

Data mesh and domain ownership

/thoughtworks

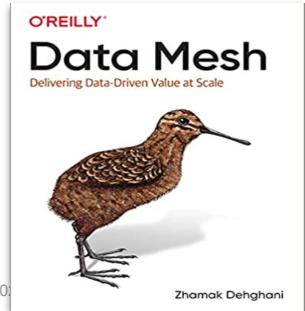


Speakers



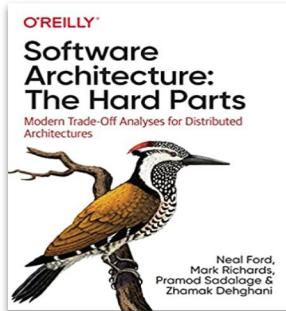
Zhamak Dehghani

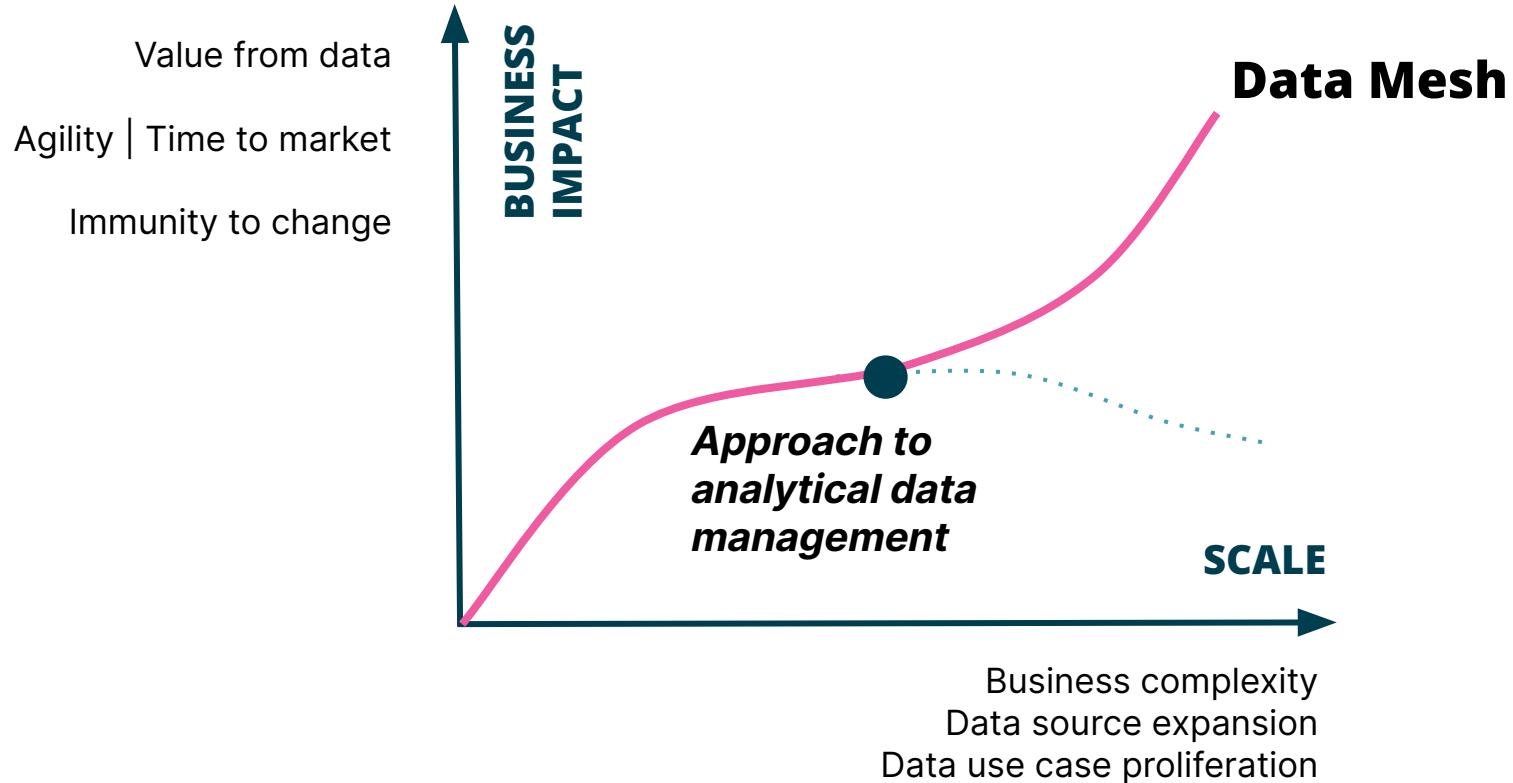
Director of Emerging Technologies,
Data Mesh Founder
Thoughtworks, North America



Danilo Sato

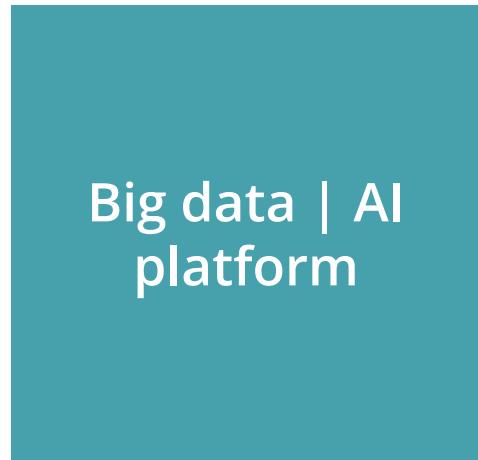
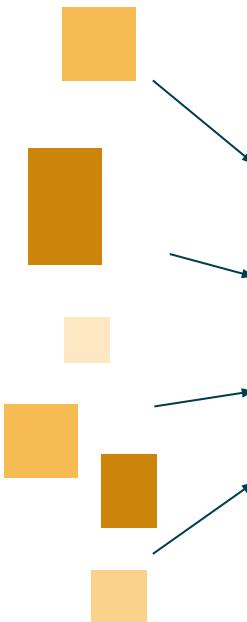
Head of Data & AI Services,
Thoughtworks UK and Europe



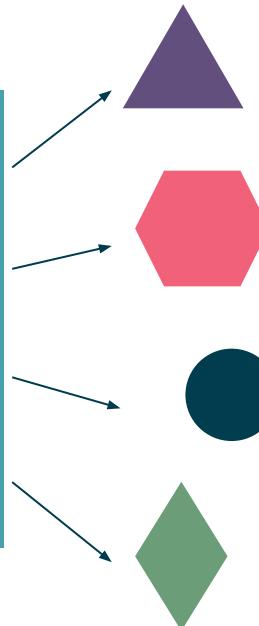


Centralized architecture

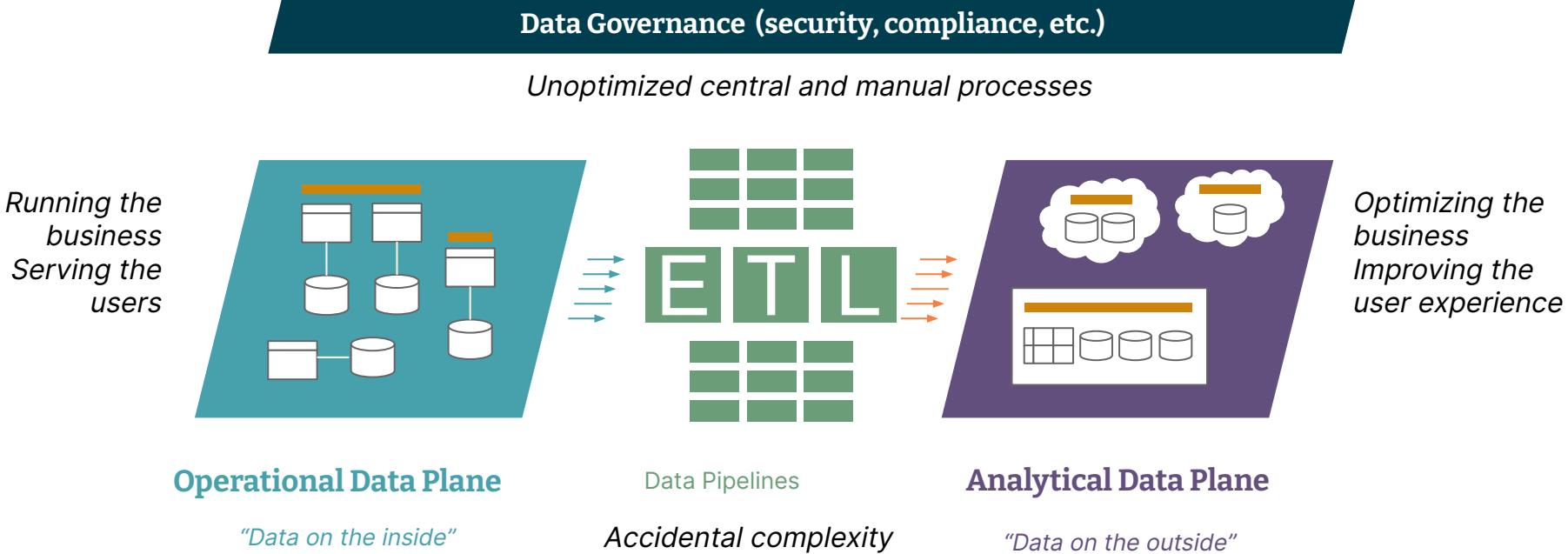
Ubiquitous Data



Innovation Agenda



Data Architecture Today



Zhamak Dehghani : @zhamakd : 2021

Data Mesh

A **decentralized sociotechnical** approach in managing and accessing **analytical data at scale**.



Data Mesh Desired State

Embedded partnership with AI and Data

Cross-functional team

Playlist domain

Data curious culture

Player domain

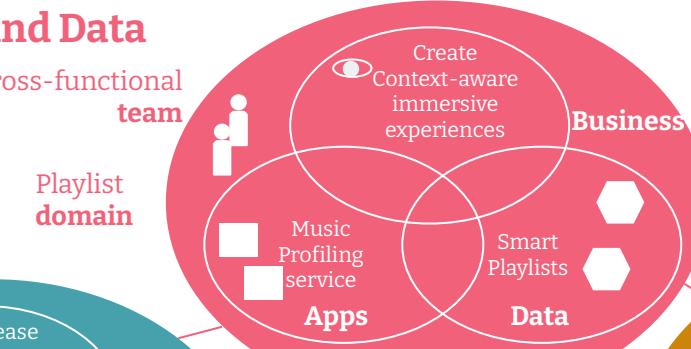
Increase Immersive experience with digital touchpoints

Player apps

Player Sessions

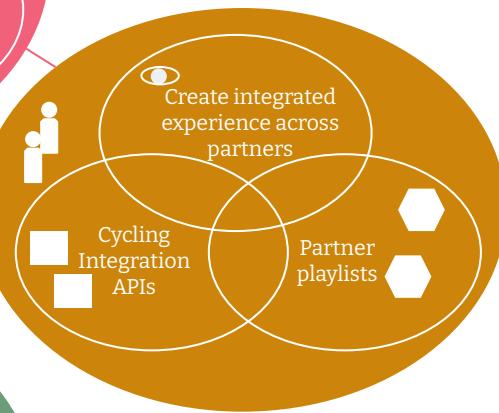
Player events

Invisible platform and policies



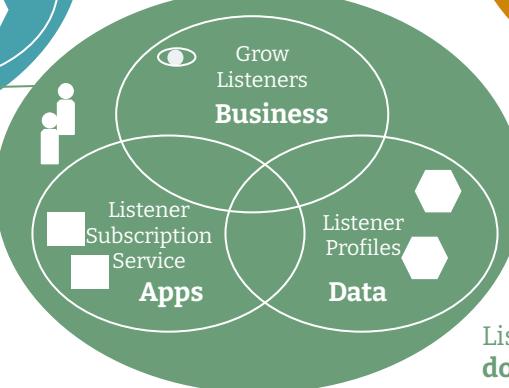
Get value through a data sharing network effect

Partner domain

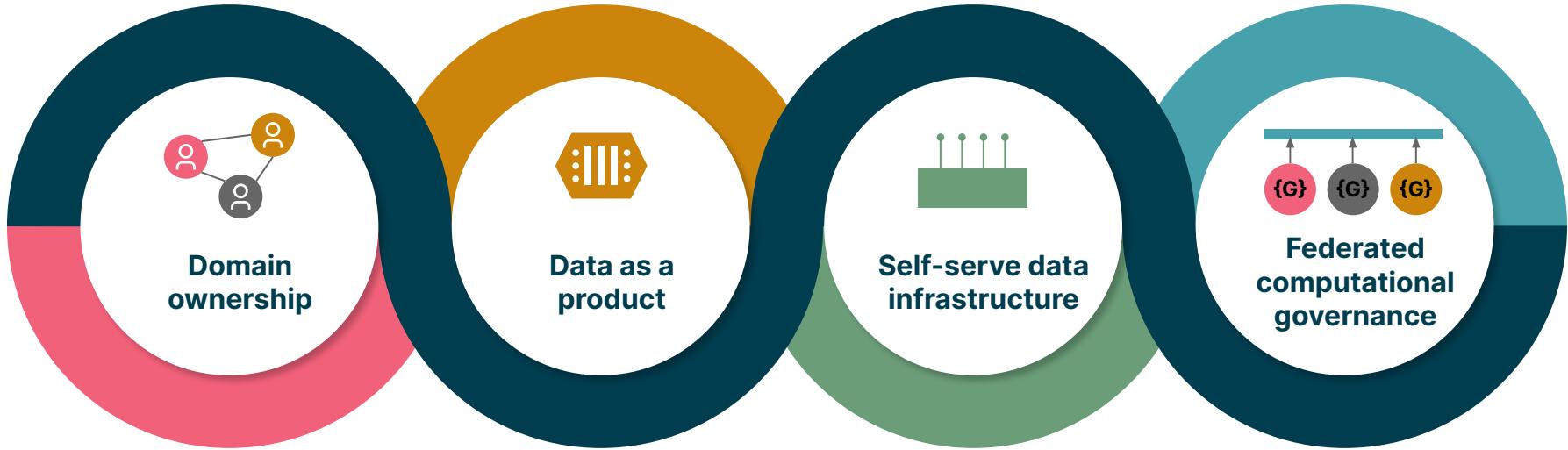


Limitless scale with autonomous data products

Listener domain

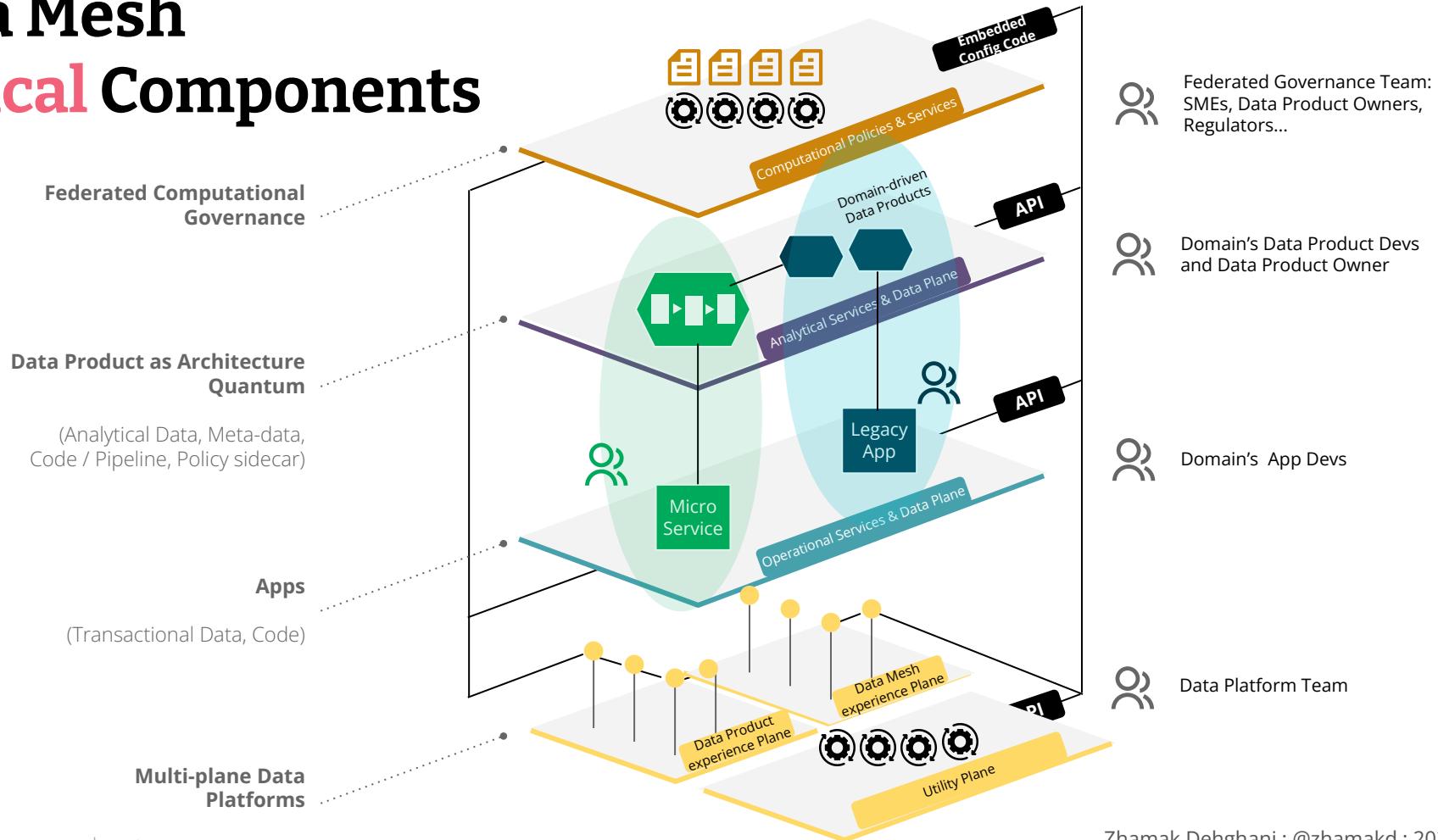


Principles of Data Mesh



Data Mesh

Logical Components



Domain Driven Design

/thoughtworks



Domain Driven Design

Tackling complexity in the heart of software

A software development approach in which we:

- Focus on the *core domain*
- Explore models in a creative collaboration of domain practitioners and software practitioners
- Speak a *ubiquitous language* within an explicitly *bounded context*



Ubiquitous
Language



Strategic
Design



Bounded
Context



Context
Maps

How to identify and define domains?

/thoughtworks

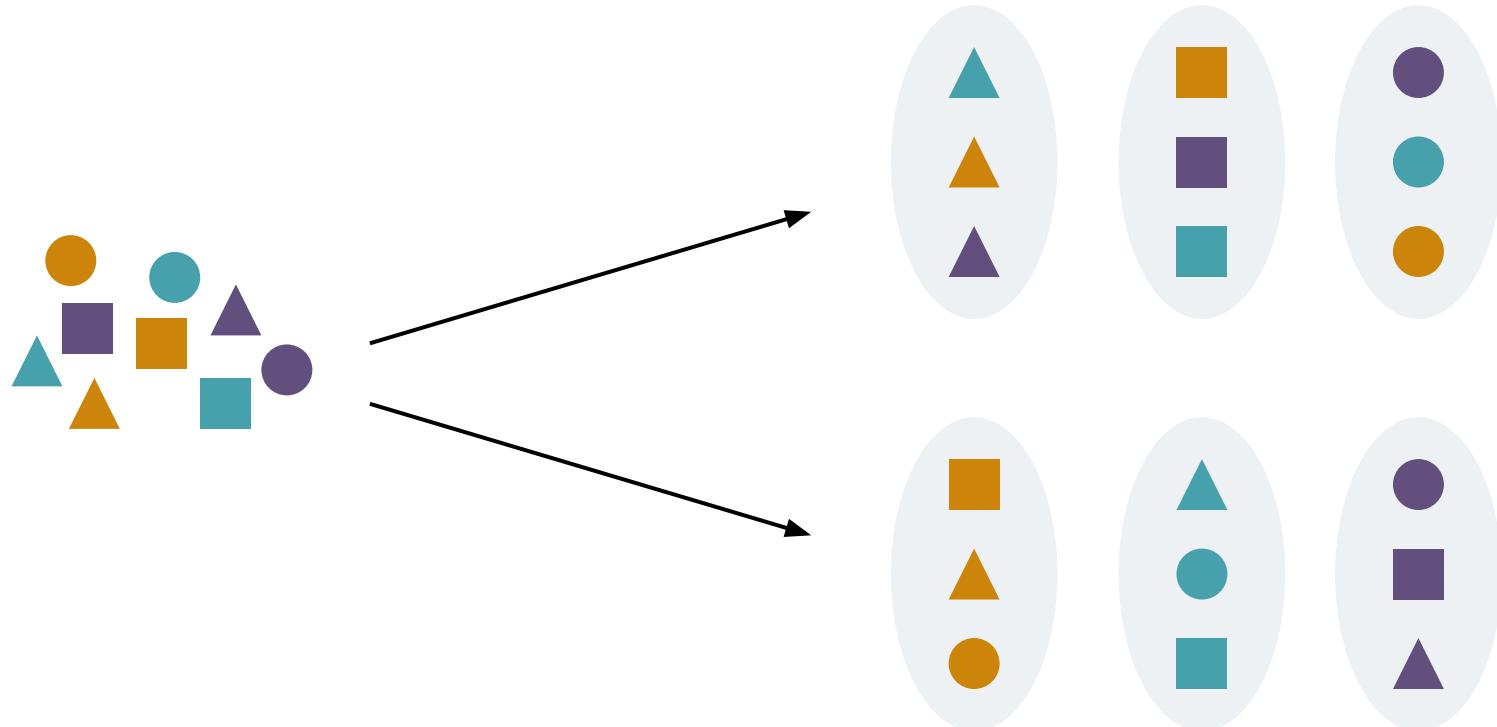




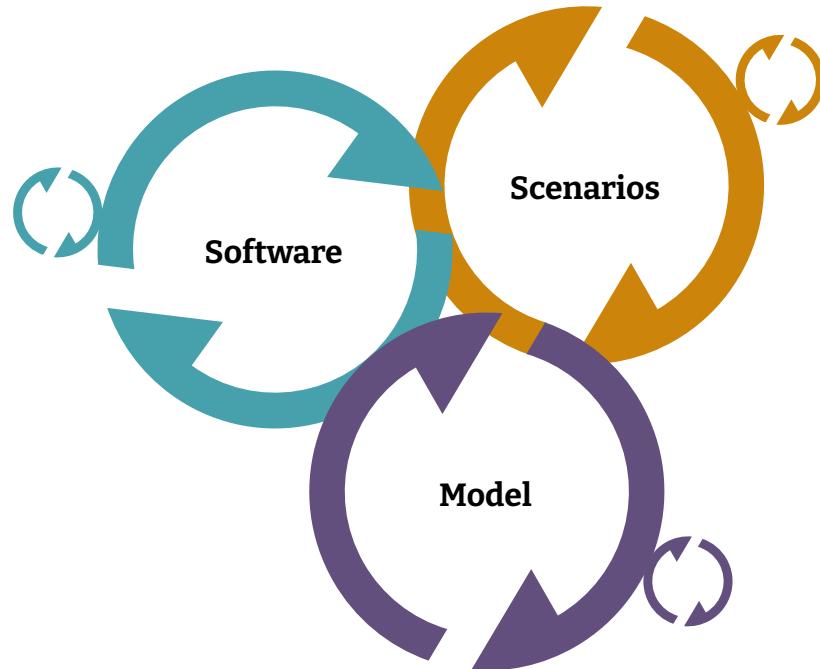
**“All models are
wrong, but some
are useful.”**

George E. P. Box

No single answer, different ways to model

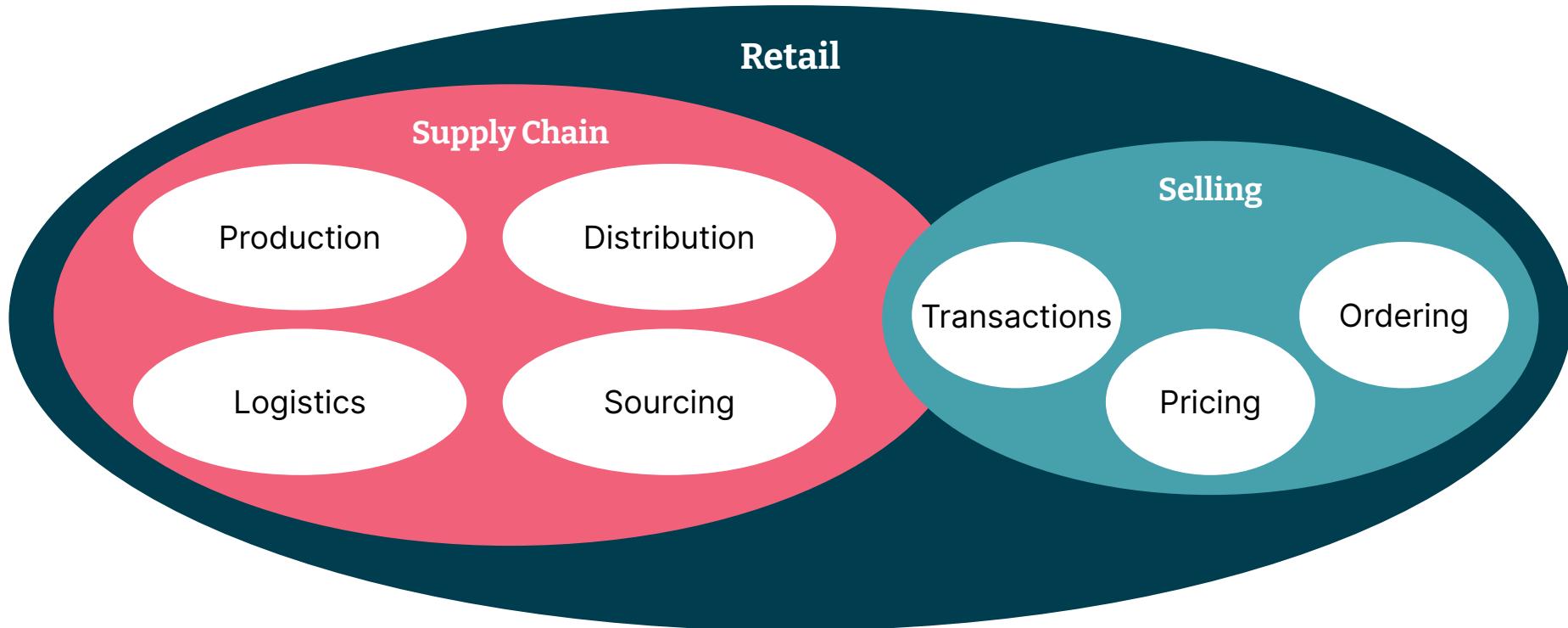


Models evolve



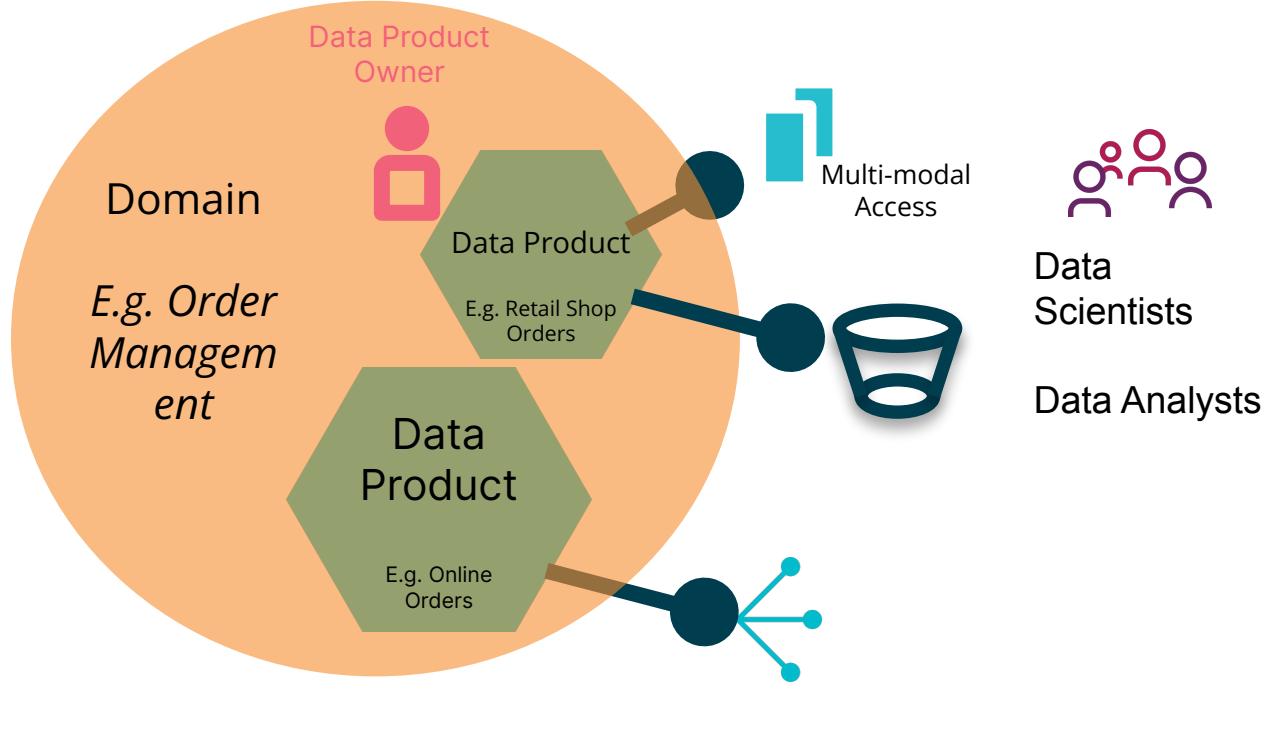
Adapted from "[Model exploration whirlpool](#)" by Eric Evans is licensed under [CC BY-SA 2.0](#)

Domains within domains



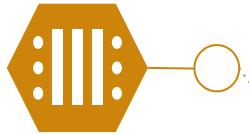
How Domains Share Their Data

Discoverable	
Understandable	
Addressable	
Secure	
Interoperable	
Trustworthy	
Natively Accessible	
Valuable on its own	



Example

Orders Data Product



/INPUTS :
Data product subscription,
collaborating service

Catalogue Service



Diversify (Domain Specific)

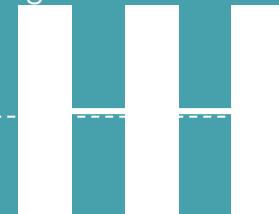
Standardize

Styles



**/DISCOVERY
/OBSERVABILITY :**
Identity, Semantic, Metrics, Documentation, ...

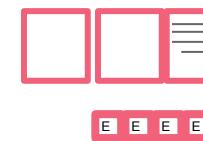
TRANSFORMATION
(e.g. Generative ML)



/OUTPUTS :
Bi-temporal multimodal streams



Recommendations



E E E E

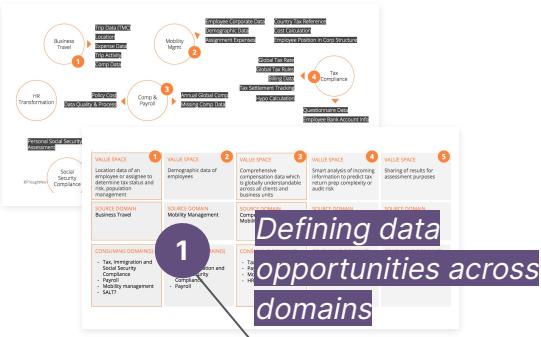


t

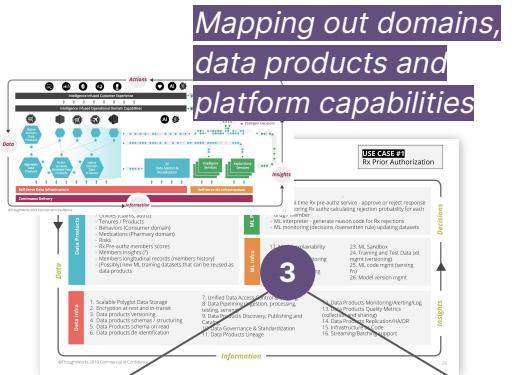


/CONTROLS:
Administrative Controls, Policies

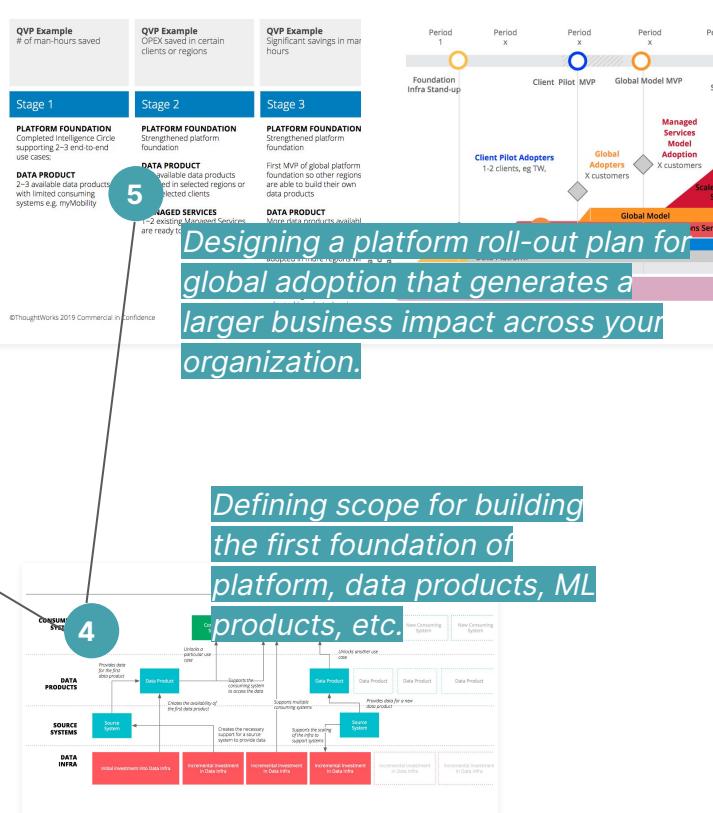
Mapping the journey from value



Defining data opportunities across domains



Articulating, breaking down, and prioritizing use cases



*Defining scope for building
the first foundation of
platform, data products, ML
products, etc.*

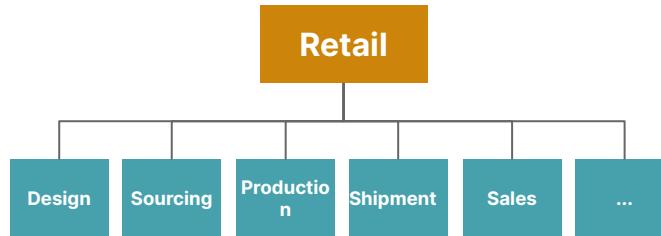
Heuristics to validate and refine domain ownership

Discover over Design



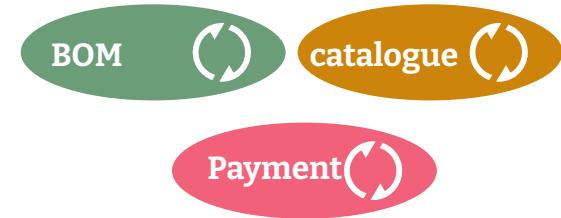
Is domain conducive
to long term
ownership?

*Does anyone care
about this domain
enough to own it?
Provider, consumer.*



Is domain aligned with business
functions?

*Is domain manufactured based on
theoretical modeling or matches reality?*



Does Domain Data have high
cohesion for ease of use and
atomic integrity.

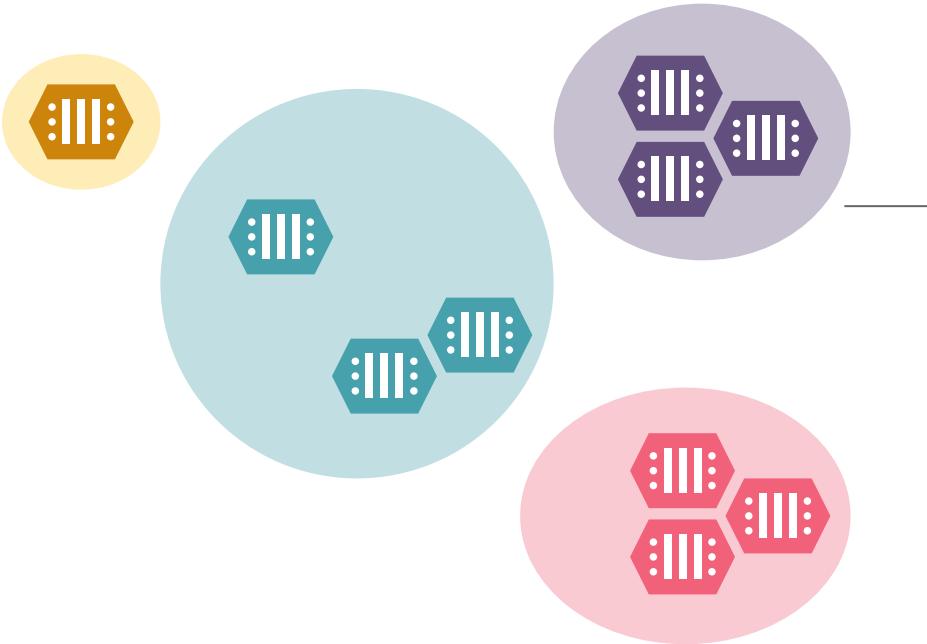
*Did we just pack data into a
domain to cater to team
allocation, or used bounded
context?*

How to structure teams and data products?

/thoughtworks



Teams, domains, and data products



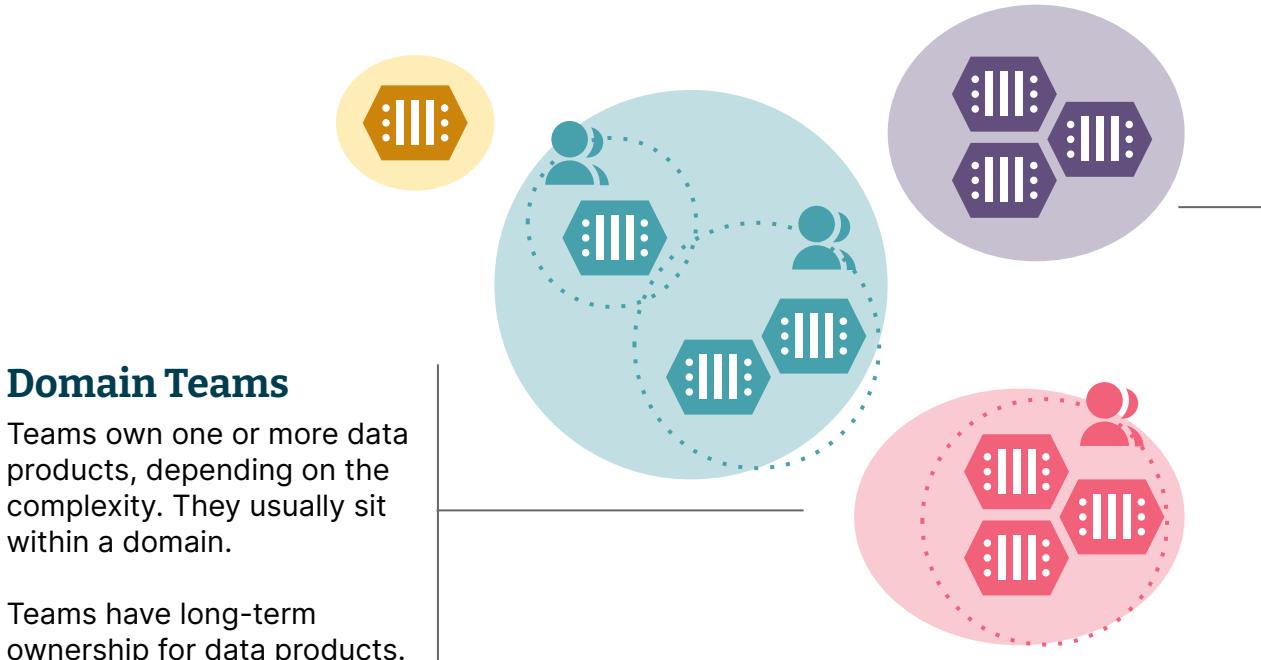
Data Products

Data products belong inside domains. A domain will usually contain many data products that can be used both within and outside its domain.

Types of data products:
source-aligned,
consumer-aligned, or aggregate
data products.

Data products don't operate in isolation.

Teams, domains, and data products



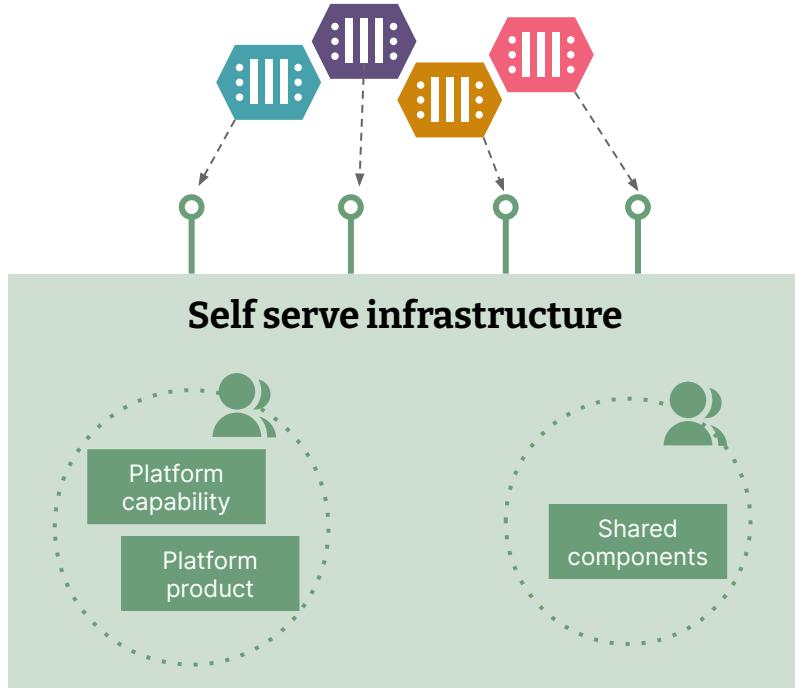
Data Products

Data products belong inside domains. A domain will usually contain many data products that can be used both within and outside its domain.

Types of data products:
source-aligned,
consumer-aligned, or aggregate
data products.

Data products don't operate in isolation

Self serve infrastructure teams



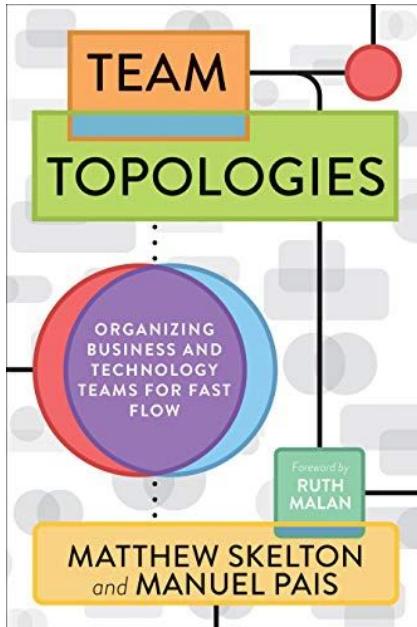
Platform Teams

To build and operate the self serve infrastructure, shared components, and platform products used by data product teams.

They are domain agnostic.

Managing and evolving ownership

Using Team Topologies



Team
Types



Cognitive
Load

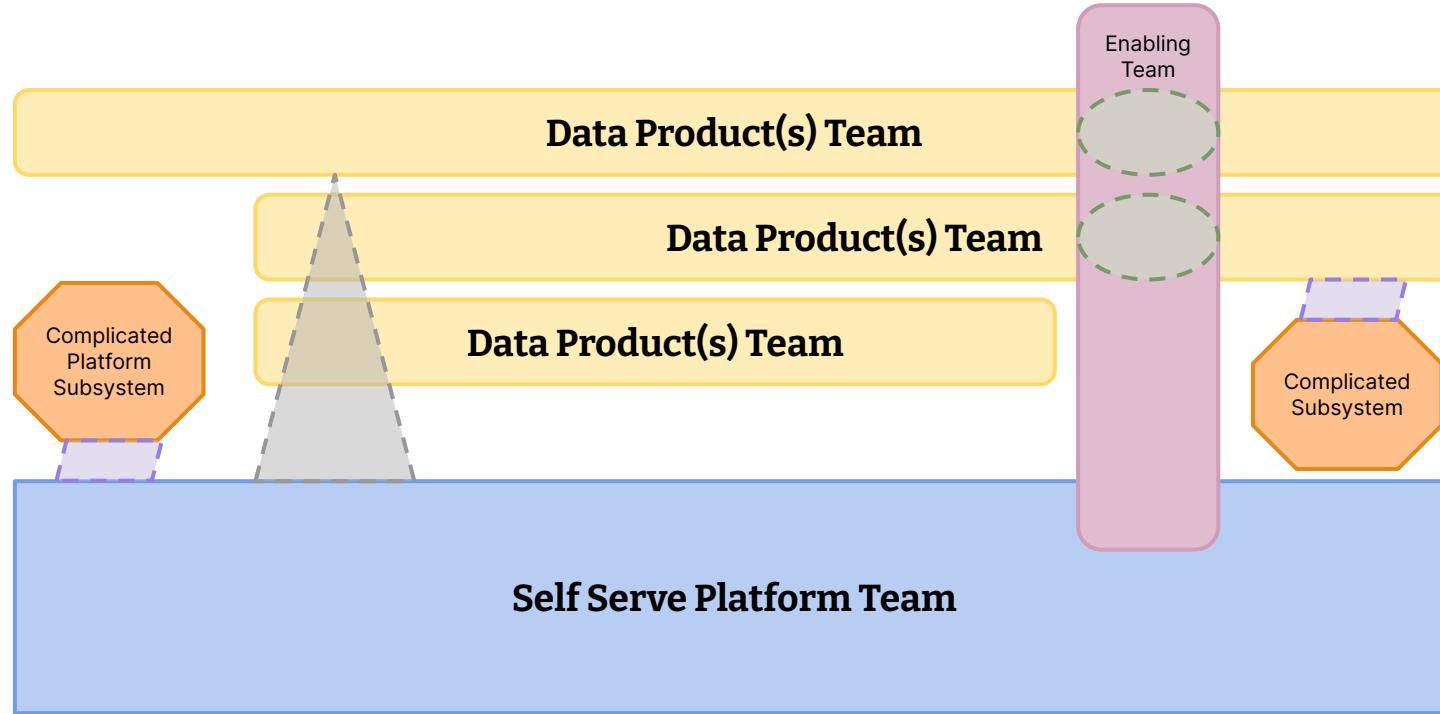


Flow of
Value



Interaction
Modes

Team topologies to manage and evolve ownership



Master Data Management

/thoughtworks





The Global Data Management Community

"Master Data is the data that provides the context for business activity data in the form of common and abstract concepts that relate to the activity. It includes the details (definitions and identifiers) of internal and external objects involved in business transactions, such as customers, products, employees, vendors, and controlled domains (code values)"
- DAMA, 2009



Employee



Product



Customer



Vendor



The Global Data Management Community

“Master Data Management (MDM) entails control over Master Data values and identifiers that enable consistent use, across systems, of the most accurate and timely data about essential business entities. The goals of MDM include ensuring *availability of accurate, current values* while reducing risks associated with ambiguous identifiers (those identified with more than one instance of an entity and those that refer to more than one entity).

- DMBOK2 (Page 356).



Accurate data



Current values



Consistent information to reduce risk

Let's look at an example

Product is a common entity that is subjected to Master Data Management



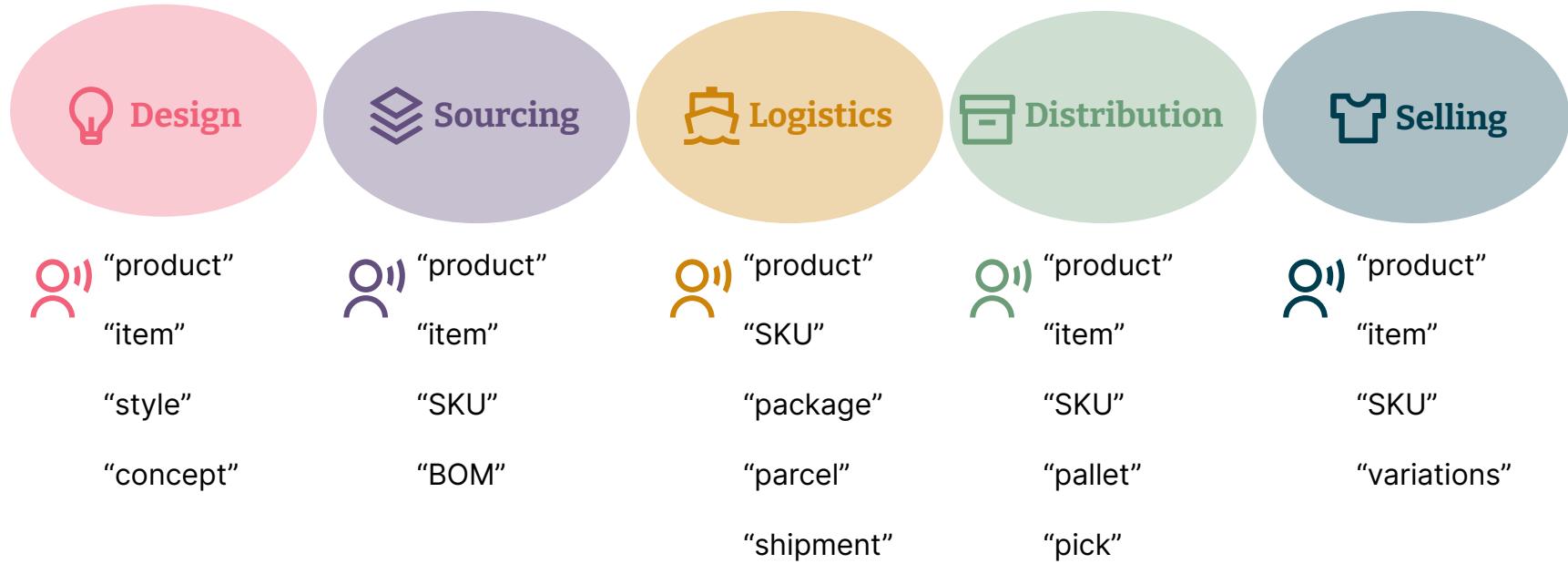
Long life cycle across entire retail domain



Has similarities in other industries

[Photo](#) by Charles Rodstrom is licensed under [CC BY-NC-ND 2.0](#)

Refining the ubiquitous language



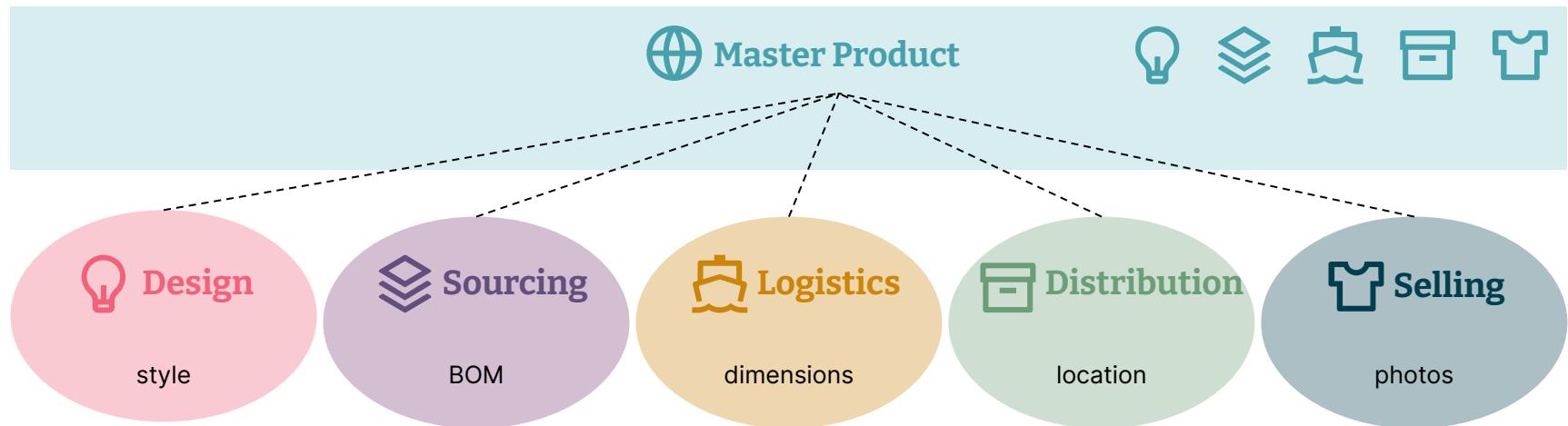
Different concepts in each domain

Contextual views of “product” attributes

Design	Sourcing	Logistics	Distribution	Selling
<ul style="list-style-type: none">• Color palettes• Fabrics / textiles• Flavours• Textures• ...	<ul style="list-style-type: none">• Suppliers• Bill of material• Cost• Size• ...	<ul style="list-style-type: none">• Packaging• Hazardous material• Customs codes• Dimensions• Weight• ...	<ul style="list-style-type: none">• Inventory• Expiry dates• Storage conditions• Warehouse location• ...	<ul style="list-style-type: none">• Channel• Photos• Price• Promotion• Stock availability• ...

MDM approach

Build a master entity to aggregate all disparate views of a product



Reality: data quality is still a problem

Despite valuable goals and good intentions, many MDM initiatives fail to deliver on the benefits.

They also tend to become a large undertaking that requires coordination across many parts of the business, creating friction and delays.

["Headache"](#) by Jose Navarro is licensed under [CC BY 2.0](#)



Common symptoms and problems

And what the user experiences

You can't solve MDM with just a technical solution. It goes beyond technology and into people and processes that cause data quality and ownership issues.



Buying a Solution



“One more place to get and explore data”

Common symptoms and problems

And what the user experiences

Data quality not fixed at the source,
propagates confusion.

Causes more problems long term.



Fixing Quality Downstream



“Why are these different?”

“Which one is right?”

Common symptoms and problems

And what the user experiences

“One model to rule them all” makes complex and large models that still are not fit for purpose and harder to understand.



Centralized Approach



“I don’t understand all these attributes”

“It still doesn’t fit my need”

Common symptoms and problems

And what the user experiences

Big initiative requires coordination across the organisation, and tend to be released in a “big bang” change.



Large “Big Bang”



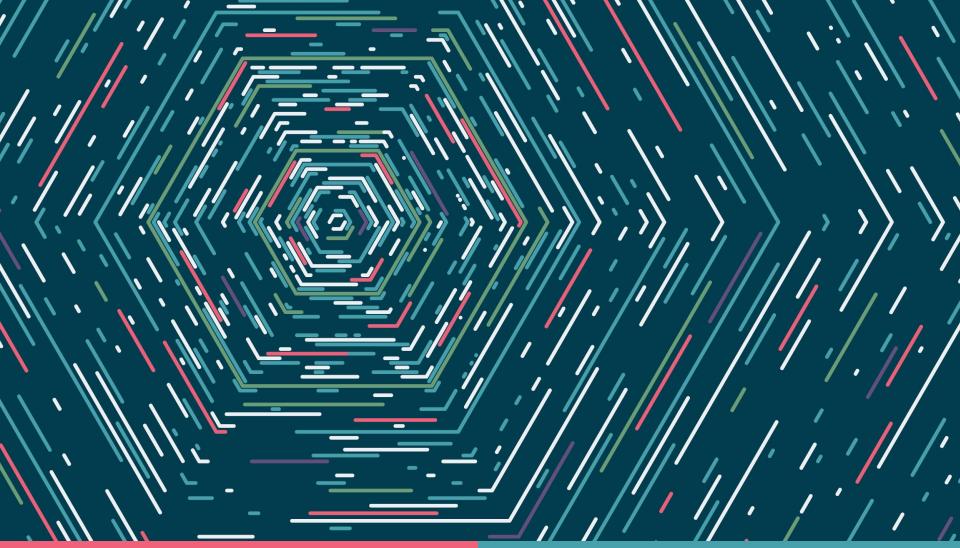
“I’ve heard about this initiative for a while, I’ll worry about it when it’s live”

MDM approach

Centralized representation creates coupling and forces synchronization



Why is it hard to create a single view of product?



1.

Too complex to represent in a single model

2.

Forces data at the most granular level

3.

Forces a single representation / shape

4.

Creates data quality issues

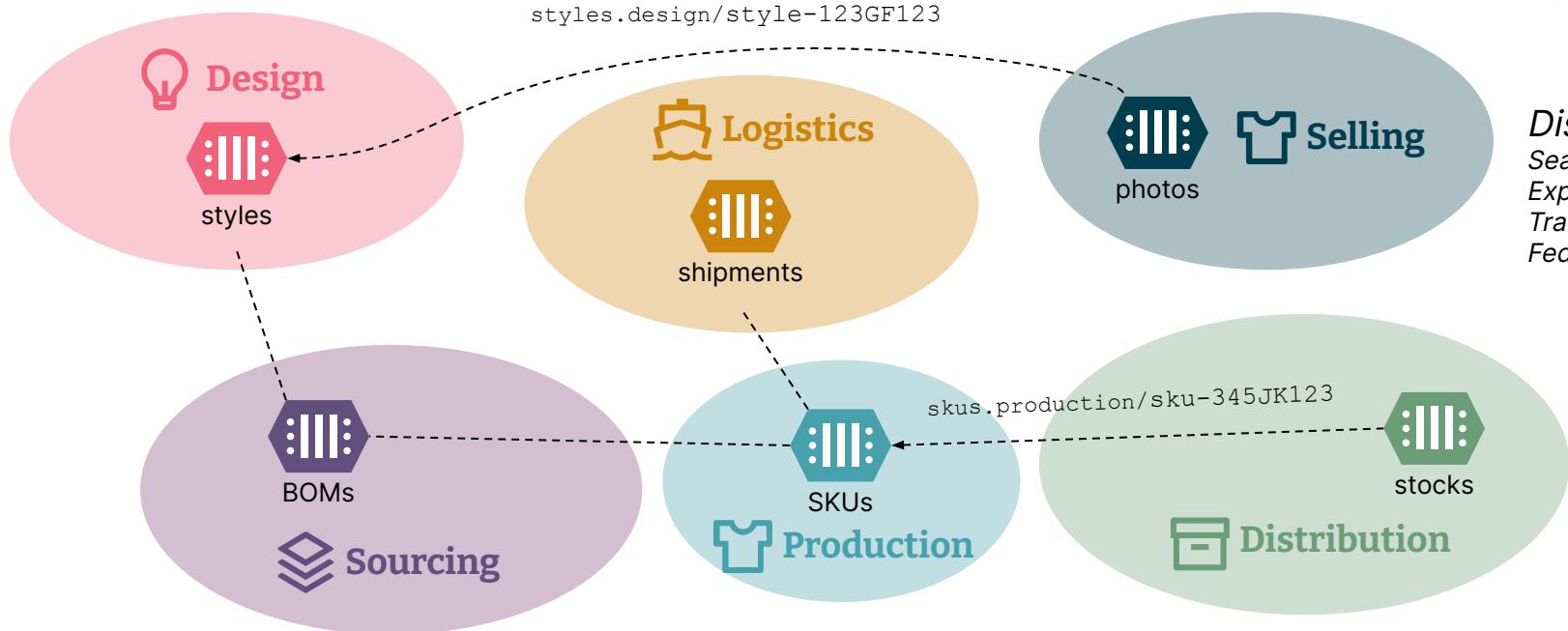
Data Mesh approach

Embrace decentralization and long-term domain-oriented ownership of data - *connect over collect*

			
Globally addressable identifiers Productionizing Domain-oriented data as <i>data products</i> require globally unique external identity system. <code>styles.design/style-12GF23</code>	Semantic linking of polysemes Linking the modeling of shared data across domains Ability to connect (over collect) <code>photos.selling { style is design.styles.id }</code>	Linking data Linking the data entities. When possible <code>photos.selling : [{ id: "photo-1231432", style: "styles.design/style-12GF13", ... }, ...]</code>	Local Bottom-up Definitions Each data product defines its links to its dependencies. The global mesh emerges from local rules.

Data Mesh approach

Federated entities to enable linking and fit-for-purpose aggregations



Discoverability:
Search (Index)
Explore
Traverse
Federated Query

Allow models to focus on different stages of a “master entity” lifecycle

Wrapping Up

/thoughtworks





Why

A solution that can scale out around the axis of organizational growth



What

Align tech, business, and data products around domains with long-term ownership of data by domains

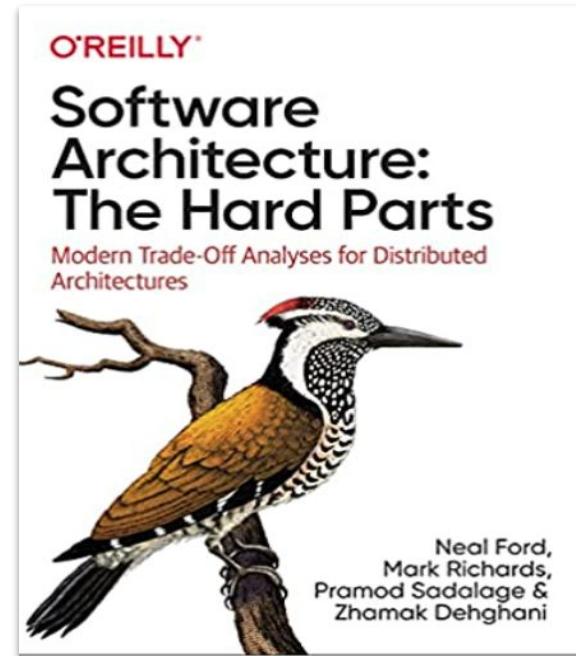
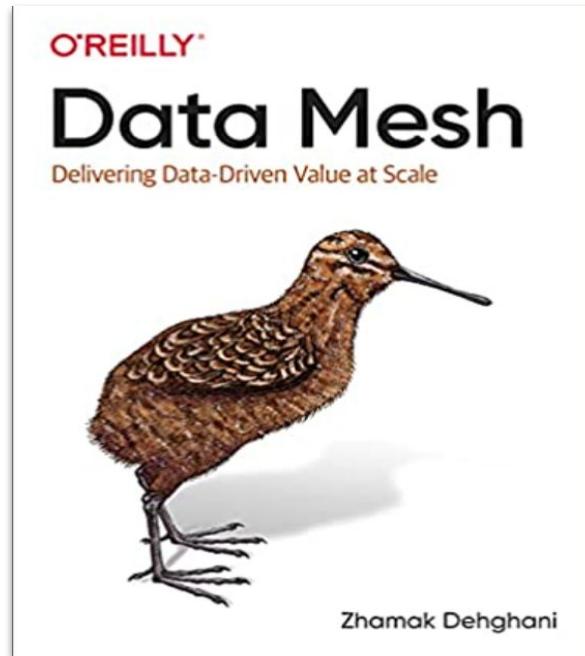


How

An evolutionary transformation from *incubation to sustaining to scale*

Choose goal fit for the stage of evolution # domains, data product & deploy a two-pronged approach for domain mapping

Learn More?



Learn More?

Virtual Conference: State of Data Mesh 2022

[View recordings](#)

Introduction Upcoming events Four principles Industry solutions Content vault

Data Mesh applies the principles of modern software engineering and the learnings from building robust, internet-scale solutions to unlock the true potential of enterprise data.

**Introduction to Data Mesh
A principled approach**

Zhamak Dehghani
@zhamakd

thoughtworks

34:52

<https://www.thoughtworks.com/en-us/what-we-do/data-and-ai/data-mesh>

Thanks! Any Questions?

Zhamak Dehghani

Director of Emerging Technologies, Thoughtworks North America

zdehghan@thoughtworks.com

 @zhamakd

Danilo Sato

Head of Data & AI Services, Thoughtworks UK

dsato@thoughtworks.com

 @dtsato

