# Goldilocks and the 3 (data mesh) bears

**@SimonAubury** 



Photo by Small Town Korea on Unsplash



#### Once upon a time ...

There were three bears, who lived together in a house of their own in a wood.

One of them was a **small bear**; one was a **middle-sized bea**r, and the other was a great, **huge bear**!

A little girl came into the house, named **Goldilocks**.



https://open.spotify.com/album/7fue7usKwgVD2YmyuJVMsA



## **Simon Aubury**

**Principal Data Engineer** 

Data enthusiast
 Doing cool things with data
 Sydney, Australia

#### /thoughtworks



## Goldilocks

#### **Fairytale star**

Exploring forest of data
 Understanding consequences of tradeoffs

#### What is data mesh?



Data is only valuable if consumers have it when when they need it.

Data Mesh is an approach on how we organise ourselves, take responsibilities as people for the vast amounts of ever growing data and its architecture.

Photo by Dorothe Wouters on Unsplash

## Is this different?



## Why Data Mesh

#### And why should I care?



 $\mathcal{O}$ Large user base fuelled by increasing need of multiple org functions to consume data for decision making Partner Ecosystem: Exchange of Data with other Ð organisations and advent of platforms Regulatory requirements are complex and tightening Lower Tech Barriers: democratise tech by lowering tech barriers. Discoverable, addressable, trustworthy, self-describing, inter-operable & secure Data Cloud nuances: Organizations are moving to cloud and よ this has a new meaning for structure of data

## **Principles of data mesh**

**Distributed domain** driven architecture





Data Mesh: Delivering data-driven value at scale book by Zhamak Dehghani

#### Do I really have to read a book?

and has a state and a damage address of the and and Bod Advances in the second sure adapte allowage sure quickly and other out her and the second and the same in

a real local

THE REAL PROPERTY AND

Name and Ard other designs, including south And the state of t and the Andrew Andrew Andrew The Port of the A spin and the spi And and a second on a second of any second of a second

spinster of states of the local division in which the real of the local division in the Design of the second se Appendix shows many 1. september of the side processing

state without third monthly share traditions for a minute well walked what eligenticals on only of the strategies within strends semblified and and same drawn with both totale property desidence on whether new algorithm monthly stilling. Hada and stores and size store hada ward story and allotty burness yads - sarrad sada real bilands you's roll motions and broad and The work thereast even your bloost meaning and And support of the state

more the process their series - on soil suscous to and an - Butrolanus Housed tonard and que and the protocology and realized as a prostant of the the most provide a deal wheely wheely and the rived we have the state at state well have been and herebols on moneyood one of and the failed and mentals and one have supported and particular of the

as man, "Assessible and Berry and it by marging and a stress born wisselshare periodenty married for

non reasons dougs sails and the set and and

a to store when generate at many at arguings & seen and

CONTRACTOR OFFICE AND

IS NOT THE ADDRESS.

sectors whethe states with the

the later of the state of the s

Image by <u>Mabel Amber</u> from <u>Pixabay</u>

## Domain Ownership

Ensuring data is managed close to its source of origin





give the originators of data authority over it

reduces the gap between producer and consumer

## Distributed domain driven architecture

Ensuring data is managed close to its source of origin

**Business domains**: Organisations are organised in business domain or functions

**Data mesh domains**: Data Mesh partitions data around business domains and gives data ownership to the domain teams. For this, domain teams should have both business and tech capabilities

**Change management**: data producers become data owners



## Data products

## Data as product means it maximises the quality, trust & utility





designed around the needs of its consumers

put the emphasis on use

## What is data as a product? Creating trust in data

Data as product means it maximises the quality and follows DATSIS principle. I.e. it is discoverable, addressable, trustworthy, self-describing, interoperable, and secure.



#### **D-Discoverable**

A- Addressable

**T-Trustworthy** 

S-Self-describing

I-Interoperable

**S-Secure** 

### Data product?



Self Describing
Discoverable
Addressable
Trustworthy

2,033 global ratings

74%

11%

7%

3% 5%

5 star

4 star

3 star

2 star

1 star



#### About this item

- From ear to toes 36 inches, filling high quality PP cotton
- Plastic eyes on the toys is very strong, will not fall off, never causing the child eat
- Big teddy bear high quality plush with well stitching, adorable rosette, surface washable
- Premium Quality and size, which is Perfect for hugging and snuggling
- Complete with a ribbon on his neck,Great gift for any occasion, such as birthday, christmas or any other holidays

@Si	mo	nAu	bury
-----	----	-----	------

## Self-serve data

## Self-serve capability in hands of domain team





creation and consumption of data products without a human

decrease lead time

## Self-serve infra

**Empowering domain teams** 

Data as a product means that there are new roles and they are responsible for data processing. For this, the teams need access to a high-level, self-serve data platform. This empowers teams with the autonomy needed to manage their domains.



Self-serve capability in hands of domain team

Allows the teams to produce and maintain their data products

## Federated data governance

Allows interoperability between different domains





make the requirements of good citizenship clear managing with constraints empowers teams

## Data to computational governance

Data governance is a critical aspect of any organization, and it becomes even more important in a distributed model

20400	U. Broth		
0	Filters 🗇	Connector - Q Search all assets 17 ©	engagement_mobile_daily
Assets	Certificate ~	All 238K ET Table 15K PE View 221 9E Materia	I Table > E brainly-bi > E users
lossary	Owners ^		, Open E Query ∨ 🐺 🗉 :
0	Q Sea Q A	engagement_mobile_daily	Description 8
(5) Insights	Kasia BM (Me)		This is a source of data about daily and monthly
	O No Owners	sessions and users on Mobile Apps. Preview	Col.
	Classifications ~	1.7K rows 5 columns	UJ View README →
	Terms v	O any the dd esseine at O	Rows Columns Size
	101110		1,072 5 58.78 KB (2
	Properties ~	Source table for agr_bigquery_textbook_data_pt_daily.	Terms
$\triangleright$	🗙 dbt 🗸 🗸	48 columns	T Reso
rkflows			Owners Que
		@ agr_tbs_dd_sessions_pl ♥ View > B brainly-bi > B textbook agg data daily	+
o		Source table for agr_bigquery_textbook_data_pl_daily.	Classification Req
Admin		48 columns	+ Prop
ha			Certificate
porting		ⓐ agr_tbs_dd_sessions_us	🔮 Verified
_		Source table for agr_bigquery_textbook_data_us_daily	R katarzyna.bodzioch- 4 months marczewska ago Pric
6.0		48 columns	2.15

#### Data governance at Brainly

## Federated data governance



<sup>&</sup>quot;Enterprise Data Model"



Data Catalog = Technical Metadata + Ownership Information + Data Lineage + Business Glossary

Data Quality = DQ Rules visualisation + Self-serve capability to create rules

Data ownership= Personas + DG Committee + Cadences + Org Incentives



## What could go wrong?

@SimonAubury

Photo by Sandy Millar on Unsplash

## Some data mesh trajectories



Scope of domain use cases

## Foundational pillars of Data Mesh

Four interconnected & non-negotiable



## Some data mesh inflections



## SHOULD WE START?

## Where to start?

Photo by Marina Shatskih on Unsplash



### DM ready reckoner



#### How Not to Mesh it Up-Barr Moses













Attendee data products



#### Data Product Stewards



## Implementation

Image by lisa runnels from Pixabay

30

#### Data mesh implementation: top down

#### Advantages:

- Clear visibility
- Ability to choose right milestones
- Organisation wide change management program is easy when there is buy in from top

#### **Considerations:**

- Large organizational impact
- Big commitment
- Organisation maturity required

#### Setting up enterprise domains & data culture

Change Management, Organisational Incentivisation, Ownership policies, Domain standards, Governance policies, et al Either set up all domains or choose one domain depending on business situation.

#### Self serve infra platform

Self serve data platform with infrastructure, product and pan enterprise data discovery capabilities

#### Federated data governance

Federated DG= Data Quality + Data Catalog + Data Ownership Data Catalog Technical Metadata + Ownership Information + Data Lineage + Business Glossary Data Quality = DQ Rules visualisation + Self-serve capability to create rules Data ownership= Personas + DG Committee + Cadences + Org Incentives

#### Data as product

Data assets discoverability, interoperability, addressability, self-description, trust, security. Contract definition of product, Self serve capability to easily and uniformly create data products.

Domain Onboarding

#### Data mesh implementation: bottom up

#### Advantages:

- Iterative process: test & learn
- Ability to show results
- Change management in small steps
- Nascent maturity orgs can also adopt this approach

#### **Considerations:**

- Ability to scale up the program
- Leadership engagement
- Long roadmap

#### Setting up enterprise domains & data culture

Change Management, Organisational Incentivisation, Ownership policies, Domain standards, Governance policies, et al Either set up all domains or choose one domain depending on business situation.

#### Bare minimum self-serve capabilities to support source oriented data products



### **Goldilocks choices**



#### A tailored Data Mesh for reMarkable

#### Data mesh implementation: hybrid



#### Advantages:

- Ability to show results across
- Faster approach

#### **Considerations:**

- Complex program management
- Leadership commitment
- Massive and rapid change management

#### Setting up enterprise domains

Change Management, Organisational Incentivisation, Ownership policies, Domain standards, Governance policies, et al



## Fairy tail ending?

Image by <u>Gerd Altmann</u> from <u>Pixabay</u>

### Takeaways



	Adopt all four pillars of Data Mesh		
Ц	Build your platform at the same time as you're building your initial data products.		
	Find one "courageous" domain to move forward		
Ц	Organisational appropriate adoption		
	Don't overbuild early		

### **Goldilocks**?



Goldilocks learns to show respect to other people's things

Data Mesh is a **framework to take responsibilities for data ...** and put the respect & needs of the data customer first.

Image by <u>Petra</u> from <u>Pixabay</u>

#### @SimonAubury

## **The-end** (any questions?)

#### Slides





Photo by Nathan Dumlao on Unsplash