



ARCHITECTING FOR AI: AI/ML AS AN ENGINE FOR DATA MESH

Cristian Southall, CTO JourneyOne
Presented to:
Digital Transformation Perth
12th July, 2022

WHAT IS AI/ML GOOD FOR?

TRANSITIONING HUMAN INTERACTION TO HUMAN OVERSIGHT

Let's imagine that, unlike blockchain, AI/ML will not contribute to solving every problem everywhere.

In the digital transformation of businesses with physical processes, we see the proliferation of network-connected field devices. That is, the Internet of Things.

We see edge and cloud computing employed to ingest this data at scale.

We see the use of field devices to actuate our assets and the adoption of the 'digital twin' pattern to emulate and simulate operations.

Complex decisions prevent us from closing this operational loop. AI/ML is the key enabler that will transition these processes from requiring human interaction to requiring human oversight.

WHAT IS AI/ML GOOD FOR?

ENHANCING HUMAN CREATION WITH HUMAN CURATION

Not all work can be automated. Not all work should be automated.

For example, strategic decision making can't reasonably be automated but AI/ML can improve and support the process by automating the generation of scenarios or testing the validity of strategic assumptions.

In this way, humans improve their capacity for creative decision making through curation of AI-generated options.

This may well impact how we practice architecture in the coming years.

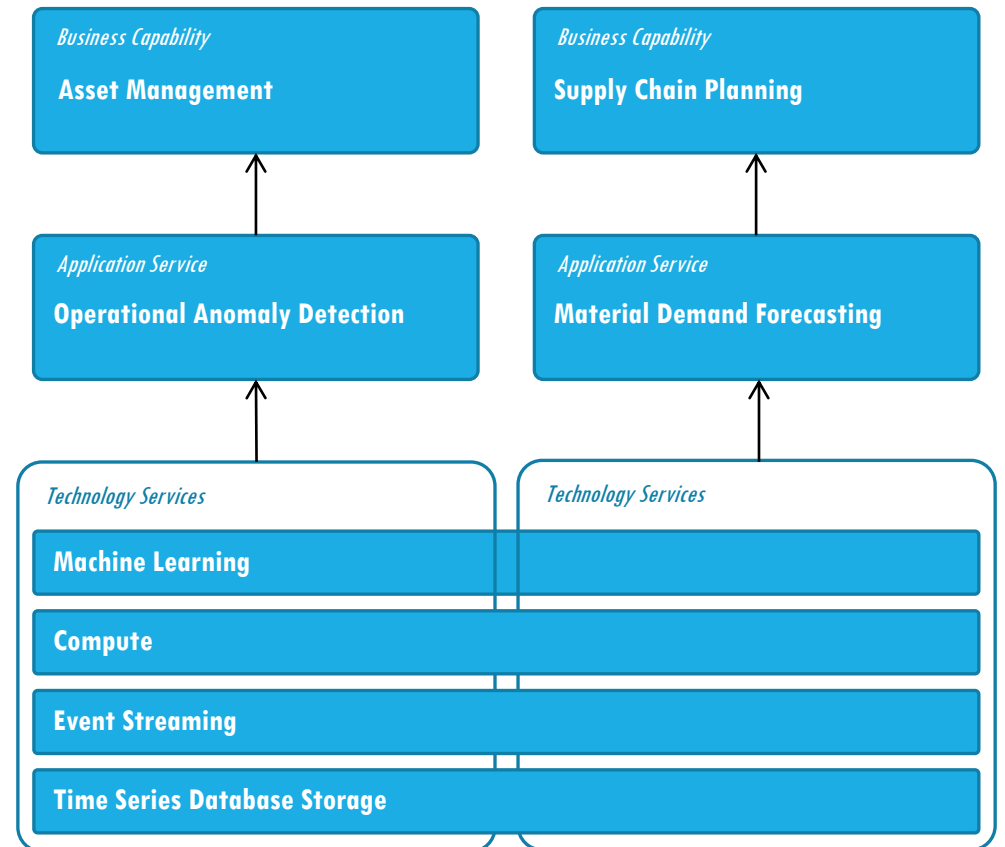
AI/ML IS NOT A BUSINESS CAPABILITY

AI/ML does not, in itself, generate direct business value.

AI/ML is one technology among many that might be composed as services used by business capabilities.

So, what does this tell us about designing and implementing AI/ML in the enterprise?

We should ...



PREFERENCE PRODUCTS OVER PLATFORMS

Technology platforms are great, but you need 'use' before you can have 're-use'.

Remember when we spent years and millions crafting 'complete' data warehouses and then years and millions in the maintenance of an entire platform and its suite of reports?

Why didn't we focus on only the components that demonstrably provided business value?

There is a customer/user somewhere that needs something. AI/ML may be the cornerstone of a solution to delight them but it won't be the only technology dependency and it likely won't satisfy the customer/user in isolation.

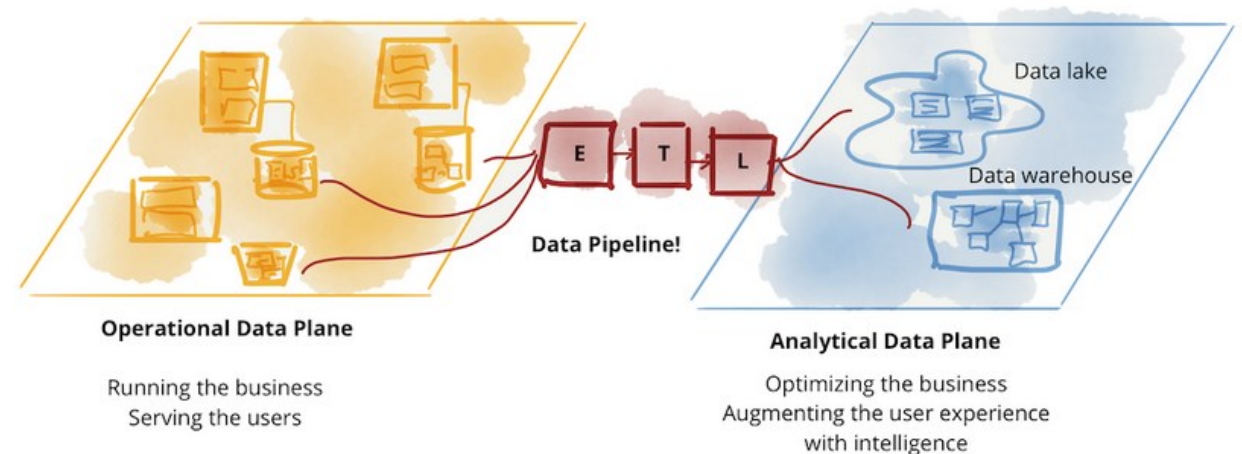
So, let's architect the user experience before architecting the system.

A CANDIDATE PATTERN: DATA MESH

Data mesh recognises two data planes: the operational and analytical.

They are integrated, yet separate. The source of fragile architectures that flow data between planes.

With respect to products over platforms: “Data mesh recognizes and respects the differences between these two planes ... However it attempts to connect these two planes under a different structure - *an inverted model and topology based on domains and not technology stack.*” (Dehghani)

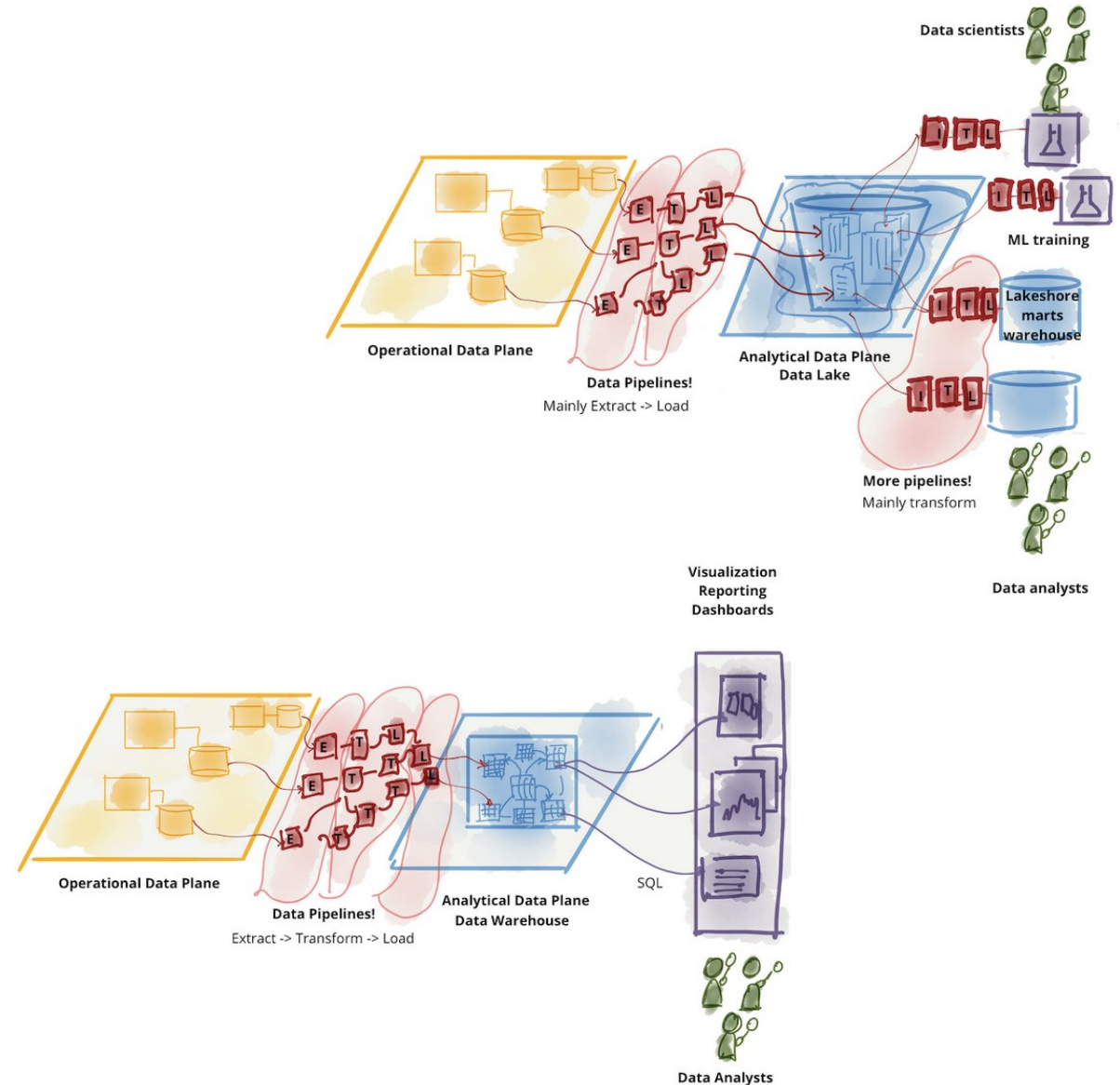


EXTENDING DATA MESH

AI/ML is considered a consumer of our data products in some views of this pattern.

This may be the case for ad-hoc ML analysis, but I would suggest AI/ML occupies two additional roles in this architecture:

- As a component of the operational data plane, augmenting operational systems and closing the 'operational loop'.
- As a component of the analytical data plane, as a tool for selecting and generating better analytical data.



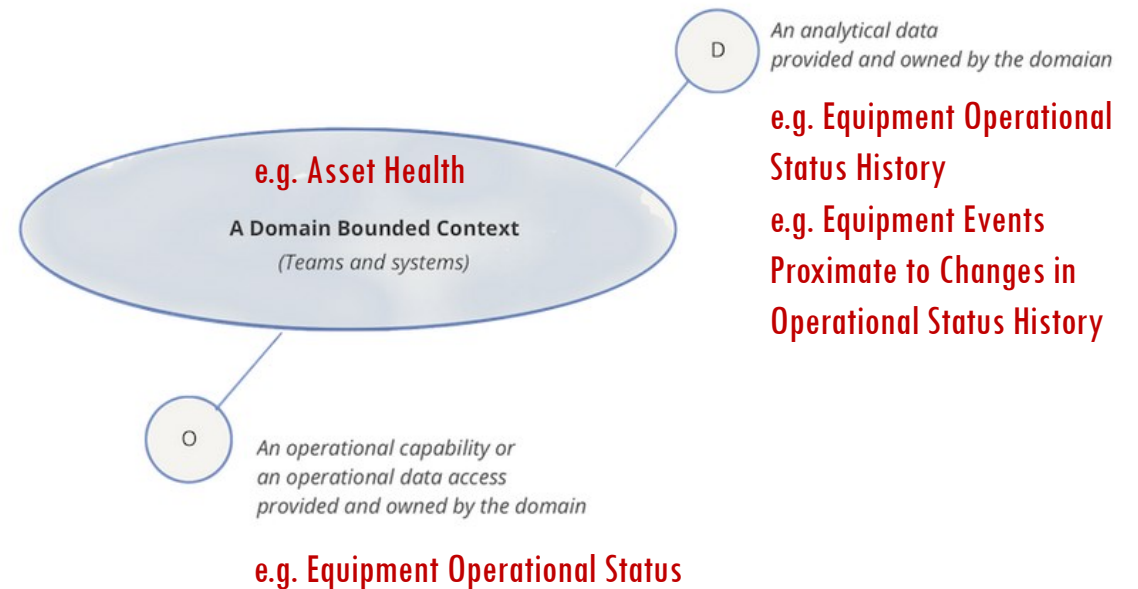
DATA MESH PRINCIPLES:

DOMAIN-ORIENTED, DECENTRALISED DATA OWNERSHIP

Promotes decentralisation and distribution of responsibility to the people who are closest to the data in order to support continuous change and scalability.

Domain composition is aligned with organisational units.

Our challenge is making the technical capability necessary for implementation available to domain data owners.



DATA MESH PRINCIPLES: DATA AS A PRODUCT

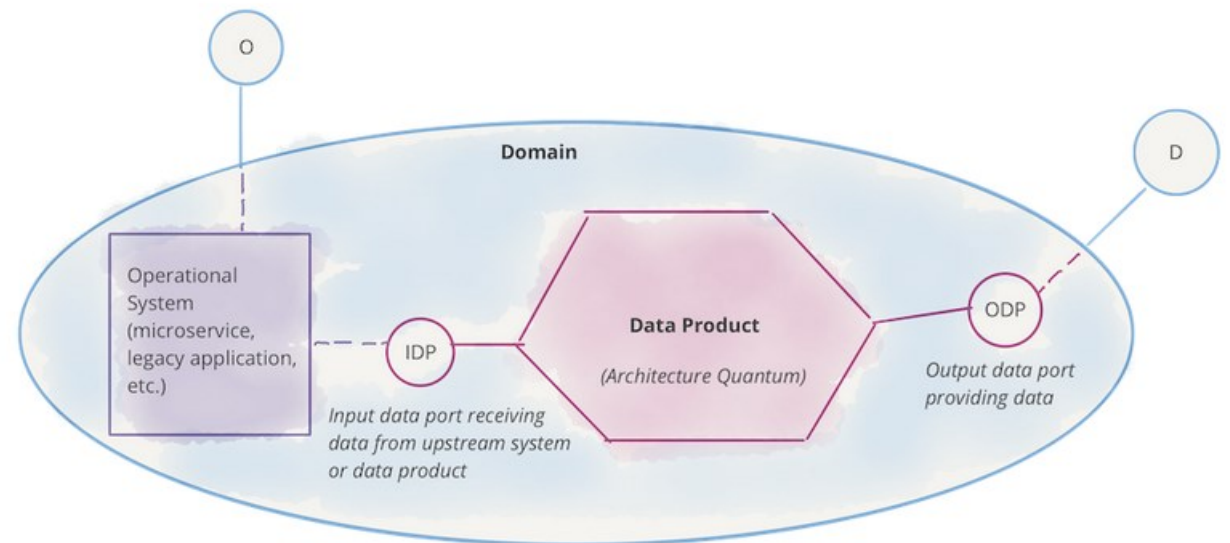
Asks us to treat data consumers as customers.

The experience of sourcing the data is considered as important as the data itself. This is what transforms a data artefact into a data product.

Data products are self-contained, comprising:

- Code
- Data/Metadata
- Infrastructure

When is AI/ML the product? When is AI/ML an enabler? In some use cases external customers consume the product, in others we are enabling internal customers.



DATA MESH PRINCIPLES: SELF-SERVE DATA INFRASTRUCTURE AS A PLATFORM

To enable re-use but preserve autonomy, data product owners should have access to high-level, domain independent infrastructure constructs. We don't need business domain experts to be infrastructure experts.

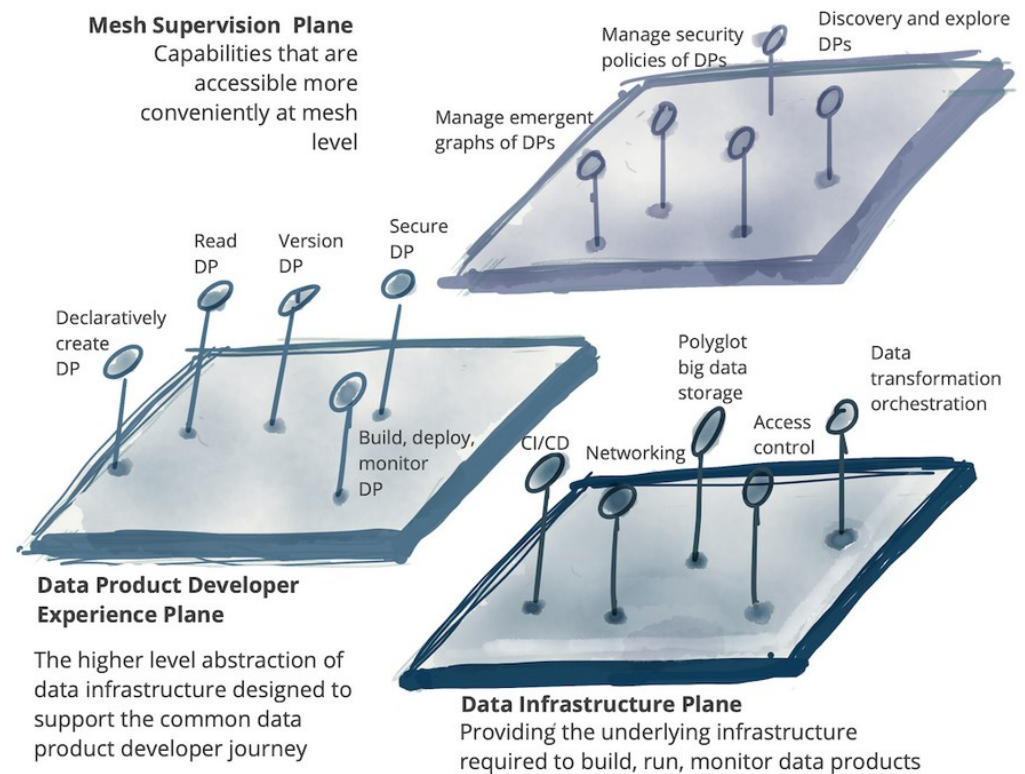
Divergence of the technology stacks we use to run operational services versus those that run analytical products has made this seem unattainable.

The good news is there is evidence of convergence:

- Consider the application of event streaming technologies like Kafka in both the operational and analytical planes.
- Consider the evolution of data platforms from proprietary suites to solutions leveraging open source technology and standards, and using the same runtimes that developers of operational services use. For example, Spark/Python.

The bad news is that this convergence is traversing the data lifecycle from acquisition to integration to storage but that the path is longer and less clear with respect to AI/ML stacks.

We can feasibly interchange application integration and data integration specialists. It's harder to see when AI/ML stack specialists will acquire generalist skills.



DATA MESH PRINCIPLES: SELF-SERVE DATA INFRASTRUCTURE AS A PLATFORM (CONTINUED)

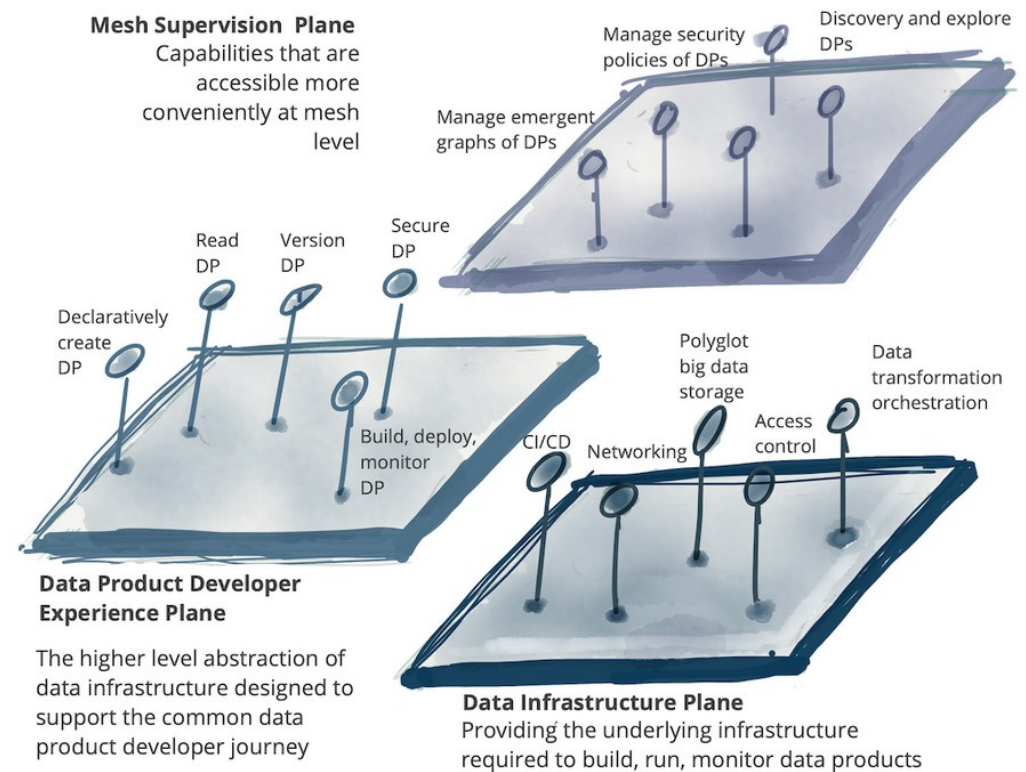
‘as-a-service’ consumption models are critical.

Not every organisation and not every domain needs a legion of data engineers, data scientists and AI/ML infrastructure specialists on the payroll.

Given a certain minimum preparation of data, AI/ML-as-a-Service may provide good-enough results with elastic consumption and less organisational commitment to scale up/down AI/ML capability.

Models are already commoditised, so as the available range increases there will be more options to buy rather than build.

Indeed, do what we do in operational systems development and ‘buy in’ services until you find a use case that warrants building infrastructure.



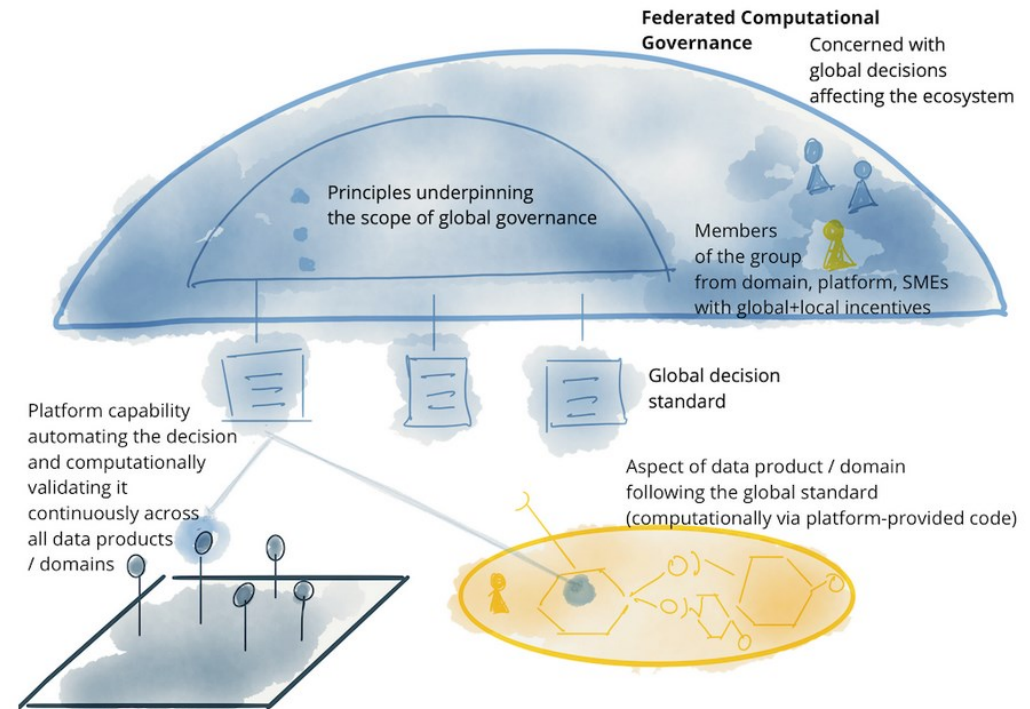
DATA MESH PRINCIPLES: FEDERATED COMPUTATIONAL GOVERNANCE

The is fundamentally about how to enable decentralisation but retain enough standardisation to preserve interoperation.

Standardisation concerns are both technical (e.g. communications protocols) and syntactic/semantic (e.g. how we might re-use key data entities across domains).

I was initially uncertain AI/ML had impact on this principle but:

- It feasible that the technology would play a key role in automating governance (e.g. data quality/integrity checking).
- AI/ML will need to be the subject of governance, especially given the ethical and practical concerns of model bias (i.e. explainable AI).



REFERENCES

<https://martinfowler.com/articles/data-mesh-principles.html>

<https://martinfowler.com/articles/data-monolith-to-mesh.html#DataAndSelf-servePlatformDesignConvergence>

<https://www.thoughtworks.com/en-au/insights/looking-glass/lens-one-partnering-with-ai>

ABOUT CRISTIAN SOUTHALL MBA BA(HONS)



As Chief Technology Officer for JourneyOne, I variously operate as a Consultant, Advisor, Architect, Product Owner and Technologist across industries and sectors, from startups to scaleups and global corporates. You can read more [here](#).

I'm a serial learner and multi-disciplinarian combining knowledge from the domains of technology, business and communications. I'm an active contributor to many professional communities and researcher in cloud-native/serverless computing, event-driven architecture, IT/OT/IIoT convergence and 'new data' capability, including data mesh.