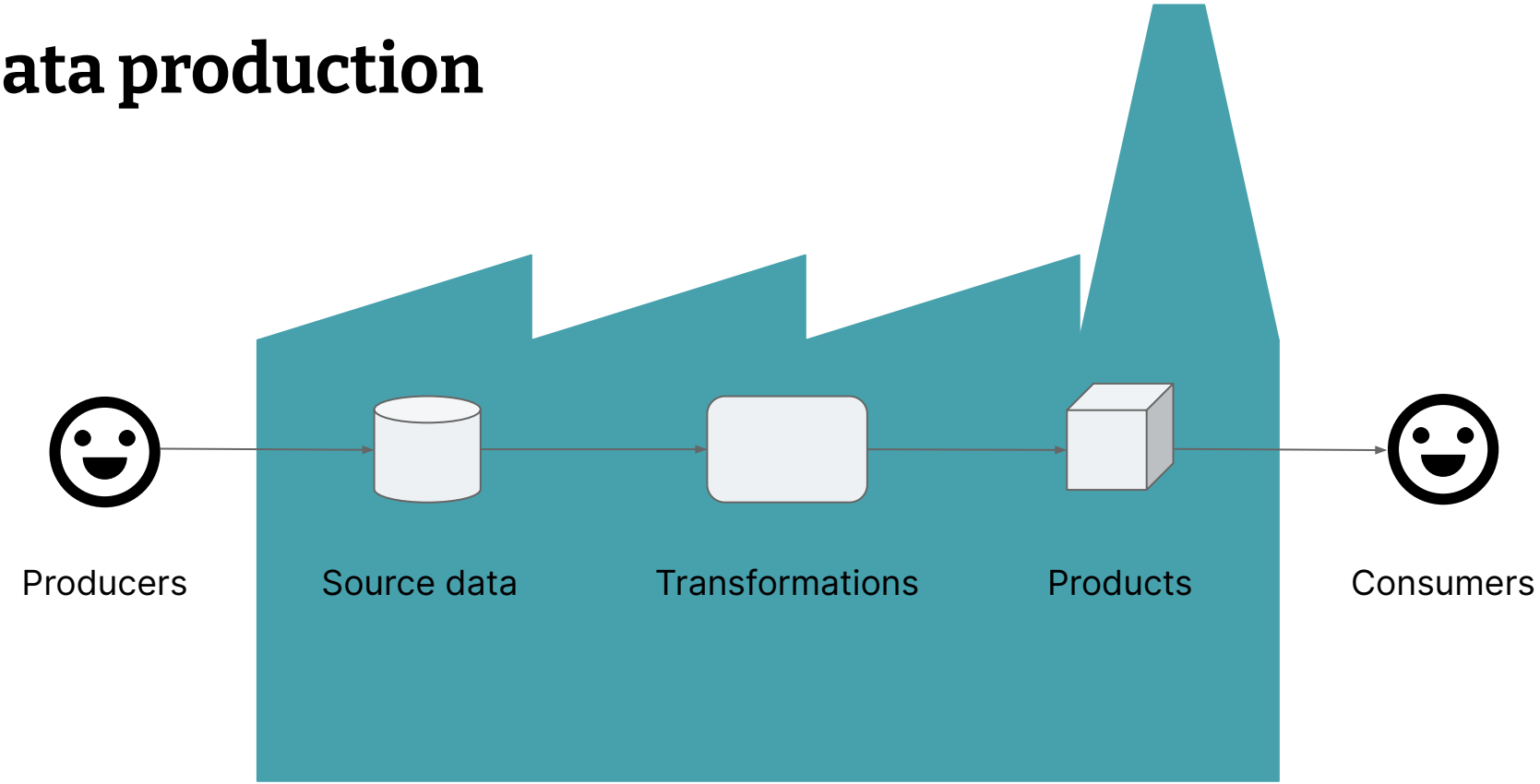
Data production and its wastes

/thoughtworks

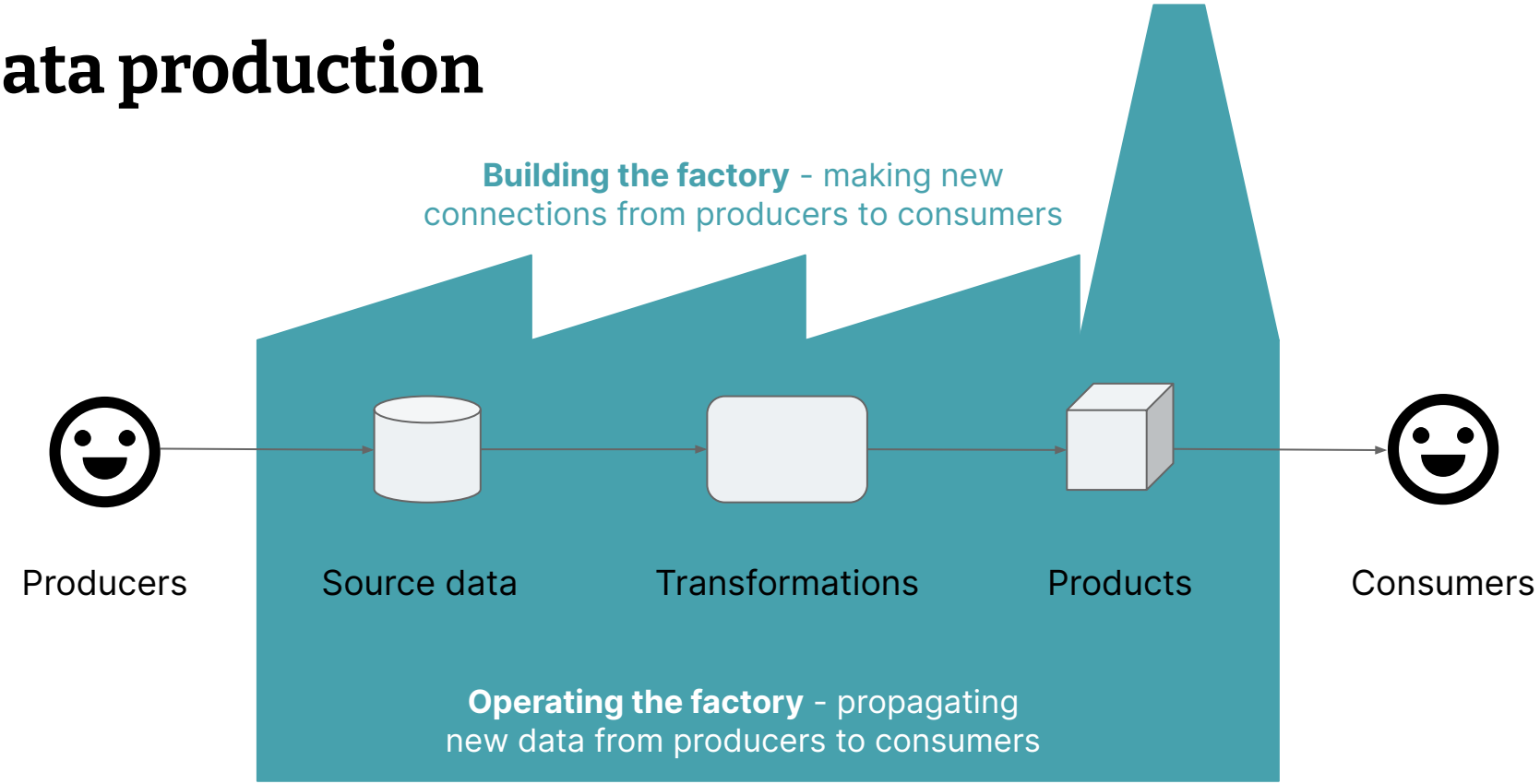
Data production



Data production



Data production



Models of waste

In data production

Build
Model waste like we do in knowledge work such as software development [1]

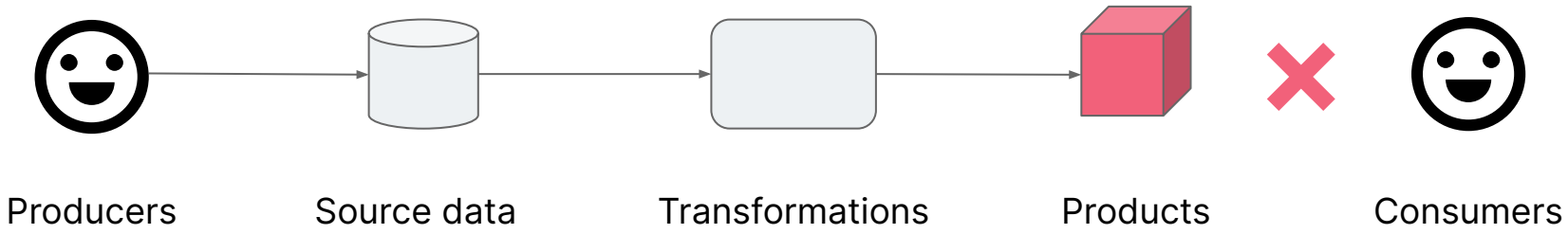
Operate
Model waste like we do manufacturing but with bits instead of atoms

1. e.g., the model due to Mary & Tom Poppendieck

Build

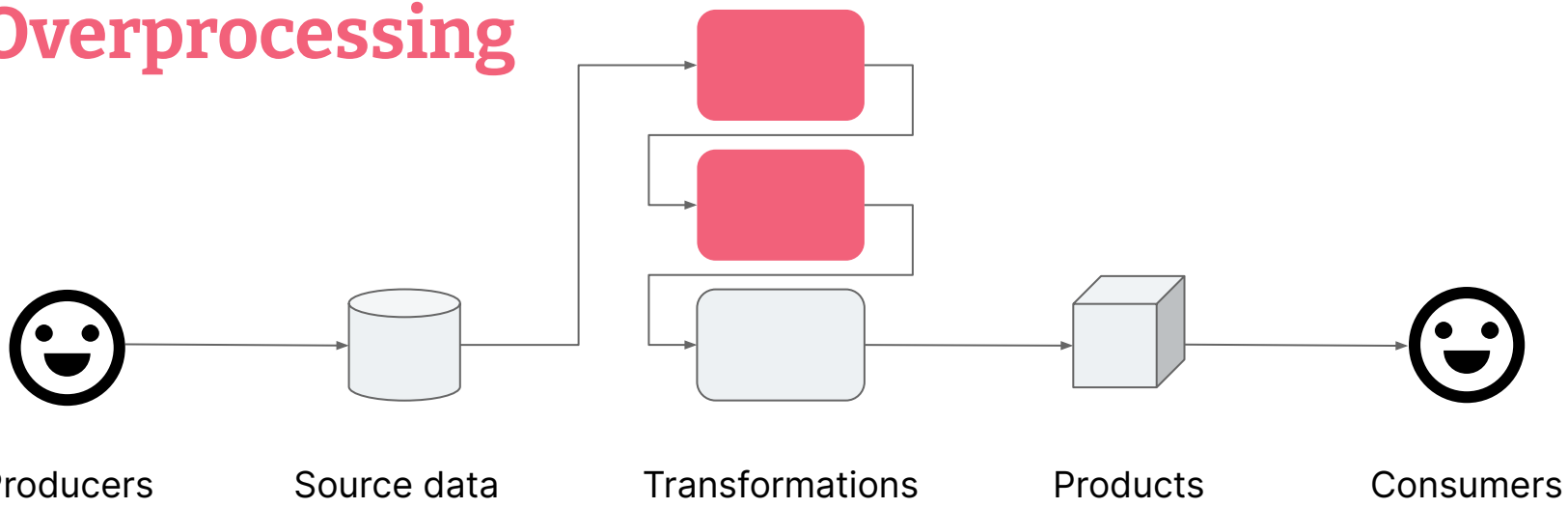
Operate

Waste of
1. Overproduction

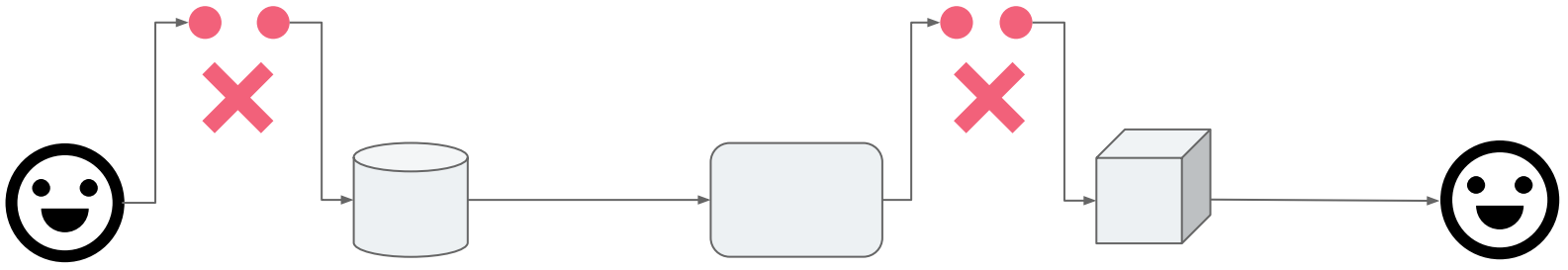


Waste of
2. Inventory

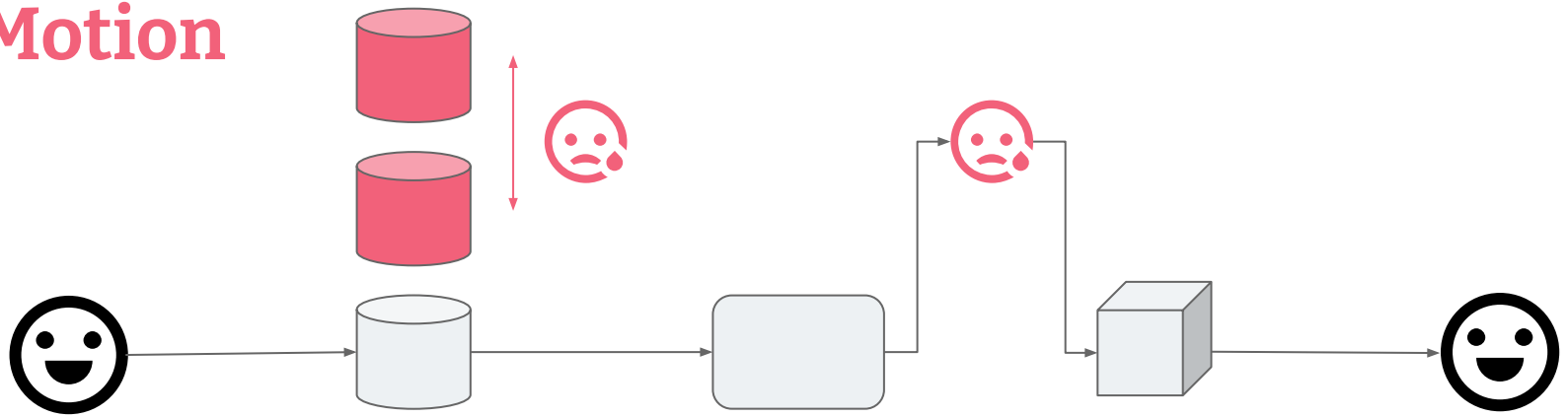
Waste of
3. Overprocessing



Waste of
4. Transportation

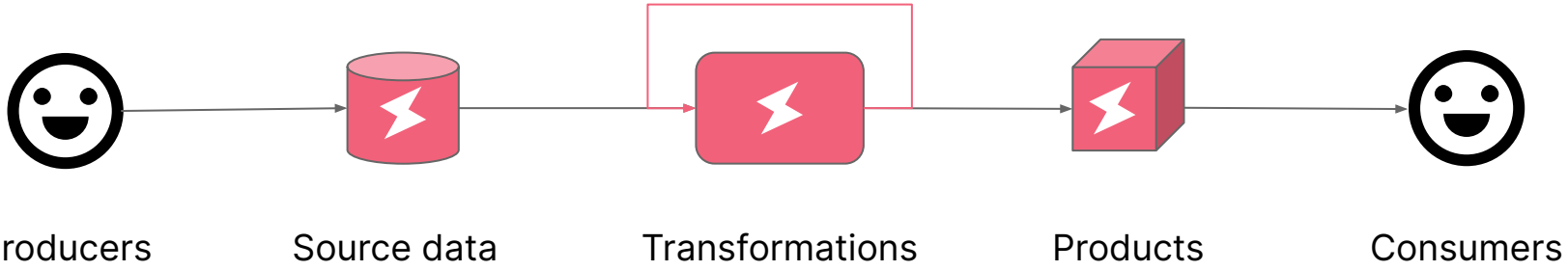


Waste of
5. Motion

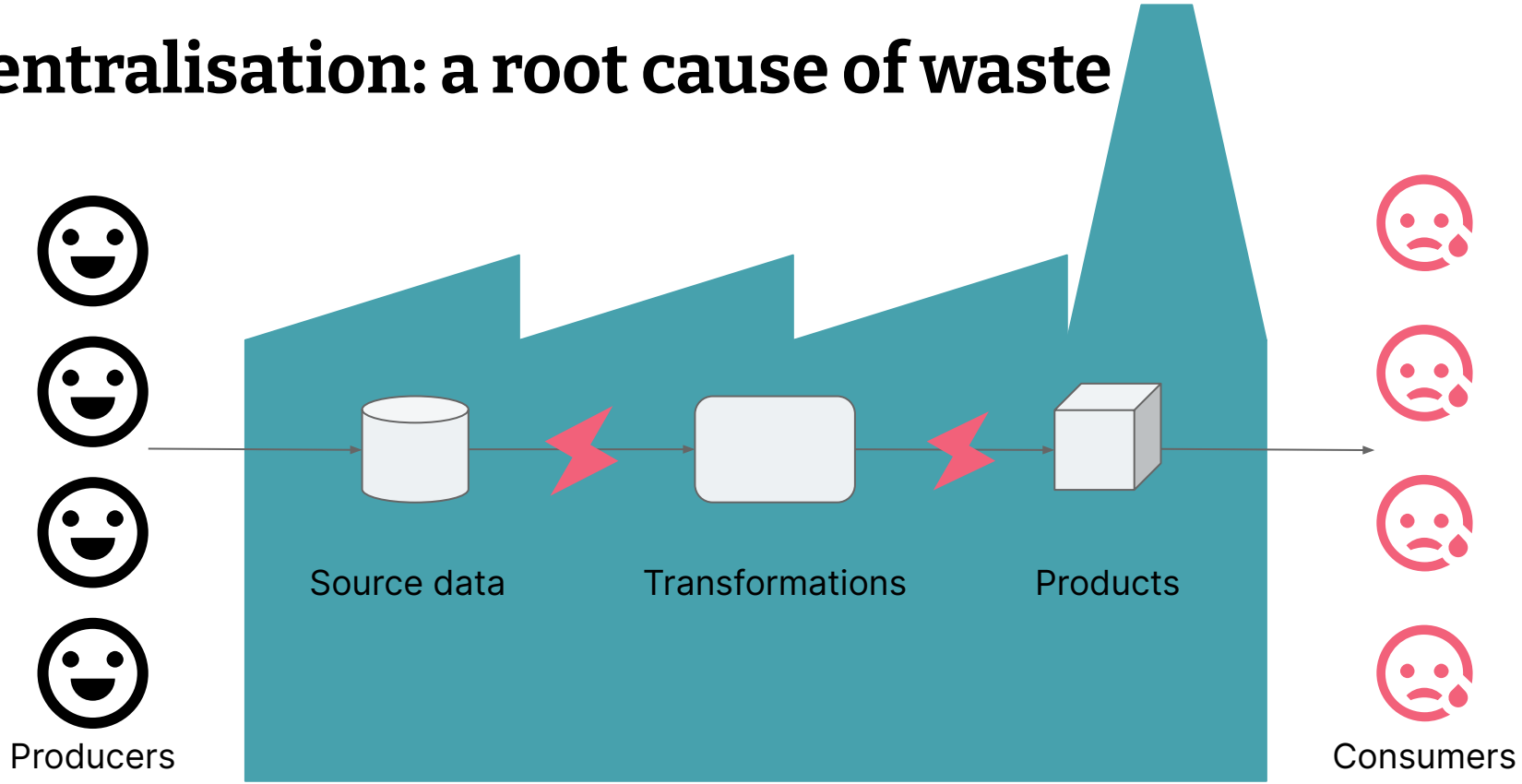


Waste of
6. Waiting

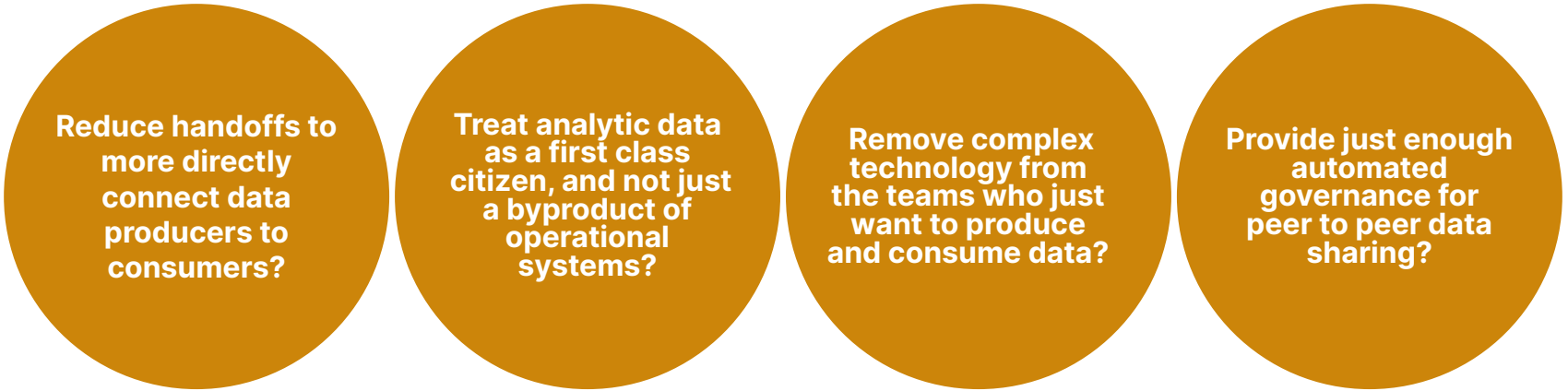
Waste of
7. Defects



Centralisation: a root cause of waste



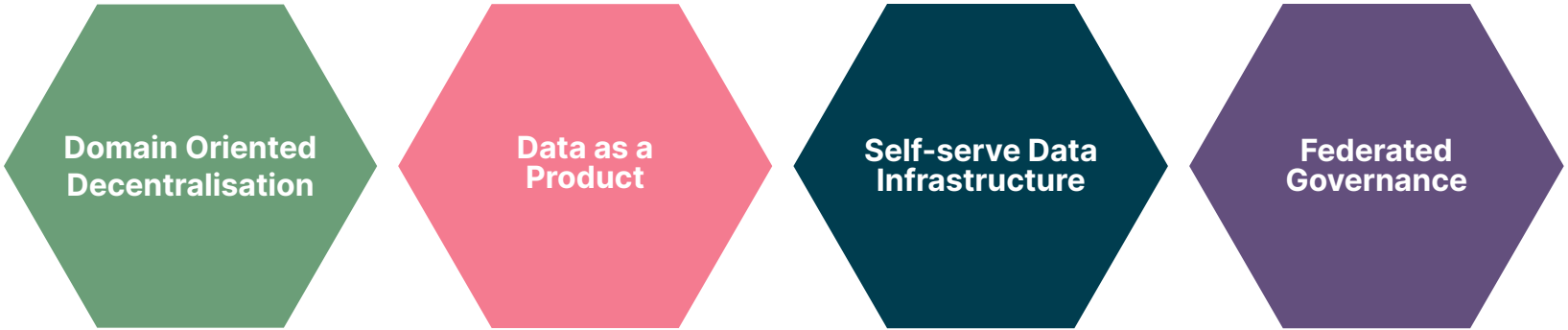
Instead of centralising data, what if we could...



What is data mesh?

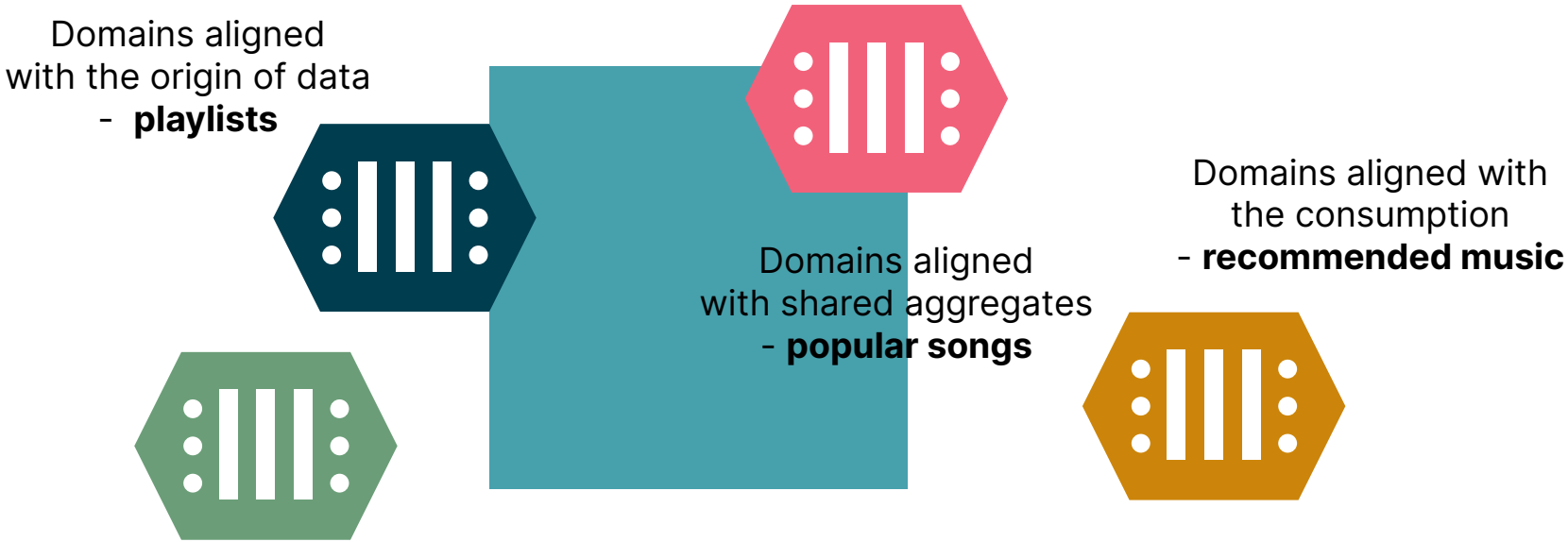
/thoughtworks

Principles of Data Mesh



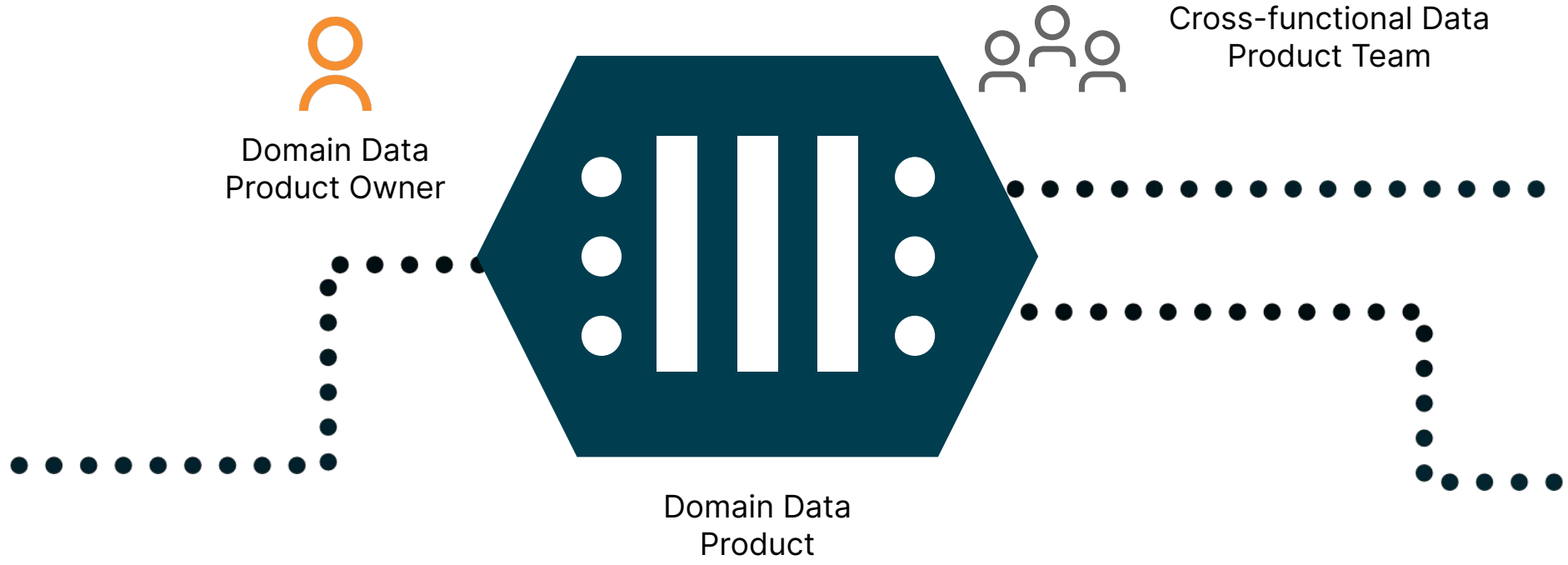
Decompose data around domains

Distribute the ownership



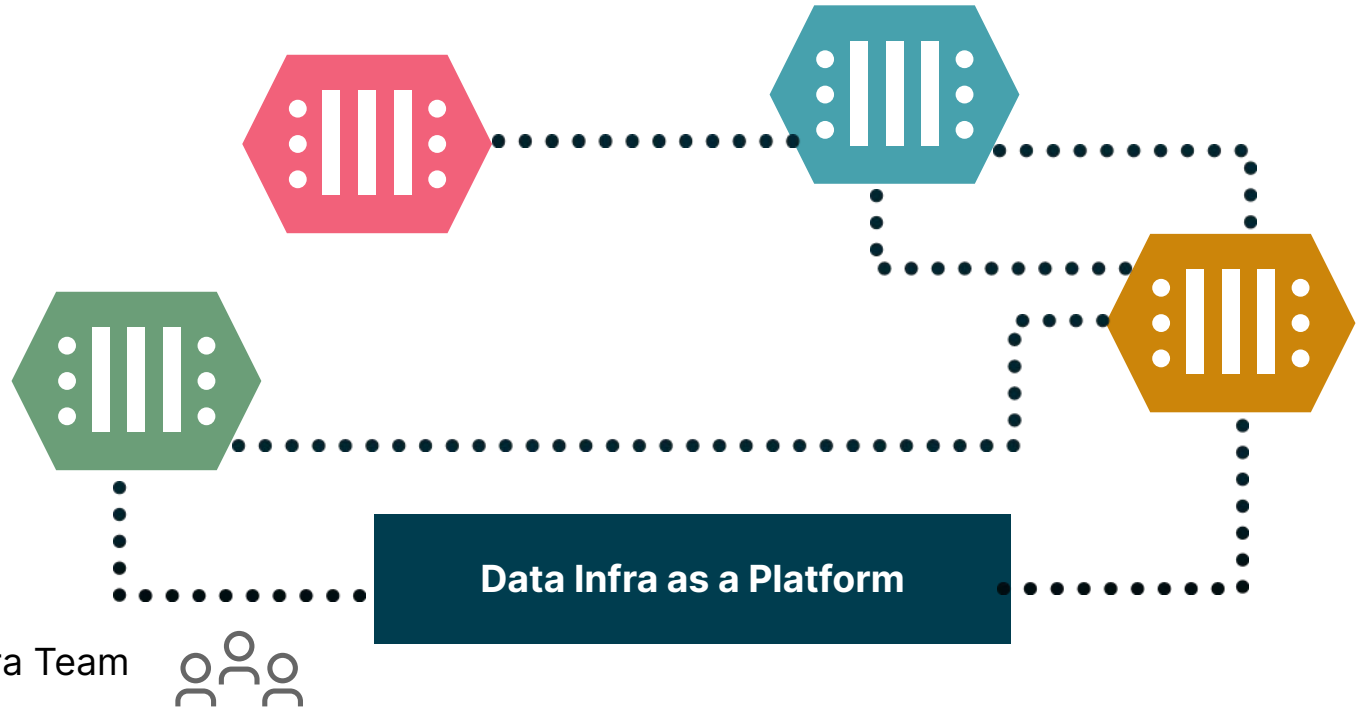
Serve data as a product

Delight the consumer with ease of data discovery and use



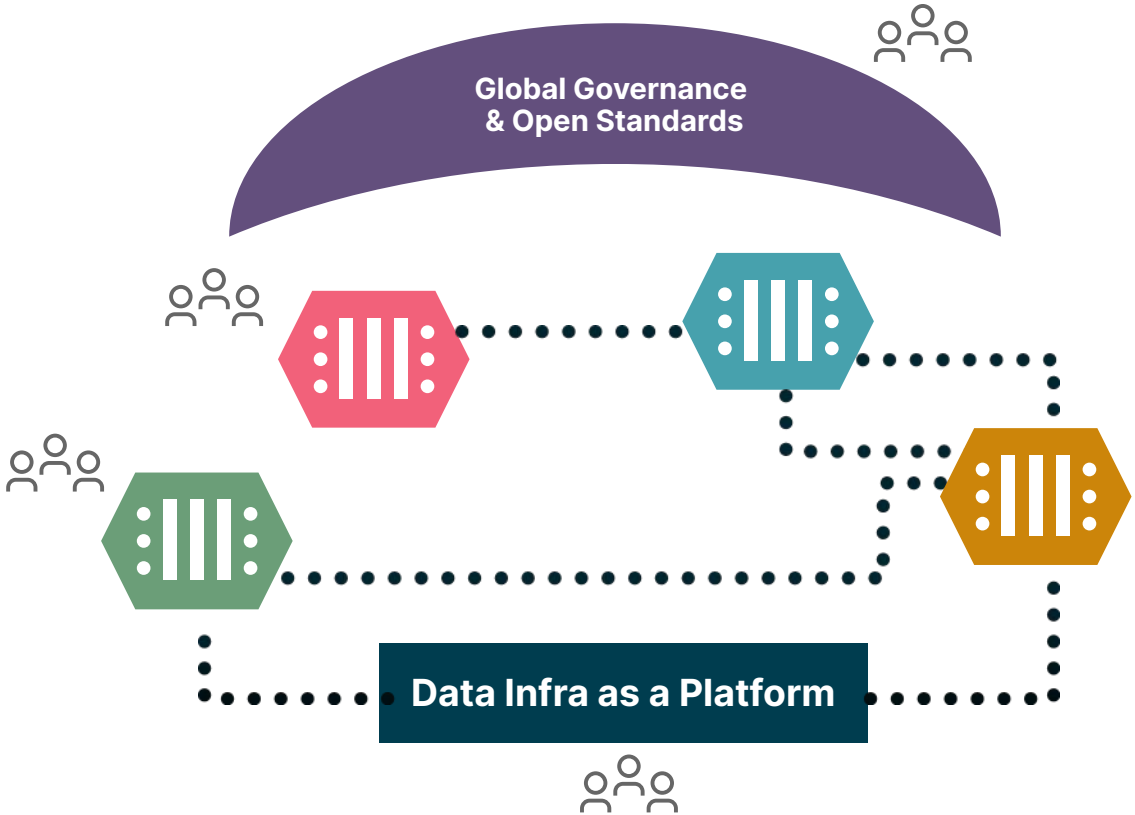
Self-serve platform

Abstract technical complexity in self-serve data infrastructure



Federated & global governance

Build an ecosystem that enables

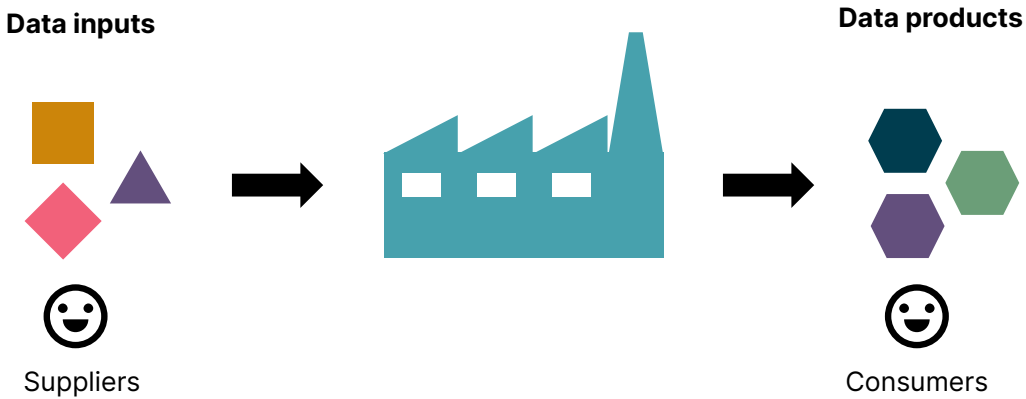


Lean Thinking and Data Mesh

/thoughtworks

Analytics as Manufacturing

Applying Lean Principles to Data Management

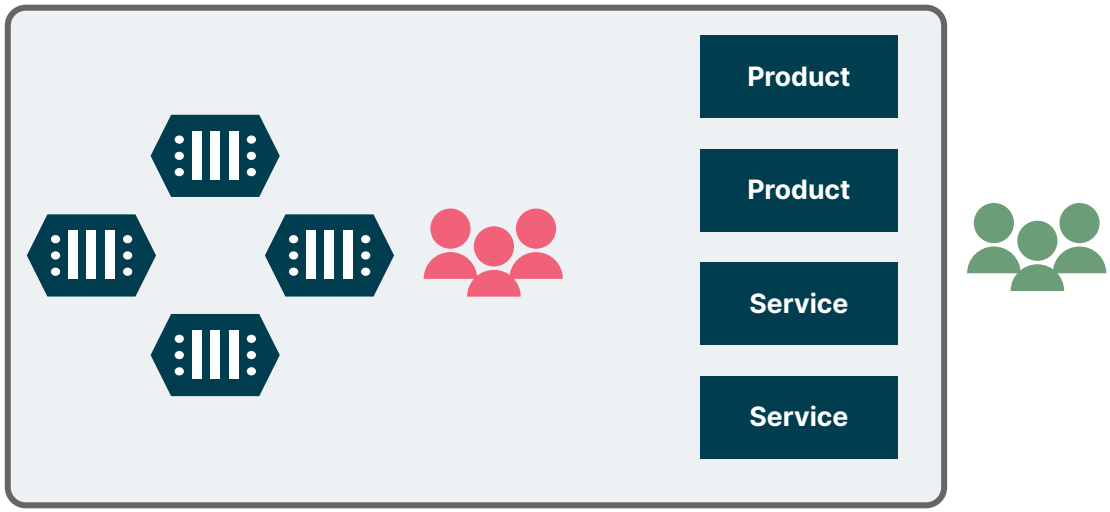


1. Define Value
2. Map the value stream
3. Create Flow
4. Establish pull
5. Continuously Improve

Lean Thinking, James P Womack & Daniel T Jones

1. Define Value

Apply product thinking to analytical data



~~*Data as asset*~~ → *Data as product*

Value is created by the producer and can only be defined by the ultimate customer.

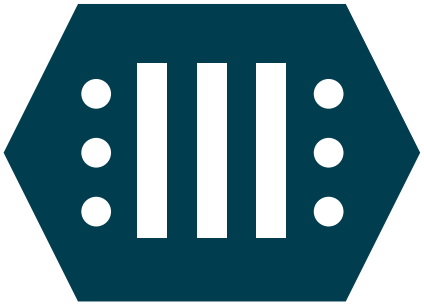
Data as Product

Treat internal data consumers as customers with different needs and personas.

1. Define Value

Apply product thinking to analytical data

- Usable
- Discoverable
- Understandable
- Trusted
- Natively Accessible

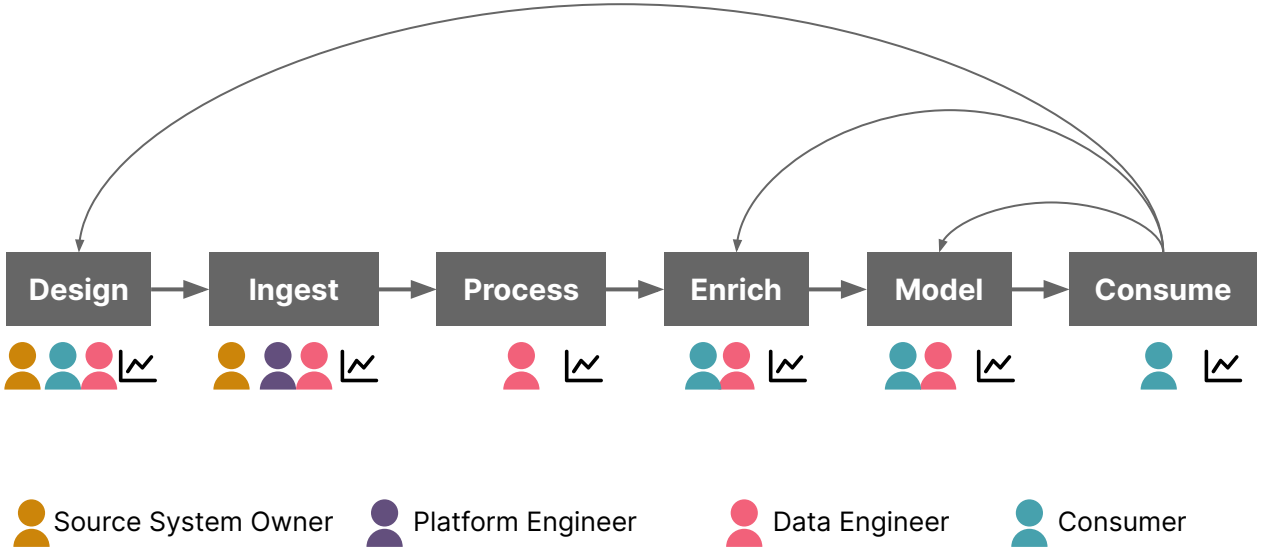


User personas

Cross-functional Data Product Roles

- Data product developers
- Domain experts
- Data product owner
- Experience Designer

2. Map the Value Stream



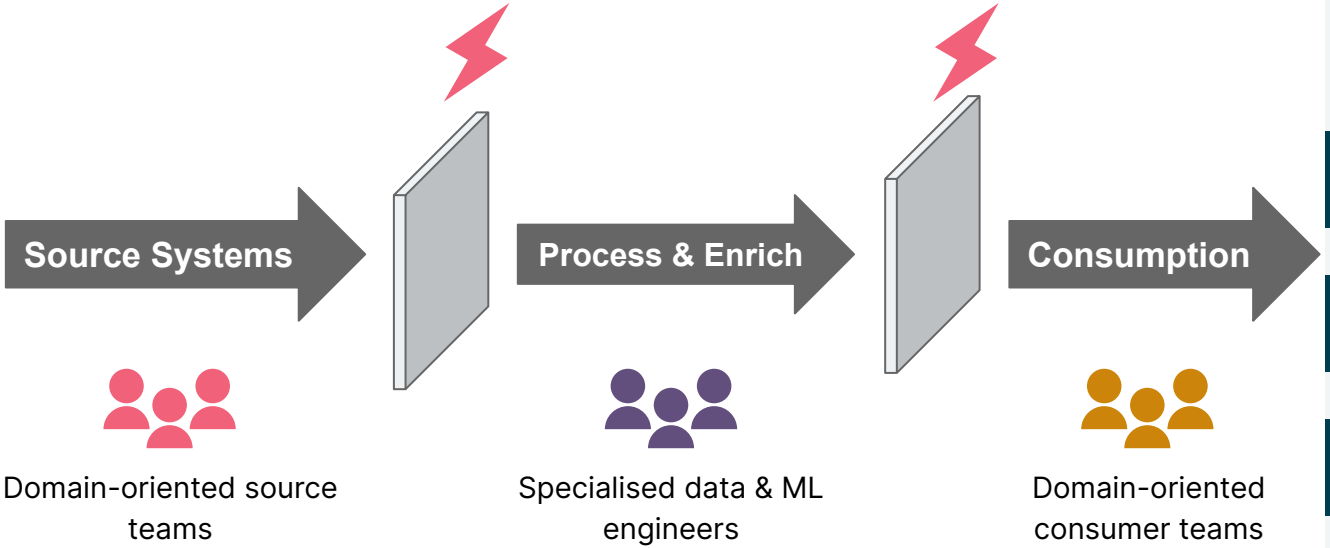
Identify steps required to produce value to consumers

For each step, identify:

- **The purpose**
- **Inputs and outputs**
- **Roles involved**
- **Metrics that indicate performance**

3. Create Flow

Barriers to flow in centralised data platforms



Technological decomposition leads to batch-and-queue inefficiencies

Increased cycle times due to hand-offs

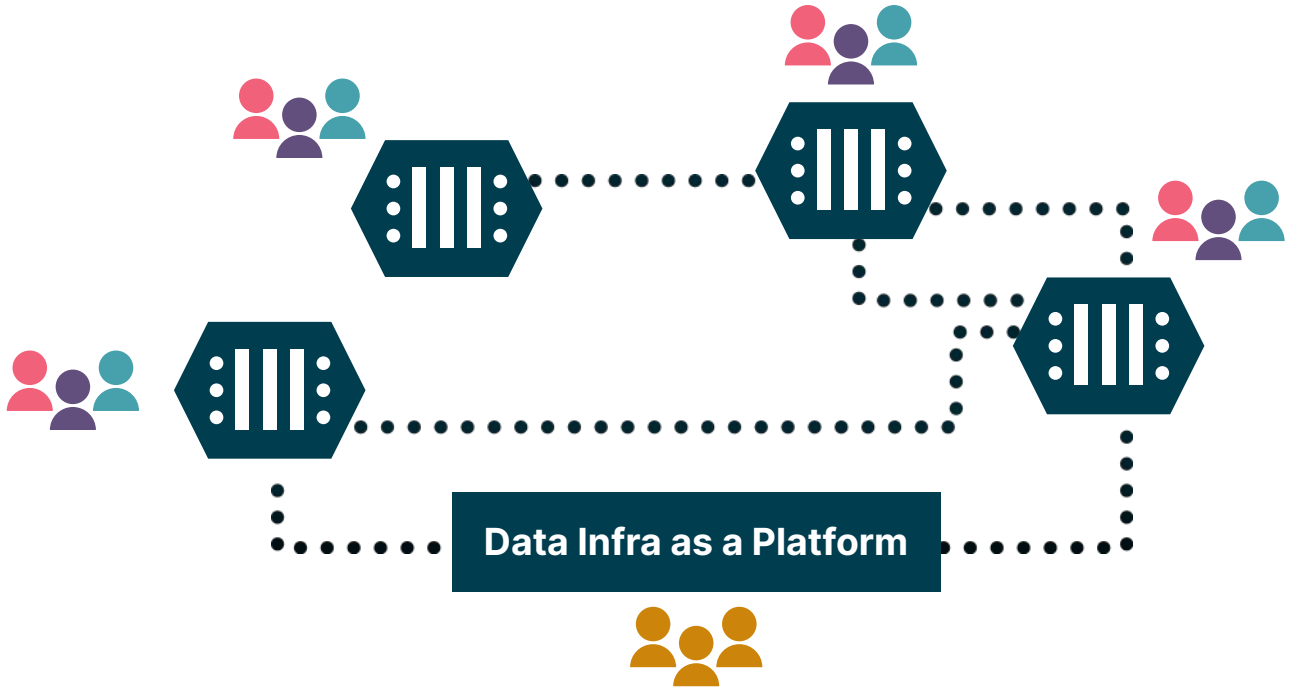
Loss of context across teams

Lack of visibility into upstream defects

Overproduction

3. Create Flow

Domain-oriented decentralisation

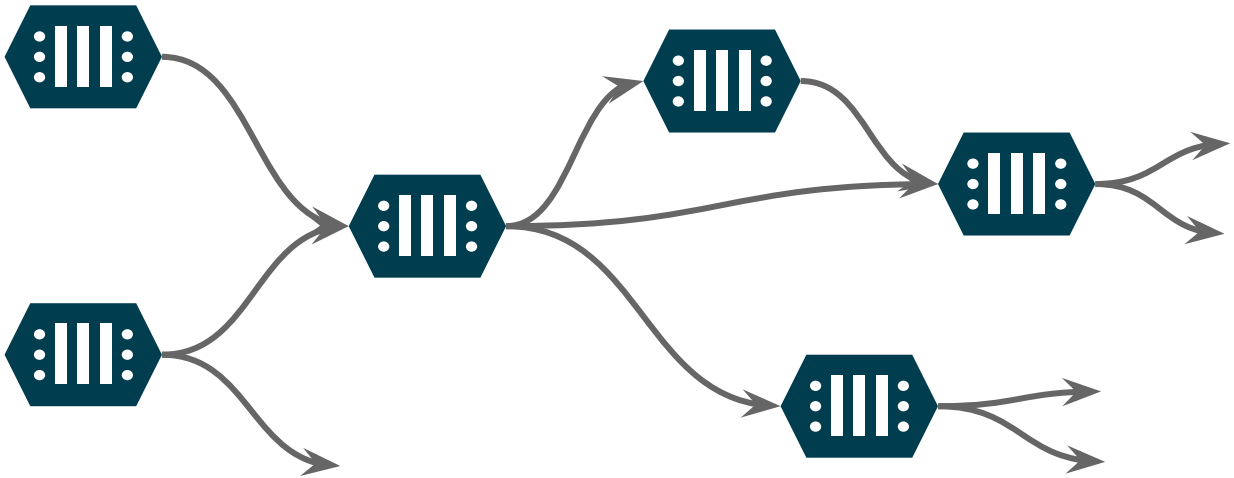


Things work better when you focus on the product and its needs, rather than the organisation of its equipment.

James P Womack, Daniel T Jones, *Lean Thinking*.

4. Establish Pull

Be Value Driven, not Data Driven



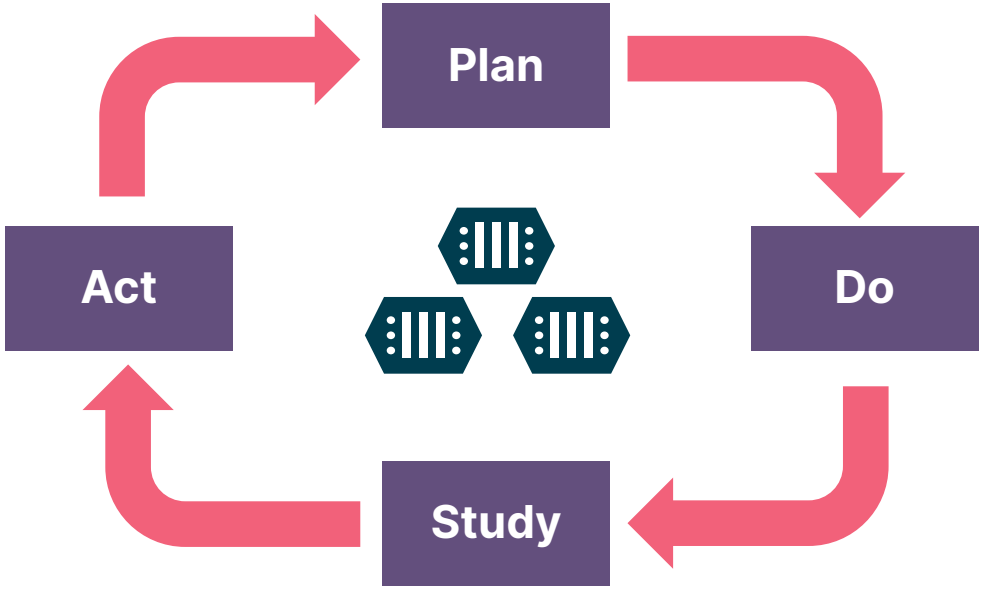
Data Flow

Consumers *pull* data from producers

Value Flow

New data products & features created only when known consumers & use-cases

5. Continuously Improve

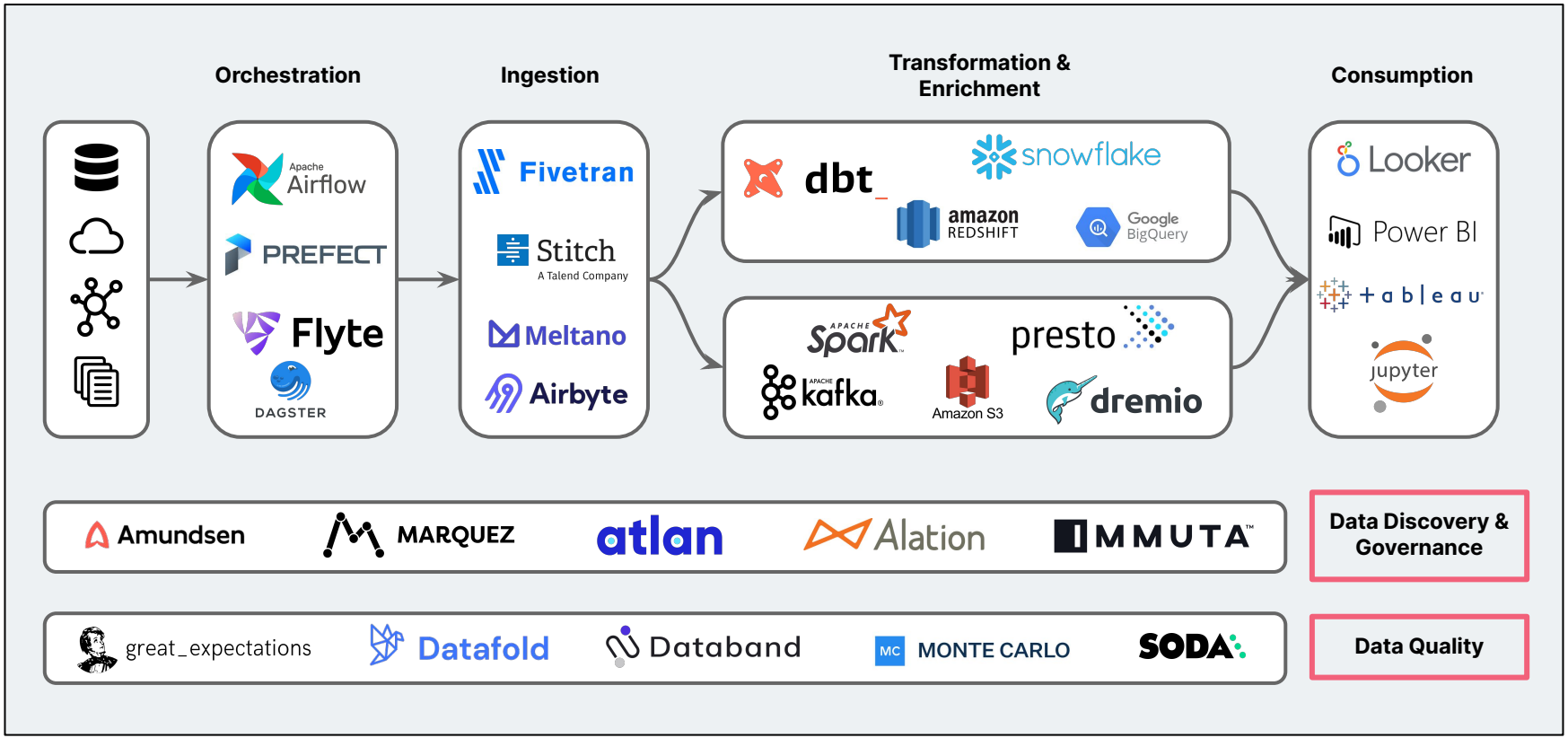


Define value-based metrics

Enable monitoring & observability

Adopt data product SLOs

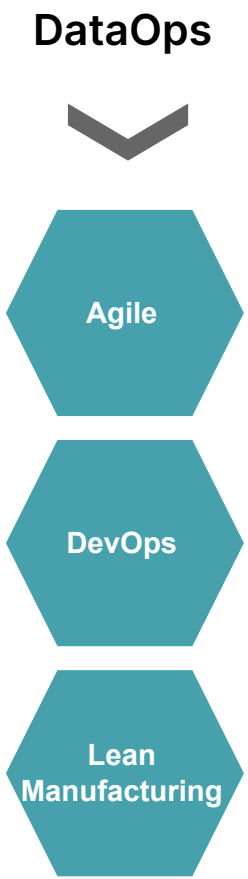
You can't buy information quality



The DataOps Manifesto

Lean in Data Management

1	Continually satisfy your customer
2	Value working analytics
3	Embrace change
4	It's a team sport
5	Daily interactions
6	Self-organize
7	Reduce heroism
8	Reflect
9	Analytics is code
10	Orchestrate
11	Make it reproducible
12	Disposable environments
13	Simplicity
14	Analytics is manufacturing
15	Quality is paramount
16	Monitor quality and performance
17	Reuse
18	Improve cycle times



https://dataopsmanifesto.org
https://datakitchen.io/the-dataops-cookbook

Thanks

/thoughtworks