



I care about data products ✨



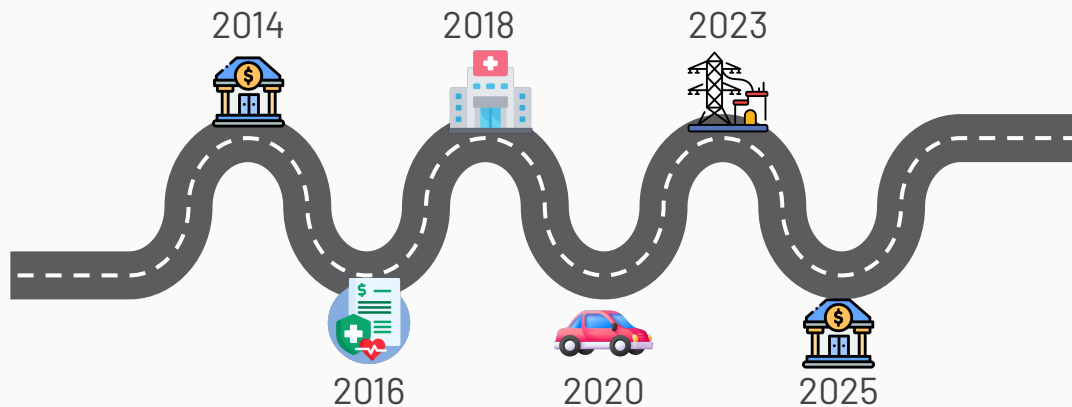
... and so should you

  @SimonAubury



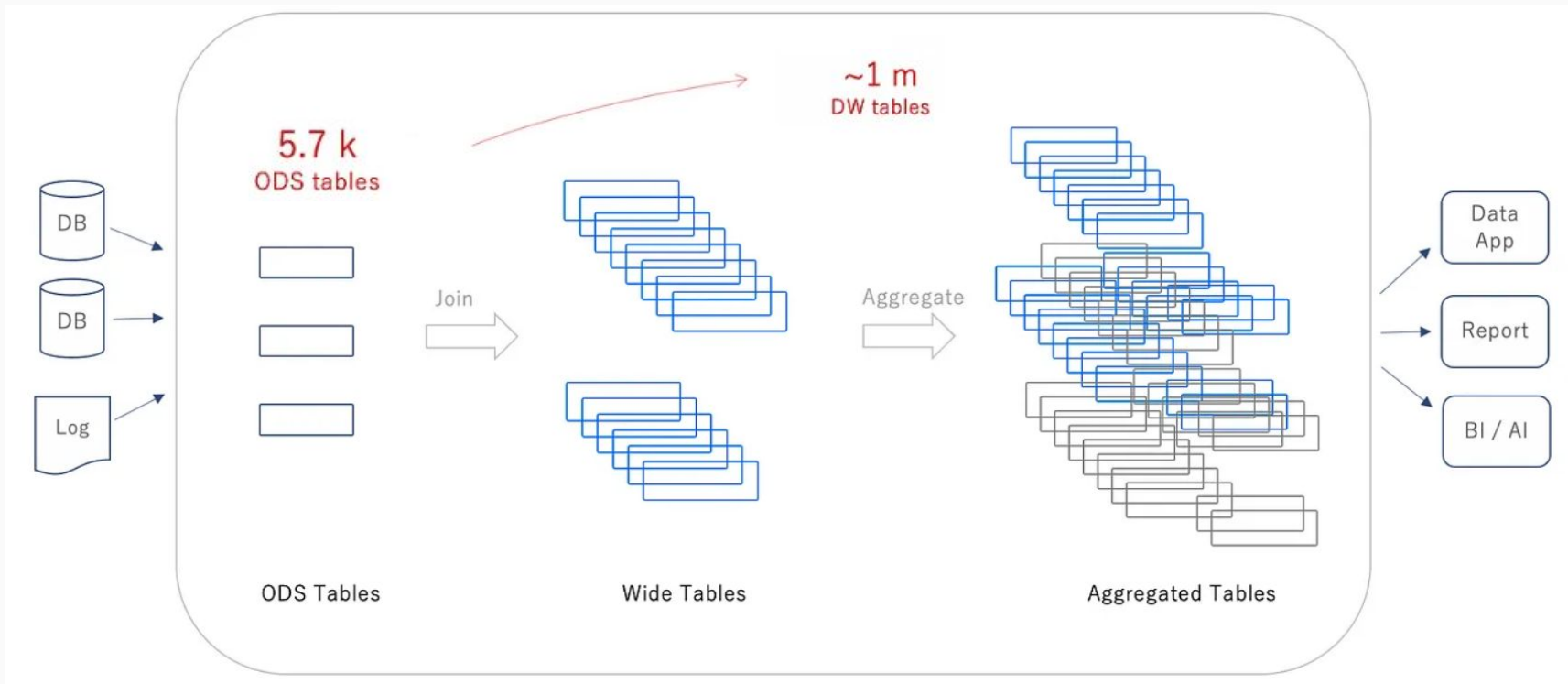
Simon Aubury

Data Geek





✧ A typical company ✧



[6k source tables in 2.5 years - Lori Lu](#)



Business teams are **working in silos**,
and data pipelines and **metrics are**
barely reused across all downstream
data applications.



When data meets
the real world



Photo by [Hassan OUAJBIR](#) on [Unsplash](#)

✧ Let's treat our data as any other product ✧



“A data product is a reusable, active, and standardized data asset designed to deliver measurable value by applying product thinking principles.”

—Jean-Georges Perrin



What's in a data product?



Data product characteristics

Discoverable

present relevant metadata information

Understandable

how to use your data product

Accessible

interfaces to access data products e.g. DB's, API's, files

Trustworthy

data fits to the usage expectations

Self-describing

effortless integration and exchange of information

Secure

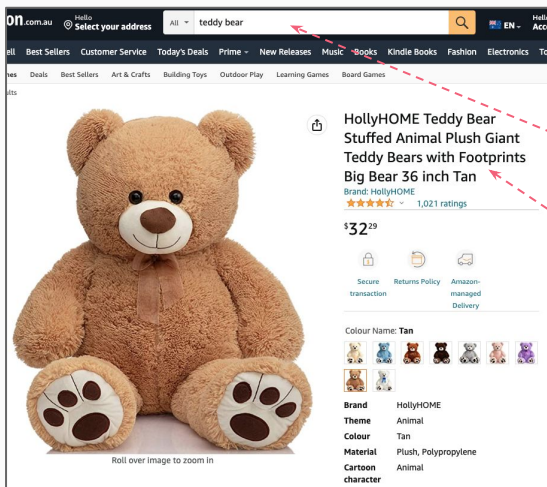
access control, encryption and privacy



Think of a product



Product

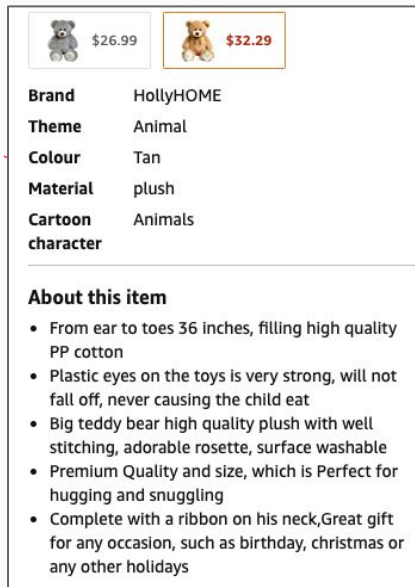


The screenshot shows an Amazon product page for a 'HollyHOME Teddy Bear Stuffed Animal Plush Giant Teddy Bears with Footprints Big Bear 36 inch Tan'. The price is \$32.29. The page includes a large image of the bear, a search bar with 'teddy bear' entered, and various navigation links. The product details section lists the brand as HollyHOME, theme as Animal, color as Tan, material as Plush, Polypropylene, and cartoon character as Animal. There are also icons for secure transaction, returns policy, and Amazon-managed delivery.

SELF DESCRIBING

DISCOVERABLE

ADDRESSABLE



This block contains product details and customer reviews. At the top, there are two price options: a smaller bear for \$26.99 and the larger bear for \$32.29. Below this is a table of product details:

Brand	HollyHOME
Theme	Animal
Colour	Tan
Material	plush
Cartoon character	Animals

Below the table is a section titled 'About this item' with a list of bullet points:

- 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

At the bottom of this block is a 'Customer reviews' section showing a 4.5 out of 5 star rating based on 2,033 global ratings. A bar chart shows the distribution of ratings:

Star Rating	Percentage
5 star	74%
4 star	11%
3 star	7%
2 star	3%
1 star	5%

UNDERSTANDABLE

TRUSTWORTHY



This block contains care instructions and manufacturer information. It lists several bullet points for washing and drying the bear:

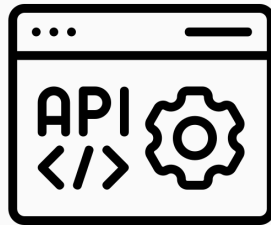
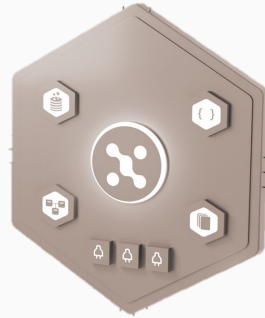
- HAND WASH WITH MILD DETERGENT.
- WASH DARK COLORS SEPARATELY.
- DO NOT DRY IN DIRECT SUNLIGHT.
- USE MILD IRON.
- DO NOT BLEACH.
- DO NOT SCRUB ON FOIL / BLOCK PRINTS.
- DO NOT TUMBLE DRY

Below the list are icons for a hand, a box, a crossed-out box, and a crossed-out box with a circle, indicating care instructions. At the bottom, it says 'MADE IN INDIA'.

What's in a product?



What's in a data product?





Orientating the data product



Data products vs. data as a product

Data product

... is a data asset, or delivered unit of data, that solves a specific business problem

THE WHAT

Data as a product

... is a mindset or approach that applies product-like thinking to a dataset

THE HOW

[Data products vs. data as a product - Kathryn Chubb](#)

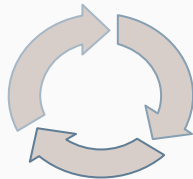


Data product perspectives



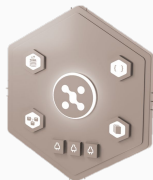
Product manager

A data product is a product offering customers a value proposition, a clear use case, and a design that aligns with an ideal customer profile



Engineer

A data product is a data asset that contains other artifacts (e.g., tables, pipelines, models). It includes data contracts to ensure the consistency and quality of inputs and outputs



Data Product



The product management view



Value proposition

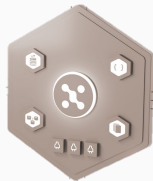
unique value and benefits

Clear use case

solves specific problems or meets particular needs

Measurable outcome

track progress using key performance indicators



Data Product



We're selling teddy bears





Value
proposition

 *ACCURATE SALES AND INVENTORY INFORMATION DRIVE
HEALTHY REVENUE GROWTH*

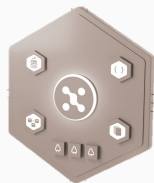
Clear use case

 *I WANT TO OPTIMIZE MY SALES PERFORMANCE AND
ENSURE INVENTORY MANAGEMENT OPTIMIZATIONS MATCH
SALES VOLUME*

Measurable
outcome

 *INCREASED REVENUE: TRACK SALES GENERATED OR REVENUE
GROWTH ATTRIBUTED TO THE DATA PRODUCT.*
 *IMPROVED CUSTOMER SATISFACTION: NPS SCORE*

The engineers view



Data Product

Data and metadata



**DATA IN
DATABASE TABLE,
DATA LAKE OBJECTS**

Code



**LOGIC,
ALGORITHMS, AND
DATA PROCESSING
PIPELINES**

Contracts



**AGREEMENT
BETWEEN THE DATA
PRODUCERS AND THE
DATA CONSUMERS**

Infra



**UNIFIED
ARCHITECTURE -
SPUN UP PURLEY TO
SUPPORT DATA
PRODUCT**

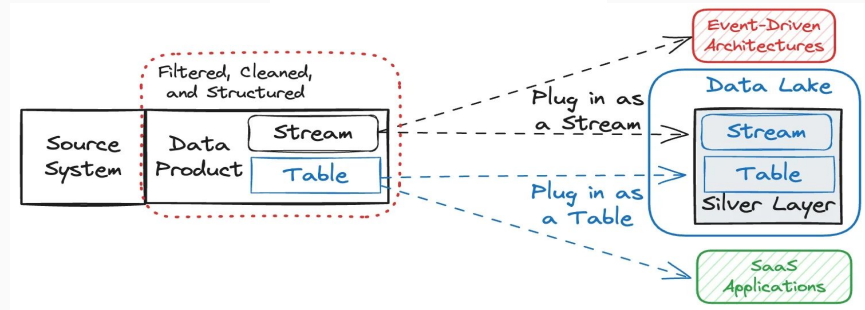
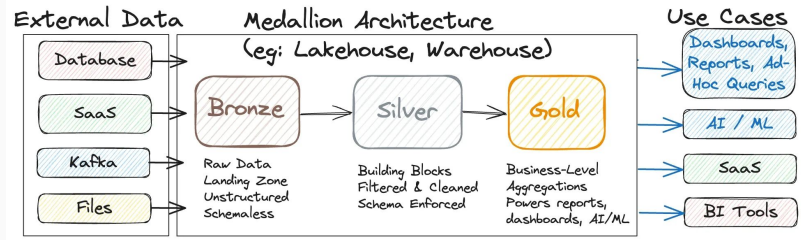


I think I've
seen this before?



Compared to the Medallion architecture

Easier, cheaper, and simpler to access data



[Rethinking the Medallion Architecture - Adam Bellemare](#)














**Let's start
building**





The mission



Solve a well-defined scope; the value generated by data products is its use and practical application



Data Product Canvas		Product Name:	Owner:	Date:	Version:
Problem  <i>Ask the right questions:</i> <ul style="list-style-type: none">• What is the problem?• Why is it a problem?• Whose problem is it?• Why, why and why...?	Data  <i>Ask the right questions:</i> <ul style="list-style-type: none">• Source• Quality• Access vs. Availability• Process / Transformation• Outputs• Test / Training / Validation	Solution  <i>Ask the right questions:</i> <ul style="list-style-type: none">• Type (Analytics, ML, IA,)• What will be the solution?• Output expected?	KPIs  <i>Ask the right questions:</i> <ul style="list-style-type: none">• How to evaluate the model?• Which metrics should be used?• How much uncertainty can we handle with?• AB Test – How?	Actions  <i>Ask the right questions:</i> <ul style="list-style-type: none">• Which actions will be used?• Which campaigns?	
	Hypothesis  <i>Ask the right questions:</i> <ul style="list-style-type: none">• What will be tested?• What are the expected responses for each of them?• What should we do from each answer? What strategy should we follow?		Actors  <i>Ask the right questions:</i> <ul style="list-style-type: none">• Who is your client?• Who are your stakeholders?• Who is your sponsor?• Who will use the solution?• Who will consume the solution?• Who will be impact with?		
Values  <i>Ask the right questions:</i> <ul style="list-style-type: none">• What is the size of your problem?• What is the baseline?• What is the uplift / savings?	Risks  <i>Ask the right questions:</i> <ul style="list-style-type: none">• What are the risks?• What these risks might block?	Performance / Impact  <i>Ask the right questions:</i> <ul style="list-style-type: none">• What is the impact? How to measure it?• Where you can see this improvement / performance? <p><i>Example:</i></p> <ul style="list-style-type: none">• Increase our customers base;• Keep it them by Reduce Churn;• Savings lost revenue + A/B Test;• Reduce cost of acquisition;			

 [br.linkedin.com/company/dataproductcanvas](https://www.linkedin.com/company/dataproductcanvas)



 <https://www.dataproductcanvas.com>
Prof. Leandro Carvalho, PhD
 [br.linkedin.com/in/leandroscarvalho](https://www.linkedin.com/in/leandroscarvalho)

Data Product Canvas - Leandro Carvalho



Source or consumer aligned

Source aligned data product

represents the data as it is in the operational system with minimal transformation



POINT OF SALES SYSTEM

Consumer aligned data products

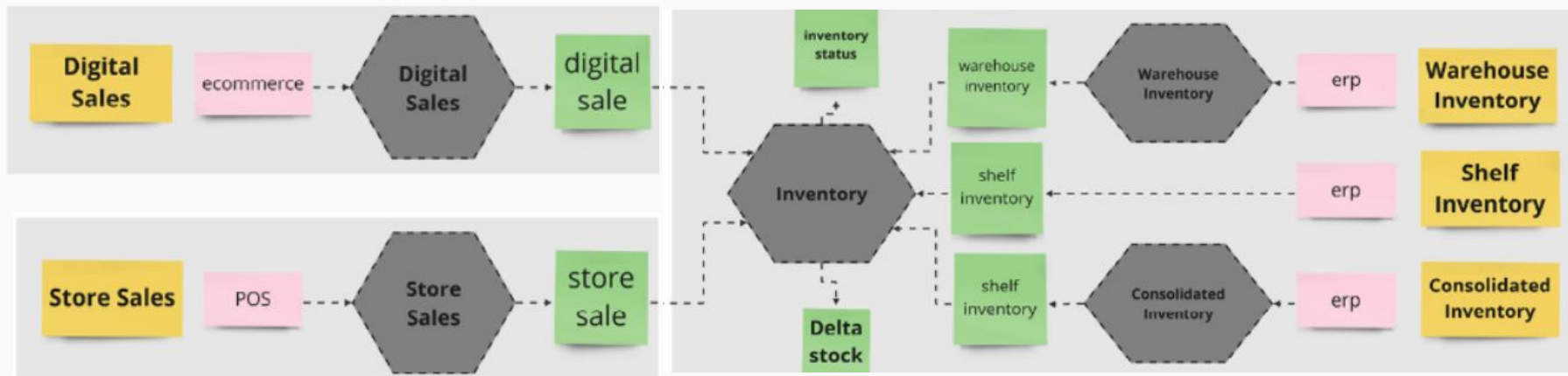
produced by business experts within the domain that generate value through the codification of business knowledge and expertise



CONSOLIDATED INVENTORY



Domain-aligned data products



Managing Data as a Product
Andrea Gioia ; Figure 4.17

How data becomes product



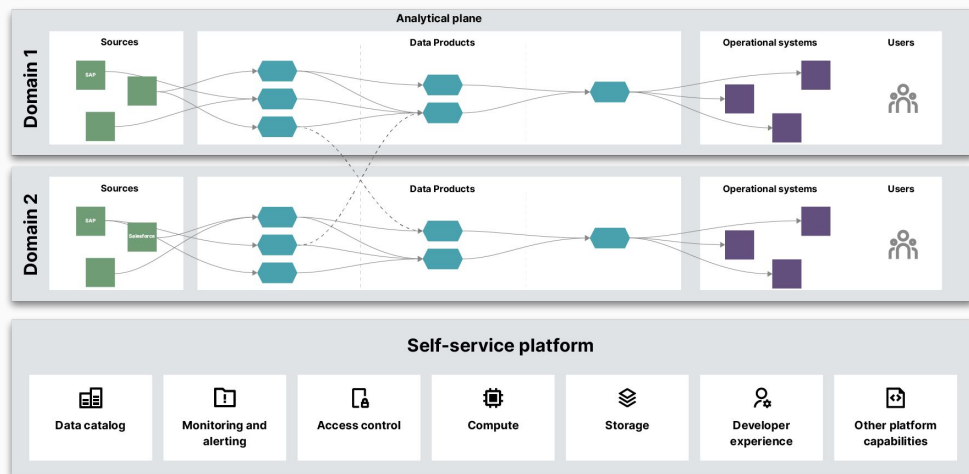
Where Exactly Data Becomes Product - Animesh Kumar and Travis Thompson



**Lots of
data products**



What about data mesh?



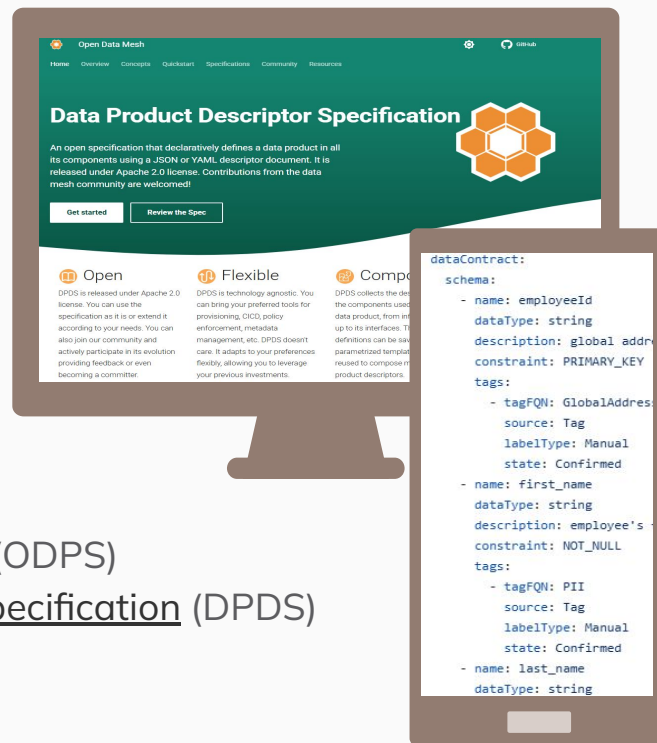
Data products are the architectural quantum of the Data Mesh.

Designed as the smallest functionally cohesive unit of the mesh, each with an independent life cycle.

[Data Mesh in practice - Thoughtworks](#)

Data product descriptor specifications

- Linux Foundation's Open Data Product Specification (ODPS)
- Open Data Mesh (ODM)'s Data Product Descriptor Specification (DPDS)
- Innoq Data Product Specification open initiative





TLDR ...






Building your first data product


Clear & achievable goal

 *SHOW PROOF OF VALUE
QUICKLY - AUTHORITY TO SCALE*

Minimum lovable product

 *TEST WITH STAKEHOLDERS -
ITERATE AND ADD FEATURES AND
METRICS*

Adoption

 *IT'S ONLY VALUABLE IF ADOPTED
- MAKE SURE YOU ARE SPENDING
TIME WITH THE STAKEHOLDER*

Existing technology

 *BUILD WITH WHAT YOU HAVE*



Thanks!



Do you have any questions?

  [@SimonAubury](https://twitter.com/SimonAubury)

Links to books, products and further resources

<https://github.com/data-engineering-helpers/data-products>

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)