

FEATURING FORRESTER

Why Building a Data Mesh Goes Beyond Technology Decisions—and How You Can Do It Right

We recently hosted a webinar entitled “Data Mesh: From Concept to Implementation”, with guests from Forrester and Lumen. We’re continuing the conversation in this Q&A with Michele Goetz of Forrester.

As businesses take a hard look at getting the most value out of their data, the concepts of data products within a data mesh architecture become essential. What is less clear, however, is how to get started on building a data mesh and how data pipelines, including streaming pipelines, can help teams produce, share and consume self-service data products. That’s all while ensuring the appropriate quality controls and security policies are applied to get the data from the right teams, to the right places, in the right format, at the right time.

[Watch the full data mesh conversation](#) 



Michele Goetz, VP, Principal Analyst, Forrester

Michele Goetz serves enterprise architects, chief data officers, and business analysts trying to navigate the complexities of data while running an insight-driven business. Her research covers artificial intelligence technologies and consultancies, semantic technology, data management strategy, data governance, and data integration.

1. What are some common challenges in building a data mesh?

The first hurdle organizations face is ownership. There are considerations for the domain, the data products, and the components and services. Decentralization, particularly in the lens of self-service and data governance, exacerbates the ownership challenge because business subject matter experts, analysts, and data engineers all have different ownership stakes and responsibilities for the data products and their assets. The next challenge is how to define a data product. It can be as simple as a data set or as complex as a machine learning model running on a streaming pipeline. Organizations need to define a value for the data product up front, to the point that the value can monetize the data. For example, if the product is a data set, how is subscribing to that data set helping internal teams to make a decision on the creditworthiness of a new customer?

2. What role does data governance (or the lack thereof) play in building a data mesh?

When we apply data mesh principles to streaming real-time use cases, data governance becomes critical. Traditional data governance approaches like data quality or master data management can introduce significant latency or cause inconsistencies and conflicts in data pipelines and schemas. To avoid these issues, the business logic to comply with policies is maintained as a service that can be subscribed to within a streaming pipeline. That way, data governance audits and rules address data trust (data and schema quality) in a pipeline and across topics in context of data consumption. We can think about this as data governance on read (applied in motion) rather than data governance on write (applied to data at rest). Data governance in this federated model provides more granular trust and control to meet the increased complexity of our digital distributed ecosystems.

3. How do you get started with implementing a data mesh?

Start with a business process. Data mesh demands context, with the business value and objective defined up front. A business process provides that. If you are onboarding a new customer, managing commerce transactions, or orchestrating a supply chain, each of these processes help define what data is needed, when it is needed, and how quickly an insight or action needs to be taken. And importantly, the business process indicates the type of data governance policies required and where and when to best apply those policies. If organizations start with the data rather than the process, they are constantly fishing for use cases, leaving data platform investments to languish.

4. Data mesh is more than just technology as it involves people and processes. With that in mind, what should tech leaders going down this path be aware of?

Data products can be shared via real-time streaming pipelines, a core data engineering responsibility. However, data value, data trust, and data governance are defined by a business use case. It is the line-of-business experts, developers, analysts and data scientists, and data stewards who have the most knowledge and awareness of how data products need to function. Centralized data engineering teams often do not have the right proximity to the line of business for this context. Decentralizing data engineers and moving them into line-of-business solution teams with application developers can alleviate this disconnect. In addition, decentralization is a key principle of data mesh, ensuring that technical resources and experts are more closely aligned to their

business subject matter expert (SME) counterparts. As business SMEs design data-driven and intelligent digital experiences, they should work closely with technical teams to build prototypes and wireframes to create proof of concepts. Those can be refined into production-ready data products faster and easier, since knowledge is shared in collaboration and technology teams are able to access available components to rapidly develop new applications.

5. Are there any anecdotes you can share or success stories you have heard from organizations that have successfully built a data mesh?

A pharmacy created a digital experience to help schedule and manage COVID testing and vaccinations. Their data mesh approach focused on this experience and customer journey to determine the data domains and ownership, and explore existing data artifacts (data sets, pipelines, policies, APIs, etc.) that could be used as starting points. Taking advantage of existing online pharmacy, scheduling, and flu vaccination capabilities allowed the pharmacy to not only support and offer COVID testing and vaccinations quickly, but also integrate that with existing customer and patient accounts for a complete patient view. While the initial platform for scheduling and vaccinations was using traditional data platform and integration capabilities, the initiative for COVID allowed the company to modernize and move to a streaming platform for real-time access and updates across all pharmacy and health services.